Herbert W. Hamber

# Quantum Gravitation

## The Feynman Path Integral Approach

Springer

Quantum Gravitation

Herbert W. Hamber

# Quantum Gravitation

## The Feynman Path Integral Approach

Springer

Prof. Dr. Herbert W. Hamber
University of California, Irvine
Department of Physics and Astronomy
Frederick Reines Hall
Irvine, CA 92697, USA

ISBN: 978-3-540-85292-6 e-ISBN: 978-3-540-85293-3

Library of Congress Control Number: 2008936266



*Cover design:* eStudio Calamar S.L.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

*To Franca, Jane and Lawrence*

# Preface

Almost a century has gone by since the discovery of general relativity and quantum mechanics, yet the goal of finding a consistent theory of quantum gravity nonetheless remains elusive. After the two major triumphs of modern quantum field theory, quantum electrodynamics and the quantization of non-abelian gauge theories (including quantum chromodynamics and the electro-weak theory) the early seventies provided high hopes that a quantum treatment of general relativity might be around the corner. However, to the dismay of many, the results of t' Hooft and Veltman conclusively established that quantum gravity is not perturbatively renormalizable, thus confirming earlier suspicions based on purely dimensional arguments. Disturbingly, the divergences which appear in gravity at one loop order in the semiclassical expansion, involving curvature squared terms, cannot be re-absorbed into a redefinition of the coupling constants, thereby making it difficult to derive unambiguous statements about the properties of the underlying quantum theory. More importantly, the now exhaustively explored examples of quantum electrodynamics and non-abelian gauge theories have established that until these ultraviolet renormalization effects are consistently and systematically brought under control, it will be very difficult to make any sort of physically relevant predictions. To this day, the ultraviolet problems of quantum gravity border on the speculative for many: after all, if quantum gravity effects are relevant at distances of the order of the Planck length ($10^{-33}$cm), then these might very well have little relevance for laboratory particle physics in the foreseeable future. But how could one so conclude without actually doing the relevant calculations? What if new, non-perturbative scales arise in the renormalization procedure, as occurs in non-abelian gauge theories?

Since the seventies, strategies that deal with the problem of ultraviolet divergences in quantum gravity have themselves diverged. Some have advocated the search for a new theory of quantum gravity, a theory which does not suffer from ultraviolet infinity problems. In supersymmetric theories, such as supergravity and ten-dimensional superstrings, new and yet unobserved particles are introduced thus reducing the divergence properties of Feynman amplitudes. In other, very restricted classes of supergravity theories in four dimensions, proponents have claimed that enough conspiracies might arise thereby making these models finite. For superstrings, which live in a ten-dimensional spacetime, one major obstacle prevails to date: what dynamical mechanism would drive the compactification of spacetime from the ten dimensional string universe to our physical four-dimensional world, or for that matter, to any other dimension less than ten?

A second approach to quantum gravity has endeavored to pursue new avenues to quantization, by introducing new quantum variables and new cutoffs, which involve quantum Hamiltonian methods based on parallel transport loops, spacetime spin foam and new types of quantum variables describing quantum dust. It is characteristic of these methods that the underlying theory is preserved, it essentially remains a quantum version of Einstein's relativistic theory, yet the ideas employed are intended to go past the perturbative treatment. While some of these innovative tools have had limited success in exploring the much simpler non-perturbative features of ordinary gauge theories, proponents of such methods argue that gravity is fundamentally different, thereby necessitating the use of new methods.

The third approach to quantum gravity, which forms the underlying topic of this book, focuses on the application of modern methods of quantum field theory, whose cornerstones include the manifestly covariant Feynman path integral approach, Wilson's modern renormalization group ideas and the development of lattice methods to define a regularized form of the path integral, which then allows non-perturbative calculations. In non-ablelian gauge theories and in the standard model of elementary particle interactions, said methods are invariably the the tools of choice; the covariant Feynman path integral approach is crucial in proving the renormalizability of non-abelian gauge theories; modern renormalization group methods establish the core of the derivation of the asymptotic freedom result and the related discussion of momentum dependence of amplitudes in terms of running coupling constants; and finally, the lattice formulation of gauge theories, which thus far provides the only convincing theoretical evidence of confinement and chiral symmetry breaking in non-abelian gauge theories.

Therefore, this book shall cover key aspects and open issues related to a consistent regularized formulation of quantum gravity from the perspective of the covariant Feynman path integral quantization. In the author's opinion, such a formulation is an important and essential step towards a quantitative and controlled investigation of the physical content of the theory.

## An Outline of the Book

This book is composed of three major sections. Part I introduces basic elements of the covariant formulation of continuum quantum gravity, with some emphasis on those issues which bear an immediate relevance for the remainder of the book. Discussion will include the nature of the spin-two field, its wave equation and possible gauge choices, the Feynman propagator, the coupling of a spin two field to matter and the implementation of a consistent local gauge invariance to all orders, ultimately leading to the Einstein gravitational action. Additional terms in the gravitational action, such as the cosmological constant and higher derivative contributions, are naturally introduced at this stage.

A section on the perturbative weak field expansion will later introduce the main aspects of background field method as applied to gravity, including issues such as the choice of field parametrization and gauge fixing. Then, results related to the structure of one- and two-loop divergences in pure gravity shall be discussed, leading up to the statement of perturbative non-renormalizability for the Einstein theory in four dimensions. One important aspect that will be stressed is that perturbative methods generally rely on a weak field expansion for the metric fluctuations, and are therefore not necessarily well suited for the investigation of potentially physically relevant regime of large metric fluctuations.

Next, the Feynman path integral for gravitation will be introduced by closely analogizing the theory with the related Yang-Mills case. The discussion brings up the thorny issue of the gravitational functional measure, a threshold requirement used to define Feynman's sum over histories, as well some other important aspects related to the convergence of the path integral and derived quantum averages, along with the origin of the conformal instability affecting the Euclidean case. Emphasis will be drawn to the strongly constrained nature of the theory, which depends on the absence of matter, and its close analogy to pure Yang-Mills theories, on a single dimensionless parameter $G\lambda$, besides the required usual short distance cutoff.

Since quantum gravity is not perturbatively renormalizable, the next question arises naturally: what other theories are not perturbatively renormalizable, and what can be derived from those theories? The following sections will thus summarize the methods of Wilson's $2+\varepsilon$ expansion as applied to gravity, expanding the deviation of the space-time dimensions from two; in such a dimension the gravitational coupling becomes dimensionless and the theory is therefore power-counting renormalizable. As an initial motivation, but also for illustrative and pedagogical purposes, the non-linear sigma model is introduced first. The latter represents a reasonably well understood perturbatively non-renormalizable theory above two dimensions which is characterized by a rich two-phase structure, and whose scaling properties in the vicinity of the fixed point can nevertheless be accurately computed in three dimensions (via the $2+\varepsilon$ expansion, as well as by other methods which include the strong coupling expansions and a variety of other lattice approximation techniques), and whose universal predictions are known to compare favorably with experiments. Within the context of gravity, which is discussed next the main results of the perturbative expansion are the existence of a nontrivial ultraviolet fixed point close to the origin above two dimensions (a phase transition in statistical field theory language), and the predictions of non-trivial universal scaling exponents in the vicinity of the new fixed point.

Generally, discussion of the quantization of gravity without referring in some detail to the Hamiltonian formulation is not possible. As in ordinary non-relativistic quantum mechanics, there are a number of important physical results which are obtained much more readily using this approach. Particularly notable, in the case of gravity, involves the nature of the Hamiltonian constraint, which implies that the total energy of a quantum gravitational system is zero, and the Wheeler-DeWitt equation, a Schrödinger-like equation for the vacuum functional, whose solution in some simple cases can be obtained using reduced phase space (minisuperspace) methods. In addition, the Hamiltonian method can be used as a starting point for a lattice description of quantum gravity, whose results may be regarded as complementary to those obtained via the Feynman path integral approach. The ambiguities that appear here as operator ordering problems have their correspondence in the path integral approach, under the rubric of issue associated with the choice of functional measure. The Hamiltonian approach also presents additional problems, including the lack of covariance due to the choice of time coordinate, and the difficulty of doing practical approximate non-perturbative calculations. Closely related to Hamiltonian approach is an array of semiclassical methods which have been used to obtain approximate cosmological solutions to the Wheeler-DeWitt equations, which are discussed later in some detail. The section ends with the exposition of some physically relevant results such as black hole radiance, and some more general issues which arise in a semiclassical treatment of quantum gravity.

Part II discusses the lattice theory of gravity based on Regge's simplicial formulation, with a primary focus on the physically relevant four-dimensional case. The starting point is a description of discrete manifolds in terms of edge lengths and incidence matrices, then moving on to a description of curvature in terms of deficit angles, thereby offering a re-formulation of continuum gravity in terms of a discrete action and a set of lattice field equations. The direct and clear correspondence between lattice quantities (edges, dihedral angles, volumes, deficit angles, etc.) and continuum operators (metric, affine connection, volume element, curvature tensor etc.) allows one to define, as an example, discrete versions of curvature squared terms which arise in higher derivative gravity theories, or more generally as radiatively induced corrections. An important element in the lattice-to-continuum correspondence is the development of the lattice weak field expansion, allowing a clear and precise identification between lattice and continuum degrees of freedom, as well as their gauge invariance properties, as illustrated for example in the weak field limit by the arbitrariness in the assignments of edge lengths used to cover a given physical geometry. On the lattice one can then see how the lattice analogs of gravitons arise naturally, and how their transverse-traceless nature is made evident.

When coupling matter fields to lattice gravity one needs to introduce new fields localized on vertices, as well as appropriate dual volumes which enter the definition of the kinetic terms for those fields. In the fermion case, it is necessary (as in the continuum) to introduce vierbein fields within each simplex, and then use an appropriate spin rotation matrix to relate spinors between neighboring simplices. In general the formulation of fractional spin fields on a simplicial lattice is useful in formulating a lattice discretization of supergravity. At this point it will be useful to compare and contrast the simplicial lattice formulation to other discrete approaches to quantum gravity such as, the hypercubic (vierbien-connection) lattice formulation and simplified fixed-edge-length approaches such as dynamical triangulations.

Subsequent sections deal with the interesting problem of what gravitational observables should look like, that is which expectation values of operators (or ratios thereof) have meaning and physical interpretation in the context of a manifestly covariant formulation, specifically in a situation where metric fluctuations are not necessarily bounded. Such averages naturally include expectation values of the integrated scalar curvature and other related quantities (involving for example curvature squared terms), as well as correlations of operators at fixed geodesic distance, sometimes referred to as bi-local operators. Another set of physical averages refer to the geometric nature of space-time itself, such as the fractal dimension. Finally, one more set of physical observables correspond to the gravitational analog of the Wilson loop, which provides information about the parallel transport of vectors, and therefore on the effective curvature, around large near-planar loops, and the correlation between particle world-lines, which gives the static gravitational potential. These quantities play an important role in the physical characterization of the two phases of gravity, as seen both in the $2+\varepsilon$ expansion and in the lattice formulation in four dimensions.

Part III of the book discusses applications of the lattice theory to non-perturbative gravity. Ultimately, investigations of the strongly coupled regime of quantum gravity where metric fluctuations cannot be assumed to be small, requires the use of numerical methods applied to the lattice theory. A discrete formulation combined with numerical tools can therefore be viewed as an essential step towards a quantitative and controlled investigation of the physical content of the theory: that is, in the same way that a discretization of a complicated ordinary differential equation can be viewed as a mean to determine the properties of its solution with arbitrary accuracy. These methods are outlined next, together with a summary of the main lattice results, showing the existence of two phases, depending on the value of the bare gravitational coupling, and in good agreement with the qualitative predictions of the $2+\varepsilon$ expansion. Specifically, lattice gravity in four dimensions is characterized by two phases: a weak coupling degenerate polymer-like phase, and a strong coupling smooth phase with small average curvature. The somewhat technical aspect of the determination of universal critical exponents and non-trivial scaling dimensions, based on finite size methods, is outlined, together with a brief discussion of how the lattice continuum limit has to be approached in the vicinity of a non-trivial ultraviolet fixed point.

The determination of non-trivial scaling dimensions in the vicinity of the fixed point leads to a discussion of the renormalization group properties of fundamental couplings, that is their scale dependence, as well as the emergence of physical renormalization group invariant quantities, such as the fundamental gravitational correlation length and the closely related gravitational condensate. These are discussed next, with an eye towards physical applications. This includes a discussion of the physical nature of the expected quantum corrections to the gravitational coupling, based, in part on an analogy to QED and QCD, on the effects of a virtual graviton cloud, and of how the two phases of lattice gravity relate to the two opposite scenarios of gravitational screening (for weak coupling, and therefore unphysical due to the branched polymer nature of this phase) versus anti-screening (for strong coupling, and therefore physical).

A final section touches on the general problem of formulating running gravitational couplings in a context that does not assume weak gravitational fields. The discussion includes a brief presentation on the topic of covariant running of $G$ based on the formalism of non-local field equations, with the scale dependence of $G$ expressed through the use of a suitable covariant d'Alembertian. Simple applications to standard metrics (static isotropic and homogeneous isotropic) are briefly summarized and their potential physical consequences and interpretation elaborated. The book ends with a general outlook on future prospects for lattice studies of quantum gravity, some open questions and work that can be done to help elucidate the relationship between discrete and continuum models, such as extending the range of problems addressed by the lattice, and providing new impetus for further developments in covariant continuum quantum gravity.

One final comment on the notation used in this book. Unless stated otherwise, the same notation is used as in (Weinberg, 1972), with the sign of the Riemann tensor reversed; the signature in the Lorentzian case is therefore $-,+,+,+$. In the Euclidean case $t = -i\tau$ on the other hand the flat metric is of course the Kronecker $\delta_{\mu\nu}$, with the same conventions as before for the Riemann tensor.

Berlin, May 2008 *Herbert W. Hamber*

# Acknowledgements

Over the years I have greatly benefitted from discussions with James Bjorken, Thibault Damour, Stanley Deser, James Hartle, Hikaru Kawai, Yoshihisa Kitazawa, Giorgio Parisi, Mike Peskin, Tullio Regge, Gabriele Veneziano and Jean Zinn-Justin, and many other colleagues, whose insights, benevolent criticisms and encouragements have affected the content of this work. Most of all I would like to thank Ruth Williams for her longstanding collaboration on a number of key topics covered extensively in this book, and for her careful reading of many parts of the manuscript. I also wish to thank Ido Adam, Hermann Nicolai, Alvaro Restuccia, Robert Schrader and Stefan Theisen for their diligent reading, constructive criticism and helpful comments on an early draft of the manuscript. Finally I wish to thank my students, who over the years were exposed, both in the quantum field theory and general relativity classes, to some of the topics appearing here now in book form.

The writing of this book would not have been possible without a generous research grant from the Max Planck Gesellschaft zur Förderung der Wissenschaften. The author wishes to thank Hermann Nicolai and the staff at the Max Planck Institut für Gravitationsphysik (Albert-Einstein-Institut) in Berlin-Potsdam for a very warm hospitality, and a memorable stay.

# Contents


**1 Continuum Formulation** ........ 1
- 1.1 General Aspects ........ 1
- 1.2 Massless Spin Two Field ........ 1
- 1.3 Wave Equation ........ 2
- 1.4 Feynman Rules ........ 11
- 1.5 One-Loop Divergences ........ 16
- 1.6 Gravity in $d$ Dimensions ........ 21
- 1.7 Higher Derivative Terms ........ 24
- 1.8 Supersymmetry ........ 33
- 1.9 Supergravity ........ 37
- 1.10 String Theory ........ 40
- 1.11 Supersymmetric Strings ........ 48

**2 Feynman Path Integral Formulation** ........ 55
- 2.1 The Path Integral ........ 55
- 2.2 Sum over Paths ........ 56
- 2.3 Eulidean Rotation ........ 58
- 2.4 Gravitational Functional Measure ........ 59
- 2.5 Conformal Instability ........ 64

**3 Gravity in $2+\varepsilon$ Dimensions** ........ 67
- 3.1 Dimensional Expansion ........ 67
- 3.2 Perturbatively Non-renormalizable Theories: The Sigma Model ........ 68
- 3.3 Non-linear Sigma Model in the Large-$N$ Limit ........ 79
- 3.4 Self-coupled Fermion Model ........ 83
- 3.5 The Gravitational Case ........ 84
- 3.6 Phases of Gravity in $2+\varepsilon$ Dimensions ........ 92
- 3.7 Running of $\alpha(\mu)$ in Gauge Theories ........ 98

**4 Hamiltonian and Wheeler-DeWitt Equation** . . . 103
4.1 Classical Initial Value Problem . . . 103
4.2 First Order Formulation . . . 104
4.3 Arnowitt-Deser-Misner (ADM) Formalism . . . 107
4.4 Orthogonal Decomposition in the Linearized Theory . . . 110
4.5 Intrinsic and Extrinsic Curvature, Hamiltonian . . . 111
4.6 Matter Source Terms . . . 112
4.7 Wheeler-DeWitt Equation . . . 114
4.8 Semiclassical Expansion of the Wheeler-DeWitt Equation . . . 115
4.9 Connection with the Feynman Path Integral . . . 117
4.10 Minisuperspace . . . 118
4.11 Solution of Simple Minisuperspace Models . . . 122
4.12 Quantum Hamiltonian for Gauge Theories . . . 127
4.13 Lattice Regularized Hamiltonian for Gauge Theories . . . 129
4.14 Lattice Hamiltonian for Quantum Gravity . . . 135

**5 Semiclassical Gravity** . . . 141
5.1 Cosmological Wavefunctions . . . 141
5.2 Semiclassical Expansion . . . 146
5.3 Pair Creation in Constant Electric Fields . . . 152
5.4 Black Hole Particle Emission . . . 154
5.5 Method of In and Out Vacua . . . 158
5.6 Complex Periodic Time . . . 163
5.7 Black Hole Evaporation . . . 166
5.8 Quantum Gravity Corrections . . . 167

**6 Lattice Regularized Quantum Gravity** . . . 169
6.1 The Lattice Theory . . . 169
6.2 General Formulation . . . 170
6.3 Volumes and Angles . . . 171
6.4 Rotations, Parallel Transports and Voronoi Loops . . . 173
6.5 Invariant Lattice Action . . . 179
6.6 Lattice Diffeomorphism Invariance . . . 184
6.7 Lattice Bianchi Identities . . . 186
6.8 Gravitational Wilson Loop . . . 188
6.9 Lattice Regularized Path Integral . . . 190
6.10 An Elementary Example . . . 195
6.11 Lattice Higher Derivative Terms . . . 198
6.12 Scalar Matter Fields . . . 203
6.13 Invariance Properties of the Scalar Action . . . 209
6.14 Lattice Fermions, Tetrads and Spin Rotations . . . 211
6.15 Gauge Fields . . . 213
6.16 Lattice Gravitino . . . 214
6.17 Alternate Discrete Formulations . . . 216
6.18 Lattice Invariance versus Continuum Invariance . . . 220

**7 Analytical Lattice Expansion Methods** . . . 225
7.1 Motivation . . . 225
7.2 Lattice Weak Field Expansion and Transverse-Traceless Modes . . . 225
7.3 Lattice Diffeomorphism Invariance . . . 234
7.4 Strong Coupling Expansion . . . 242
7.5 Gravitational Wilson Loop . . . 248
7.6 Discrete Gravity in the Large-*d* Limit . . . 260
7.7 Mean Field Theory . . . 269

**8 Numerical Studies** . . . 273
8.1 Nonperturbative Gravity . . . 273
8.2 Observables, Phase Structure and Critical Exponents . . . 274
8.3 Invariant Local Gravitational Averages . . . 276
8.4 Invariant Correlations at Fixed Geodesic Distance . . . 278
8.5 Wilson Lines and Static Potential . . . 281
8.6 Scaling in the Vicinity of the Critical Point . . . 284
8.7 Physical and Unphysical Phases . . . 285
8.8 Numerical Determination of the Scaling Exponents . . . 290
8.9 Renormalization Group and Lattice Continuum Limit . . . 295
8.10 Curvature Scales . . . 300
8.11 Gravitational Condensate . . . 302

**9 Scale Dependent Gravitational Couplings** . . . 305
9.1 Renormalization Group and Scale Dependence of $G$ . . . 305
9.2 Effective Field Equations . . . 305
9.3 Poisson's Equation and Vacuum Polarization Cloud . . . 308
9.4 Static Isotropic Solution . . . 311
9.5 Cosmological Solutions . . . 314
9.6 Quantum Gravity and Mach's Principle . . . 321

**References** . . . 325

**Index** . . . 335

# Chapter 1
# Continuum Formulation

## 1.1 General Aspects

The Lagrangian for the massless spin-two field can be constructed in close analogy to what one does in the case of electromagnetism. In gravity the electromagnetic interaction $e\, j \cdot A$ is replaced by a term

$$\tfrac{1}{2}\, \kappa\, h_{\mu\nu}(x)\, T^{\mu\nu}(x), \tag{1.1}$$

where $\kappa$ is a constant to be determined later, $T^{\mu\nu}$ is the conserved energy-momentum tensor

$$\partial_\mu\, T^{\mu\nu}(x) = 0 \ , \tag{1.2}$$

associated with the sources, and $h_{\mu\nu}(x)$ describes the gravitational field. It will be shown later that $\kappa$ is related to Newton's constant $G$ by $\kappa = \sqrt{16\pi G}$.

## 1.2 Massless Spin Two Field

As far as the pure gravity part of the action is concerned, one has in principle four independent quadratic terms one can construct out of the first derivatives of $h_{\mu\nu}$, namely

$$\begin{array}{lcl} \partial_\sigma h_{\mu\nu}\, \partial^\sigma h^{\mu\nu} & , & \partial^\nu h_{\mu\nu}\, \partial_\sigma h^{\mu\sigma} \ , \\ \partial^\nu h_{\mu\nu}\, \partial^\mu h_\sigma^{\ \sigma} & , & \partial^\mu h_\nu^{\ \nu}\, \partial_\mu h_\sigma^{\ \sigma} \ . \end{array} \tag{1.3}$$

The term $\partial_\sigma h_{\mu\nu}\, \partial^\nu h^{\mu\sigma}$ need not be considered separately, as it can be shown to be equivalent to the second term in the above list, after integration by parts. After combining these four terms into an action

$$\int dx\ [\ a\, \partial_\sigma h_{\mu\nu}\, \partial^\sigma h^{\mu\nu} + b\, \partial^\nu h_{\mu\nu}\, \partial_\sigma h^{\mu\sigma}$$

$$+c\,\partial^\nu h_{\mu\nu}\,\partial^\mu h_\sigma^{\ \sigma} + d\,\partial^\mu h_\nu^{\ \nu}\,\partial_\mu h_\sigma^{\ \sigma}$$
$$+\tfrac{1}{2}\,\kappa h_{\mu\nu}\,T^{\mu\nu}\,]\ , \tag{1.4}$$

and performing the required variation with respect to $h_{\alpha\beta}$, one obtains for the field equations

$$\begin{aligned}
&2a\ \partial_\sigma\partial^\sigma h_{\alpha\beta}\\
+&\ \ b\ \ (\partial_\beta\partial^\sigma h_{\alpha\sigma} + \partial_\alpha\partial^\sigma h_{\beta\sigma})\\
+&\ \ c\ \ (\partial_\alpha\partial_\beta h_\sigma^{\ \sigma} + \eta_{\alpha\beta}\,\partial_\mu\partial_\nu h^{\mu\nu})\\
+&\,2d\ \eta_{\alpha\beta}\,\partial_\mu\partial^\mu h_\sigma^{\ \sigma}\\
&\quad = \tfrac{1}{2}\,\kappa\,T_{\alpha\beta}\ ,
\end{aligned} \tag{1.5}$$

with $\eta_{\alpha\beta} = \mathrm{diag}(-1,1,1,1)$. Consistency requires that the four-divergence of the above expression give zero on both sides, $\partial^\beta(\ldots) = 0$. After collecting terms of the same type, one is led to the three conditions

$$\begin{aligned}
(2a+b)\,\partial_\sigma\partial^\sigma\partial^\beta\,h_{\alpha\beta} &= 0\\
(b+c)\,\partial_\alpha\partial^\beta\partial^\sigma\,h_{\beta\sigma} &= 0\\
(c+2d)\,\partial_\alpha\partial_\beta\partial^\beta\,h_\sigma^{\ \sigma} &= 0\ ,
\end{aligned} \tag{1.6}$$

with unique solution (up to an overall constant, which can be reabsorbed into $\kappa$) $a = -\frac{1}{4}$, $b = \frac{1}{2}$, $c = -\frac{1}{2}$ and $d = \frac{1}{4}$. As a result, the quadratic part of the Lagrangian for the pure gravitational field is given by

$$\begin{aligned}
\mathscr{L}_{sym} = &-\tfrac{1}{4}\,\partial_\sigma h_{\mu\nu}\,\partial^\sigma h^{\mu\nu} + \tfrac{1}{2}\,\partial^\nu h_{\mu\nu}\,\partial_\sigma h^{\mu\sigma}\\
&-\tfrac{1}{2}\,\partial^\nu h_{\mu\nu}\,\partial^\mu h_\sigma^{\ \sigma} + \tfrac{1}{4}\,\partial^\mu h_\nu^{\ \nu}\,\partial_\mu h_\sigma^{\ \sigma}\ .
\end{aligned} \tag{1.7}$$

## 1.3 Wave Equation

One notices that the field equations of Eq. (1.5) take on a particularly simple form if one introduces trace reversed variables $\bar{h}_{\mu\nu}(x)$,

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \tfrac{1}{2}\,\eta_{\mu\nu}\,h_\sigma^{\ \sigma} \tag{1.8}$$

and

$$\bar{T}_{\mu\nu} = T_{\mu\nu} - \tfrac{1}{2}\,\eta_{\mu\nu}\,T_\sigma^{\ \sigma}\ . \tag{1.9}$$

In the following it will be convenient to write the trace as $h = h^\sigma_{\ \sigma}$ so that $\bar{h}^\sigma_{\ \sigma} = -h$, and define the d'Alembertian as $\Box = \partial_\mu\partial^\mu = \nabla^2 - \partial_t^2$. Then the field equations become simply

$$\Box\,h_{\mu\nu} - 2\,\partial_\nu\,\partial^\sigma\,\bar{h}_{\mu\sigma} = -\kappa\,\bar{T}_{\mu\nu}\ . \tag{1.10}$$

One important aspect of the field equations is that they can be shown to be invariant under a local gauge transformation of the type

$$h'_{\mu\nu} = h_{\mu\nu} + \partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu \ , \tag{1.11}$$

involving an arbitrary gauge parameter $\varepsilon_\mu(x)$. This invariance is therefore analogous to the local gauge invariance in QED, $A'_\mu = A_\mu + \partial_\mu \varepsilon$. In the quantum theory it implies the existence of Ward identities. Furthermore, it suggests choosing a suitable gauge (analogous to the familiar Lorentz gauge $\partial^\mu A_\mu = 0$) in order to simplify the field equations, for example

$$\partial^\sigma \bar{h}_{\mu\sigma} = 0 \ , \tag{1.12}$$

which is usually referred to as the harmonic gauge condition. Then the field equations in this gauge become simply

$$\Box h_{\mu\nu} = -\kappa \bar{T}_{\mu\nu} \ . \tag{1.13}$$

These can then be easily solved in momentum space ($\Box \to -k^2$) to give

$$h_{\mu\nu} = \kappa \frac{1}{k^2} \bar{T}_{\mu\nu} \ , \tag{1.14}$$

or, in terms of the original $T_{\mu\nu}$,

$$h_{\mu\nu} = \kappa \frac{1}{k^2} \left( T_{\mu\nu} - \tfrac{1}{2} \eta_{\mu\nu} T_\sigma{}^\sigma \right) \ . \tag{1.15}$$

It should be clear that this gauge is particularly convenient for practical calculations, since then graviton propagation is given simply by a factor of $1/k^2$; later on gauge choices will be introduced where this is no longer the case.

Next one can compute the amplitude for the interaction of two gravitational sources characterized by energy-momentum tensors $T$ and $T'$. From Eqs. (1.1) and (1.14) one has

$$\tfrac{1}{2} \kappa T'_{\mu\nu} h^{\mu\nu} = \tfrac{1}{2} \kappa^2 T'_{\mu\nu} \frac{1}{k^2} \bar{T}^{\mu\nu} \ , \tag{1.16}$$

which can be compared to the electromagnetism result $j'_\mu \frac{1}{k^2} j^\mu$.

To fix the value of the parameter $\kappa$ it is easiest to look at the static case, for which the only non-vanishing component of $T_{\mu\nu}$ is $T_{00}$. Then

$$\tfrac{1}{2} \kappa^2 T'_{00} \frac{1}{k^2} \bar{T}^{00} = \tfrac{1}{2} \kappa^2 T'_{00} \frac{1}{k^2} \left( T_{00} - \tfrac{1}{2} \eta_{00} T_0{}^0 \right) \ . \tag{1.17}$$

For two bodies of mass $M$ and $M'$ the static instantaneous amplitude (by inverse Fourier transform, thus replacing $\frac{4\pi}{\mathbf{k}^2} \to \frac{1}{r}$) then becomes

$$-\tfrac{1}{2} \kappa^2 \tfrac{1}{2} M' \frac{1}{4\pi r} M' \ , \tag{1.18}$$

which, by comparison to the expected Newtonian potential energy $-GMM'/r$, gives the desired identification $\kappa = \sqrt{16\pi G}$.

The pure gravity part of the action in Eq. (1.7) only propagates transverse traceless modes (shear waves). These correspond quantum mechanically to a particle of zero mass and spin two, with two helicity states $h = \pm 2$, as shown for example in (Weinberg, 1972) by looking at the nature of plane wave solutions to the wave equation in the harmonic gauge. Helicity 0 and $\pm 1$ appear initially, but can be made to vanish by a suitable choice of coordinates.

It is relatively easy to see that the particle described by the field $h_{\mu\nu}$ has spin two. Consider the wave equation for $h_{\mu\nu}$ in the absence of sources, and subject to the harmonic gauge condition

$$\Box h_{\mu\nu} \; = \; 0 \;, \qquad \partial_\mu h^\mu{}_\nu \; = \; \tfrac{1}{2} \partial_\nu h^\lambda{}_\lambda \;\; . \tag{1.19}$$

A plane wave

$$h_{\mu\nu}(x) \; = \; e_{\mu\nu}(k) \, e^{ik\cdot x} + e^*_{\mu\nu}(k) \, e^{-ik\cdot x} \;\; , \tag{1.20}$$

will satisfy the wave equation and the harmonic gauge condition, provided $k^2 = 0$ and

$$k_\mu \, e^\mu{}_\nu \; = \; \tfrac{1}{2} k_\nu \, e^\lambda{}_\lambda \;\; , \tag{1.21}$$

respectively. Initially the polarization tensor $e_{\mu\nu} = e_{\nu\mu}$ has ten independent components, which then get reduced to six by the imposition of the harmonic gauge condition. These get further reduced to just two when one allows for local coordinate changes $x^\mu \to x^\mu + \varepsilon^\mu(x)$, which here modify $h_{\mu\nu}$ according to

$$h'^{\mu\nu} \; = \; h^{\mu\nu} - \partial^\mu \varepsilon^\nu - \partial^\nu \varepsilon^\mu \;\; . \tag{1.22}$$

To see the effects of this invariance, one first expands $\varepsilon^\mu(x)$ as well in plane waves,

$$\varepsilon^\mu(x) \; = \; i\,\varepsilon(k) \, e^{ik\cdot x} - i\,\varepsilon(k) \, e^{-ik\cdot x} \;\; , \tag{1.23}$$

so that under the gauge transformation $h_{\mu\nu} \to h'_{\mu\nu}$ one has

$$h'_{\mu\nu} \; = \; e'_{\mu\nu} \, e^{ik\cdot x} + e'^*_{\mu\nu} \, e^{-ik\cdot x} \;\; , \tag{1.24}$$

with new polarization tensor

$$e'_{\mu\nu} \; = \; e_{\mu\nu} + k_\mu \, \varepsilon_\nu + k_\nu \, \varepsilon_\mu \;\; . \tag{1.25}$$

One can explicitly verify that the new polarization tensor $e'_{\mu\nu}$ still obeys the harmonic gauge condition, $k_\mu e'^\mu{}_\nu = \frac{1}{2} k_\nu e'^\lambda{}_\lambda$, which shows that the residual gauge invariance can indeed be used to eliminate four additional degrees of freedom.

To proceed further it will be useful to consider a wave propagating along the 3 direction, for which $k^1 = k^2 = 0$ and $k^3 = k^0 = k > 0$. The harmonic gauge condition of Eq. (1.21) then gives the four equations

$$
\begin{aligned}
&e_{01}+e_{31}=0 \quad (\nu=1)\\
&e_{02}+e_{32}=0 \quad (\nu=2)\\
&e_{03}+e_{33}=\tfrac{1}{2}(e_{11}+e_{22}+e_{33}-e_{00}) \quad (\nu=3)\\
&e_{00}+e_{03}=-\tfrac{1}{2}(e_{11}+e_{22}+e_{33}-e_{00}) \quad (\nu=0)\ ,
\end{aligned}
\tag{1.26}
$$

which can be used to solve for $e_{01}$, $e_{02}$, $e_{03}$ and $e_{22}$ :

$$
e_{01} = -e_{31}\ , \quad e_{02} = -e_{32}\ , \quad e_{03} = -\tfrac{1}{2}(e_{00}+e_{33})\ , \quad e_{22} = -e_{11}\ , \tag{1.27}
$$

so that there are six leftover independent variables $e_{11}$, $e_{13}$, $e_{33}$, $e_{12}$, $e_{23}$ and $e_{00}$. Under the gauge transformation of Eq. (1.25) one then has

$$
\begin{aligned}
e'_{11} &= e_{11} & e'_{12} &= e_{12}\\
e'_{13} &= e_{13}+k\varepsilon_1 & e'_{23} &= e_{23}+k\varepsilon_2\\
e'_{33} &= e_{33}+2k\varepsilon_3 & e'_{00} &= e_{00}-2k\varepsilon_0\ ,
\end{aligned}
\tag{1.28}
$$

which shows that, for this choice of wave, only $e_{11}$ and $e_{12}$ are physical; the remaining polarization components $e'_{13}$, $e'_{23}$, $e'_{33}$ and $e'_{00}$ can be set to zero by a suitable choice of $\varepsilon_1$, $\varepsilon_2$, $\varepsilon_3$ and $\varepsilon_0$.

To determine the spin of the particle, one considers a spatial rotation $R_\mu{}^\nu(\theta)$ around the 3 axis,

$$
R_\mu{}^\nu(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 & 0\\ -\sin\theta & \cos\theta & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1 \end{pmatrix}\ , \tag{1.29}
$$

which, due to the particular choice of wave vector $k^\mu$, leaves $k^\mu$ invariant, $R_\mu{}^\nu k_\nu = k_\mu$. The polarization tensor then transforms as

$$
e'_{\mu\nu} = R_\mu{}^\lambda R_\nu{}^\sigma\, e_{\lambda\sigma}\ . \tag{1.30}
$$

But its properties are most easily expressed in terms of the following helicity amplitudes

$$
\begin{aligned}
t_\pm &= e_{11} \pm i e_{12}\\
v_\pm &= e_{31} \mp i e_{32}\ .
\end{aligned}
\tag{1.31}
$$

Under the rotation $R_\mu{}^\nu(\theta)$ these transform in a simple way,

$$
\begin{aligned}
t'_\pm &= \exp(\pm 2i\theta)\, t_\pm\\
v'_\pm &= \exp(\pm i\theta)\, v_\pm\\
e'_{33} &= e_{33}\\
e'_{00} &= e_{00}\ .
\end{aligned}
\tag{1.32}
$$

In general one expects a wave $\psi$ to transform as $\psi \to e^{ih\theta}\psi$ under a rotation by an angle $\theta$ about the direction of propagation if it has helicity $h$.

Then the results of Eq. (1.32) implies that the two physical modes $t_\pm$ (made out of polarizations $e_{11}$ and $e_{12}$) have helicity $\pm 2$. On the other hand the two unphysical modes $v_\pm$ (made out of polarizations $e_{13}$ and $e_{23}$, and which can be made to vanish by a suitable choice of gauge function $\varepsilon^\mu$) have helicity $\pm 2$. Finally, the remaining two unphysical modes $e_{33}$ and $e_{00}$ have helicity zero. The fact that only two helicities $\pm 2$ are physical implies that one is dealing with a particle of zero mass and spin two.

One would expect the gravitational field $h_{\mu\nu}$ to carry energy and momentum, which would be described by a tensor $\tau_{\mu\nu}(h)$. As in the case of electromagnetism, where one has

$$T^{(em)}_{\alpha\beta} = F_{\alpha\gamma} F_\beta{}^\gamma - \tfrac{1}{4}\,\eta_{\alpha\beta} F_{\gamma\delta} F^{\gamma\delta} \quad , \tag{1.33}$$

one would also expect such a tensor to be quadratic in the gravitational field $h_{\mu\nu}$. A suitable candidate for the energy-momentum tensor of the gravitational field is

$$\tau_{\mu\nu} = \frac{1}{8\pi G}\left( -\tfrac{1}{4}\, h_{\mu\nu}\, \partial^\lambda \partial_\lambda h^\sigma{}_\sigma + \dots \right) \quad , \tag{1.34}$$

where the dots indicate 37 possible additional terms, involving schematically, either terms of the type $h\,\partial^2 h$, or of the type $(\partial h)^2$. Such a $\tau_{\mu\nu}$ term would have to be added on the r.h.s. of the field equations in Eq. (1.10), and would therefore act as an additional source for the gravitational field (see Fig. 1.1). But the resulting field equations would then no longer invariant under Eq. (1.11), and one would have to change therefore the gauge transformation law by suitable terms of order $h^2$, so as to ensure that the new field equations would still satisfy a local gauge invariance. In other words, all these complications arise because the gravitational field carries energy and momentum, and therefore gravitates.

Ultimately, a complete and satisfactory answer to these recursive attempts at constructing a consistent, locally gauge invariant, theory of the $h_{\mu\nu}$ field is found in Einstein's non-linear General Relativity theory, as shown in (Feynman, 1962; Boulware and Deser, 1975). The full theory is derived from the Einstein-Hilbert action

$$I_E \; = \; \frac{1}{16\pi G} \int dx \sqrt{g(x)}\, R(x) \quad , \tag{1.35}$$

which generalized Eq. (1.7) beyond the weak field limit. Here $\sqrt{g}$ is the square root of the determinant of the metric field $g_{\mu\nu}(x)$, with $g = -\det g_{\mu\nu}$, and $R$ the scalar curvature. The latter is related to the Ricci tensor $R_{\mu\nu}$ and the Riemann tensor $R_{\mu\nu\lambda\sigma}$ by

$$\begin{aligned} R_{\mu\nu} &= g^{\lambda\sigma} R_{\lambda\mu\sigma\nu} \\ R &= g^{\mu\nu} g^{\lambda\sigma} R_{\mu\lambda\nu\sigma} \quad , \end{aligned} \tag{1.36}$$

where $g^{\mu\nu}$ is the matrix inverse of $g_{\mu\nu}$,

$$g^{\mu\lambda}\, g_{\lambda\nu} = \delta^\mu_\nu \quad . \tag{1.37}$$

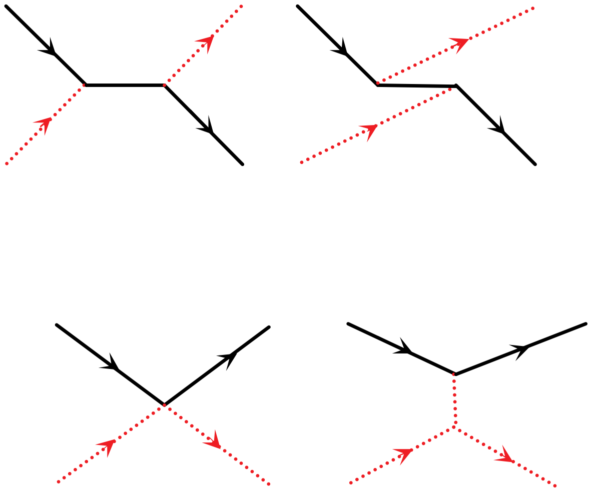


**Fig. 1.1** Lowest order diagrams illustrating the gravitational analogue to Compton scattering. *Continuous lines* indicate a matter particle, *short dashed lines* a graviton. Consistency of the theory requires that the two bottom diagrams be added to the two on the top.

In terms of the affine connection $\Gamma^{\lambda}_{\mu\nu}$, the Riemann tensor $R^{\lambda}_{\mu\nu\sigma}(x)$ is given by

$$R^{\lambda}_{\mu\nu\sigma} = \partial_{\nu}\Gamma^{\lambda}_{\mu\sigma} - \partial_{\sigma}\Gamma^{\lambda}_{\mu\nu} + \Gamma^{\eta}_{\mu\sigma}\Gamma^{\lambda}_{\nu\eta} - \Gamma^{\eta}_{\mu\nu}\Gamma^{\lambda}_{\sigma\eta} \ , \tag{1.38}$$

and therefore

$$R_{\mu\nu} = \partial_{\sigma}\Gamma^{\sigma}_{\mu\nu} - \partial_{\nu}\Gamma^{\sigma}_{\mu\sigma} + \Gamma^{\lambda}_{\sigma\lambda}\Gamma^{\sigma}_{\mu\nu} - \Gamma^{\lambda}_{\sigma\nu}\Gamma^{\sigma}_{\mu\lambda} \ , \tag{1.39}$$

with the affine connection $\Gamma^{\lambda}_{\mu\nu}(x)$ in turn constructed from components of the metric field $g_{\mu\nu}$

$$\Gamma^{\lambda}_{\mu\nu} = \tfrac{1}{2}\, g^{\lambda\sigma}\left(\partial_{\mu}g_{\nu\sigma} + \partial_{\nu}g_{\mu\sigma} - \partial_{\sigma}g_{\mu\nu}\right) \ . \tag{1.40}$$

The following algebraic symmetry properties of the Riemann tensor will be of use later

$$R_{\mu\nu\lambda\sigma} = -R_{\nu\mu\lambda\sigma} = -R_{\mu\nu\sigma\lambda} = R_{\nu\mu\sigma\lambda} \tag{1.41}$$

$$R_{\mu\nu\lambda\sigma} = R_{\lambda\sigma\mu\nu} \tag{1.42}$$

$$R_{\mu\nu\lambda\sigma} + R_{\mu\lambda\sigma\nu} + R_{\mu\sigma\nu\lambda} = 0 \ . \tag{1.43}$$

In addition, the components of the Riemann tensor satisfy the differential Bianchi identities

$$\nabla_\alpha R_{\mu\nu\beta\gamma} + \nabla_\beta R_{\mu\nu\gamma\alpha} + \nabla_\gamma R_{\mu\nu\alpha\beta} = 0 \ , \tag{1.44}$$

with $\nabla_\mu$ the covariant derivative. It is known that these, in their contracted form, ensure the consistency of the field equations. From the expansion of the Einstein-Hilbert gravitational action in powers of the deviation of the metric from the flat metric $\eta_{\mu\nu}$, using

$$\begin{aligned} R_{\mu\nu} &= \tfrac{1}{2}(\partial^2 h_{\mu\nu} - \partial_\alpha \partial_\mu h^\alpha{}_\nu - \partial_\alpha \partial_\nu h^\alpha{}_\nu + \partial_\mu \partial_\nu h^\alpha{}_\alpha) + O(h^2) \\ R &= \partial^2 h^\mu{}_\mu - \partial_\alpha \partial_\mu h^{\alpha\mu} + O(h^2) \ , \end{aligned} \tag{1.45}$$

one has for the action contribution

$$\begin{aligned} \sqrt{g} R = & -\tfrac{1}{4}\, \partial_\sigma h_{\mu\nu}\, \partial^\sigma h^{\mu\nu} + \tfrac{1}{2}\, \partial^\nu h_{\mu\nu}\, \partial_\sigma h^{\mu\sigma} \\ & -\tfrac{1}{2}\, \partial^\nu h_{\mu\nu}\, \partial^\mu h_\sigma^\sigma + \tfrac{1}{4}\, \partial^\mu h_\nu^\nu\, \partial_\mu h_\sigma^\sigma + O(h^3) \ , \end{aligned} \tag{1.46}$$

again up to total derivatives. This last expression is in fact the same as Eq. (1.7). The correct relationship between the original graviton field $h_{\mu\nu}$ and the metric field $g_{\mu\nu}$ is

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + \kappa h_{\mu\nu}(x) \ . \tag{1.47}$$

If, as is often customary, one rescales $h_{\mu\nu}$ in such a way that the $\kappa$ factor does not appear on the r.h.s., then both the $g$ and $h$ fields are dimensionless.

The weak field invariance properties of the gravitational action of Eq. (1.11) are replaced in the general theory by general coordinate transformations $x^\mu \to x'^\mu$, under which the metric transforms as a covariant second rank tensor

$$g'_{\mu\nu}(x') = \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu}\, g_{\rho\sigma}(x) \ , \tag{1.48}$$

which leaves the infinitesimal proper time interval $d\tau$ with

$$d\tau^2 = -g_{\mu\nu}\, dx^\mu dx^\nu \ , \tag{1.49}$$

invariant. In their infinitesimal form, coordinate transformations are written as

$$x'^\mu = x^\mu + \varepsilon^\mu(x) \ , \tag{1.50}$$

under which the metric at the same point $x$ then transforms as

$$\delta g_{\mu\nu}(x) = -g_{\lambda\nu}(x)\, \partial_\mu \varepsilon^\lambda(x) - g_{\lambda\mu}(x)\, \partial_\nu \varepsilon^\lambda(x) - \varepsilon^\lambda(x)\, \partial_\lambda g_{\mu\nu}(x) \ , \tag{1.51}$$

and which is usually referred to as the Lie derivative of $g$. The latter generalizes the weak field gauge invariance property of Eq. (1.11) to all orders in $h_{\mu\nu}$.

For infinitesimal coordinate transformations, one can gain some additional physical insight by decomposing the derivative of the small coordinate change $\varepsilon_\mu$ in Eq. (1.50) as

$$\frac{\partial \varepsilon_\mu}{\partial x^\nu} = s_{\mu\nu} + a_{\mu\nu} + t_{\mu\nu} \tag{1.52}$$

with

$$\begin{aligned} s_{\mu\nu} &= \tfrac{1}{d}\,\eta_{\mu\nu}\,\partial\cdot\varepsilon \\ a_{\mu\nu} &= \tfrac{1}{2}\,(\partial_\mu\varepsilon_\nu - \partial_\nu\varepsilon_\mu) \\ t_{\mu\nu} &= \tfrac{1}{2}\,(\partial_\mu\,\varepsilon_\nu + \partial_\nu\,\varepsilon_\mu - \tfrac{2}{d}\,\eta_{\mu\nu}\,\partial\cdot\varepsilon)\ . \end{aligned} \tag{1.53}$$

Then $s_{\mu\nu}(x)$ can be thought of describing local scale transformations, $a_{\mu\nu}(x)$ is written in terms of an antisymmeric tensor and therefore describes local rotations, while $t_{\mu\nu}(x)$ contains a traceless symmetric tensor and describes local shears.

Since both the scalar curvature $R(x)$ and the volume element $dx\sqrt{g(x)}$ are separately invariant under the general coordinate transformations of Eqs. (1.48) and (1.50), both of the following action contributions are acceptable

$$\begin{aligned} &\int dx\,\sqrt{g(x)} \\ &\int dx\,\sqrt{g(x)}\,R(x)\ , \end{aligned} \tag{1.54}$$

the first being known as the cosmological constant contribution (as it represents the total space-time volume). In the weak field limit, the first, cosmological constant term involves

$$\sqrt{g} = 1 + \tfrac{1}{2}\,h_\mu{}^\mu + \tfrac{1}{8}\,h_\mu{}^\mu h_\nu{}^\nu - \tfrac{1}{4}\,h_{\mu\nu}h^{\mu\nu} + O(h^3)\ , \tag{1.55}$$

which is easily obtained from the matrix formula

$$\sqrt{\det g} = \exp(\tfrac{1}{2}\,\mathrm{tr}\ln g) = \exp[\tfrac{1}{2}\,\mathrm{tr}\ln(\eta + h)]\ , \tag{1.56}$$

after expanding out the exponential in powers of $h_{\mu\nu}$. We have also reverted here to the more traditional way of performing the weak field expansion (i.e. without factors of $\kappa$),

$$\begin{aligned} g_{\mu\nu} &= \eta_{\mu\nu} + h_{\mu\nu} \\ g^{\mu\nu} &= \eta^{\mu\nu} - h_{\mu\nu} + h_\mu{}^\alpha\,h_{\alpha\nu} + \ldots \end{aligned} \tag{1.57}$$

with $\eta_{\mu\nu}$ the flat metric. The reason why such a $\sqrt{g}$ cosmological constant term was not originally included in the construction of the Lagrangian of Eq. (1.7) is that it does not contain derivatives of the $h_{\mu\nu}$ field. It is in a sense analogous to a mass term, but does have the very important property that it does not break the local gauge invariance.

This is a good place to discuss another issue. There is an old question of whether the graviton is exactly massless, or whether it has possibly a very small mass $m$. It is clear that the mass has to be very small, otherwise it would cause observable deviations from the Newtonian potential used to describe successfully large galactic cluster scales. In principle there are a number of well known problems that arise when a (Pauli-Fierz) spin-two mass term

$$\tfrac{1}{2} m^2 \left[ h^{\mu\nu} h_{\mu\nu} - (h^{\mu}{}_{\mu})^2 \right] \ , \tag{1.58}$$

is added to the gravitational Lagrangian. As in electromagnetism, such a term violates the local gauge invariance of Eq. (1.11), and in the general theory it spoils general coordinate invariance and the principle of equivalence. Since a massive spin-two particle has five polarization states and a massless one two, one has a clear mismatch in the number of physical degrees of freedom, even as $m \to 0$, as discussed originally in (van Dam and Veltman, 1970; Zakharov, 1970). In particular these authors point out the fact that the discontinuity which appears in the classical theory when the graviton mass goes to zero implies that the bending of light by the sun for massive gravitons is only $3/4$ of the experimentally confirmed general relativistic effect, thereby ruling out the possibility of a massive graviton, no matter how small its mass is. Furthermore the discontinuity does not seem to go away when a cosmological constant term is included (Dilkes et al, 2001). The latter acts in some way as a mass-like term, but does not increase the number of polarization states of the graviton since it does not break general covariance, so that a persistence of the discontinuity would be expected, for fixed cosmological constant. The problem is not entirely new, as it arises in gauge theories as well, where one finds that the only way to give the gauge particle a mass without spoiling local gauge invariance is via the Higgs mechanism. This result would suggest that a smooth limit might be obtained if a graviton mass is generated spontaneously by some sort of dynamical mechanism.

In the general theory, the energy-momentum tensor for matter $T_{\mu\nu}$ is most suitably defined in terms of the variation of the matter action $I_{\mathrm{matter}}$,

$$\delta I_{\mathrm{matter}} = \tfrac{1}{2} \int dx \sqrt{g} \, \delta g_{\mu\nu} \, T^{\mu\nu} \ , \tag{1.59}$$

and is conserved if the matter action is a scalar,

$$\nabla_{\mu} T^{\mu\nu} = 0 \ . \tag{1.60}$$

Variation of the gravitational Einstein-Hilbert action of Eq. (1.35), with the matter part added, then leads to the field equations

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R + \lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \ . \tag{1.61}$$

Here we have also added a cosmological constant term, with a scaled cosmological constant $\lambda = \lambda_0 / 16\pi G$, which follows from adding to the gravitational action a term $\lambda_0 \int \sqrt{g}$.[1]

[1] The present experimental value for Newton's constant is $\hbar G/c^3 = (1.61624(12) \times 10^{-33} cm)^2$. Recent observational evidence [reviewed in (Damour, 2006)] suggests a non-vanishing positive cosmological constant $\lambda$, corresponding to a vacuum density $\rho_{vac} \approx (2.3 \times 10^{-3} eV)^4$ with $\rho_{vac}$ related to $\lambda$ by $\lambda = 8\pi G \rho_{vac}/c^4$. As can be seen from the field equations, $\lambda$ has the same dimensions as a curvature. One has from observation $\lambda \sim 1/(10^{28} cm)^2$, so this new curvature length scale is comparable to the size of the visible universe $\sim 4.4 \times 10^{28} cm$.

One can exploit the freedom under general coordinate transformations $x'^{\mu} = f(x^{\mu})$ to impose a suitable coordinate condition, such as

$$\Gamma^{\lambda} \equiv g^{\mu\nu} \Gamma^{\lambda}_{\mu\nu} = 0 \ , \tag{1.62}$$

which is seen to be equivalent to the following gauge condition on the metric

$$\partial_{\mu} (\sqrt{g} g^{\mu\nu}) = 0 \ , \tag{1.63}$$

and therefore equivalent, in the weak field limit, to the harmonic gauge condition introduced previously in Eq. (1.12).

## 1.4 Feynman Rules

The Feynman rules represent the standard way to do perturbative calculations in quantum gravity. To this end one first expands again the action out in powers of the field $h_{\mu\nu}$ and separates out the quadratic part, which gives the graviton propagator, from the rest of the Lagrangian which gives the $O(h^3), O(h^4)\ldots$ vertices. To define the graviton propagator one also requires the addition of a gauge fixing term and the associated Faddeev-Popov ghost contribution (Feynman, 1962; Faddeev and Popov, 1967). Since the diagrammatic calculations are performed using dimensional regularization, one first needs to define the theory in $d$ dimensions; at the end of the calculations one will be interested in the limit $d \to 4$.

So first one expands around the *d-dimensional* flat Minkowski space-time metric, with signature given by $\eta_{\mu\nu} = \mathrm{diag}(-1,1,1,1,\ldots)$. The Einstein-Hilbert action in $d$ dimensions is given by a generalization of Eq. (1.35)

$$I_{\mathrm{E}} = \frac{1}{16\pi G} \int d^d x \sqrt{g(x)} R(x) \ , \tag{1.64}$$

with again $g(x) = -\det(g_{\mu\nu})$ and $R$ the scalar curvature; in the following it will be assumed, at least initially, that the bare cosmological constant $\lambda_0$ is zero. The simplest form of matter coupled in an invariant way to gravity is a set of spinless scalar particles of mass $m$, with action

$$I_{\mathrm{m}} = \tfrac{1}{2} \int d^d x \sqrt{g(x)} \left[ -g^{\mu\nu}(x) \partial_{\mu}\phi(x) \partial_{\nu}\phi(x) - m^2 \phi^2(x) \right] \ . \tag{1.65}$$

In Feynman diagram perturbation theory the metric $g_{\mu\nu}(x)$ is expanded around the flat metric $\eta_{\mu\nu}$, by writing again

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + \sqrt{16\pi G} h_{\mu\nu}(x) \ . \tag{1.66}$$

The quadratic part of the Lagrangian [see Eq. (1.7)] is then

$$\mathscr{L} = -\tfrac{1}{4}\,\partial_\mu h_{\alpha\beta}\,\partial^\mu h^{\alpha\beta} + \tfrac{1}{8}\,(\partial_\mu h^\alpha{}_\alpha)^2 + \tfrac{1}{2}\,C_\mu^2 + \tfrac{1}{2}\,\kappa\,h_{\mu\nu}\,T^{\mu\nu} + \mathscr{L}_{\mathrm{gf}} + \ldots \tag{1.67}$$

where the dots indicate terms that are either total derivatives, or higher order in $h$. A suitable gauge fixing term $C_\mu$ is given by

$$C_\mu \equiv \partial_\alpha h^\alpha{}_\mu - \tfrac{1}{2}\,\partial_\mu h^\alpha{}_\alpha \ . \tag{1.68}$$

Without such a term the quadratic part of the gravitational Lagrangian of Eq. (1.7) would contain a zero mode $h_{\mu\nu} \sim \partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu$, due to the gauge invariance of Eq. (1.11), which would make the graviton propagator ill defined.

The gauge fixing contribution $\mathscr{L}_{\mathrm{gf}}$ itself will be written as the sum of two terms,

$$\mathscr{L}_{\mathrm{gf}} = -\tfrac{1}{2}\,C_\mu^2 + \mathscr{L}_{\mathrm{ghost}} \ , \tag{1.69}$$

with the first term engineered so as to conveniently cancel the $+\frac{1}{2}C_\mu^2$ in Eq. (1.67) and thus give a well defined graviton propagator. Note incidentally that this gauge is not the harmonic gauge condition of Eq. (1.12), and is usually referred to instead as the DeDonder gauge. The second term is determined as usual from the variation of the gauge condition under an infinitesimal gauge transformation of the type in Eq. (1.11)

$$\delta C_\mu = \partial^2 \varepsilon_\mu + O(\varepsilon^2) \ , \tag{1.70}$$

which leads to the lowest order ghost Lagrangian

$$\mathscr{L}_{\mathrm{ghost}} = -\,\partial_\mu \bar{\eta}_\alpha\,\partial^\mu \eta^\alpha + O(h^2) \ , \tag{1.71}$$

where $\eta_\alpha$ is the spin-one anticommuting ghost field, with propagator

$$D^{(\eta)}_{\mu\nu}(k) = \frac{\eta_{\mu\nu}}{k^2} \ . \tag{1.72}$$

In this gauge the graviton propagator is finally determined from the surviving quadratic part of the pure gravity Lagrangian, which is

$$\mathscr{L}_0 = -\tfrac{1}{4}\,\partial_\mu h_{\alpha\beta}\,\partial^\mu h^{\alpha\beta} + \tfrac{1}{8}\,(\partial_\mu h^\alpha{}_\alpha)^2 \ . \tag{1.73}$$

The latter can be conveniently re-written in terms of a matrix $V$

$$\mathscr{L}_0 = -\tfrac{1}{2}\,\partial_\lambda h_{\alpha\beta}\,V^{\alpha\beta\mu\nu}\,\partial^\lambda h_{\mu\nu} \tag{1.74}$$

with

$$V_{\alpha\beta\mu\nu} = \tfrac{1}{2}\,\eta_{\alpha\mu}\eta_{\beta\nu} - \tfrac{1}{4}\,\eta_{\alpha\beta}\eta_{\mu\nu} \ . \tag{1.75}$$

The matrix $V$ can easily be inverted, for example by re-labeling rows and columns via the correspondence

$$11 \to 1,\ \ 22 \to 2,\ \ 33 \to 3, \ldots\ 12 \to 5,\ \ 13 \to 6,\ \ 14 \to 7\ \ \ldots \tag{1.76}$$

and the graviton Feynman propagator in $d$ dimensions is then found to be of the form

$$D_{\mu\nu\alpha\beta}(k) = \frac{\eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha} - \frac{2}{d-2}\eta_{\mu\nu}\,\eta_{\alpha\beta}}{k^2} \quad , \tag{1.77}$$

with a suitable $i\varepsilon$ prescription to correctly integrate around poles in the complex $k$ space. Equivalently the whole procedure could have been performed from the start with an Euclidean metric $\eta_{\mu\nu} \to \delta_{\mu\nu}$ and a complex time coordinate $t = -i\tau$ with hardly any changes of substance. The simple pole in the graviton propagator at $d = 2$ serves as a reminder of the fact that, due to the Gauss-Bonnet identity, the gravitational Einstein-Hilbert action of Eq. (1.64) becomes a topological invariant in two dimensions.

Higher order correction in $h$ to the Lagrangian for pure gravity then determine to order $h^3$ the three-graviton vertex, to order $h^4$ the four-graviton vertex, and so on. Because of the $\sqrt{g}$ and $g^{\mu\nu}$ terms in the action, there are an infinite number of vertices in $h$.

Had one included a cosmological constant term as in Eq. (1.55), which can also be expressed in terms of the matrix $V$ as

$$\sqrt{g} \;=\; 1 + \tfrac{1}{2}\,h_{\mu\mu} - \tfrac{1}{2}h_{\alpha\beta}V^{\alpha\beta\mu\nu}h_{\mu\nu} + O(h^3) \;\; , \tag{1.78}$$

then the expression in Eq. (1.74) would have read

$$\mathscr{L}_0 = \lambda_0(1 + \kappa\,\tfrac{1}{2}h^{\alpha}{}_{\alpha}) + \tfrac{1}{2}\,h_{\alpha\beta}\,V^{\alpha\beta\mu\nu}\,(\partial^2 + \lambda_0\kappa^2)h_{\mu\nu} \;\; , \tag{1.79}$$

with $\kappa^2 = 16\pi G$. Then the graviton propagator would have been remained the same, except for the replacement $k^2 \to k^2 - \lambda_0\kappa^2$. In this gauge it would correspond to the exchange of a particle of mass $\mu^2 = -\lambda_0\kappa^2$. The term linear in $h$ can be interpreted as a uniform constant source for the gravitational field. But one needs to be quite careful, since for non-vanishing cosmological constant flat space $g_{\mu\nu} \sim \eta_{\mu\nu}$ is no longer a solution of the vacuum field equations and the problem becomes a bit more subtle: one needs to expand around the correct vacuum solutions in the presence of a $\lambda$-term, which are no longer constant.

Another point needs to be made here. One peculiar aspect of perturbative gravity is that there is no unique way of doing the weak field expansions, and one can have therefore different sets of Feynman rules, even apart from the choice of gauge condition, depending on how one chooses to do the expansion for the metric.

For example, the structure of the scalar field action of Eq. (1.65) suggests to define instead the small fluctuation graviton field $h_{\mu\nu}(x)$ via

$$\tilde{g}^{\mu\nu}(x) \equiv g^{\mu\nu}(x)\sqrt{g(x)} = \eta^{\mu\nu} + K\,h^{\mu\nu}(x) \;\; , \tag{1.80}$$

with $K^2 = 32\pi G$ (Faddeev and Popov, 1974; Capper et al, 1973). Here it is $h^{\mu\nu}(x)$ that should be referred to as “the graviton field”. The change of variables from the $g_{\mu\nu}$’s to the $g^{\mu\nu}(x)\sqrt{g(x)}$’s involves a Jacobian, which can be taken to be one in dimensional regularization. There is one obvious advantage of this expansion over the previous one, namely that it leads to considerably simpler Feynman rules, both for the graviton vertices and for the scalar-graviton vertices, which can be advantageous when computing one-loop scattering amplitudes of scalar particles (Hamber and Liu, 1985). Even the original gravitational action has a simpler form in terms of the variables of Eq. (1.80) as shown originally in (Goldberg, 1958).

Again, when performing Feynman diagram perturbation theory a gauge fixing term needs to be added in order to define the propagator, for example of the form

$$\frac{1}{K^2}\left(\partial_\mu \sqrt{g}\, g^{\mu\nu}\right)^2 \quad . \tag{1.81}$$

In this new framework the bare graviton propagator is given simply by

$$D_{\mu\nu\alpha\beta}(k) = \frac{\eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha} - \eta_{\mu\nu}\eta_{\alpha\beta}}{2k^2} \quad , \tag{1.82}$$

which should be compared to Eq. (1.77) (the extra factor of one half here is just due to the convention in the choice of $K$). One notices that now there are no factors of $1/(d-2)$ for the graviton propagator in $d$ dimensions. But such factors appear instead in the expression for the Feynman rules for the graviton vertices, and such $(d-2)^{-1}$ pole terms appear therefore regardless of the choice of expansion field. For the three-graviton and two ghost-graviton vertex the relevant expressions are quite complicated. The three-graviton vertex is given by

$$\begin{aligned}
& U(q_1,q_2,q_3)_{\alpha_1\beta_1,\alpha_2\beta_2,\alpha_3\beta_3} = \\
& -\frac{K}{2}\Big[ q^2_{(\alpha_1} q^3_{\beta_1)} \left( 2\eta_{\alpha_2(\alpha_3}\eta_{\beta_3)\beta_2} - \tfrac{2}{d-2}\eta_{\alpha_2\beta_2}\eta_{\alpha_3\beta_3} \right) \\
& \quad + q^1_{(\alpha_2} q^3_{\beta_2)} \left( 2\eta_{\alpha_1(\alpha_3}\eta_{\beta_3)\beta_1} - \tfrac{2}{d-2}\eta_{\alpha_1\beta_1}\eta_{\alpha_3\beta_3} \right) \\
& \quad + q^1_{(\alpha_3} q^2_{\beta_3)} \left( 2\eta_{\alpha_1(\alpha_2}\eta_{\beta_2)\beta_1} - \tfrac{2}{d-2}\eta_{\alpha_1\beta_1}\eta_{\alpha_2\beta_2} \right) \\
& \quad + 2q^3_{(\alpha_2}\eta_{\beta_2)(\alpha_1}\eta_{\beta_1)(\alpha_3} q^2_{\beta_3)} + 2q^1_{(\alpha_3}\eta_{\beta_3)(\alpha_2}\eta_{\beta_2)(\alpha_1} q^3_{\beta_1)} + 2q^2_{(\alpha_1}\eta_{\beta_1)(\alpha_3}\eta_{\beta_3)(\alpha_2} q^1_{\beta_2)} \\
& \quad + q^2\cdot q^3 \left( \tfrac{2}{d-2}\eta_{\alpha_1(\alpha_2}\eta_{\beta_2)\beta_1}\eta_{\alpha_3\beta_3} + \tfrac{2}{d-2}\eta_{\alpha_1(\alpha_3}\eta_{\beta_3)\beta_1}\eta_{\alpha_2\beta_2} - 2\eta_{\alpha_1(\alpha_2}\eta_{\beta_2)(\alpha_3}\eta_{\beta_3)\beta_1} \right) \\
& \quad + q^1\cdot q^3 \left( \tfrac{2}{d-2}\eta_{\alpha_2(\alpha_1}\eta_{\beta_1)\beta_2}\eta_{\alpha_3\beta_3} + \tfrac{2}{d-2}\eta_{\alpha_2(\alpha_3}\eta_{\beta_3)\beta_2}\eta_{\alpha_1\beta_1} - 2\eta_{\alpha_2(\alpha_1}\eta_{\beta_1)(\alpha_3}\eta_{\beta_3)\beta_2} \right) \\
& \quad + q^1\cdot q^2 \left( \tfrac{2}{d-2}\eta_{\alpha_3(\alpha_1}\eta_{\beta_1)\beta_3}\eta_{\alpha_2\beta_2} + \tfrac{2}{d-2}\eta_{\alpha_3(\alpha_2}\eta_{\beta_2)\beta_3}\eta_{\alpha_1\beta_1} - 2\eta_{\alpha_3(\alpha_1}\eta_{\beta_1)(\alpha_2}\eta_{\beta_2)\beta_3} \right) \Big] \quad .
\end{aligned} \tag{1.83}$$

The ghost-graviton vertex is given by

$$V(k_1,k_2,k_3)_{\alpha\beta,\lambda\mu} = K\left[ -\eta_{\lambda(\alpha}k_{1\beta)}k_{2\mu} + \eta_{\lambda\mu}k_{2(\alpha)}k_{3\beta)} \right] \quad , \tag{1.84}$$

and the two scalar-one graviton vertex is given by

$$\frac{K}{2}\left( p_{1\mu}p_{2\nu} + p_{1\nu}p_{2\mu} - \frac{2}{d-2}\, m^2\, \eta_{\mu\nu} \right) \quad , \tag{1.85}$$

where the $p_1, p_2$ denote the four-momenta of the incoming and outgoing scalar field, respectively. Finally the two scalar-two graviton vertex is given by

$$\frac{K^2 m^2}{2(d-2)} \left( \eta_{\mu\lambda}\eta_{\nu\sigma} + \eta_{\mu\sigma}\eta_{\nu\lambda} - \frac{2}{d-2}\,\eta_{\mu\nu}\eta_{\lambda\sigma} \right) , \tag{1.86}$$

where one pair of indices $(\mu, \nu)$ is associated with one graviton line, and the other pair $(\lambda, \sigma)$ is associated with the other graviton line. These rules follow readily from the expansion of the gravitational action to order $G^{3/2}$ $(K^3)$, and of the scalar field action to order $G$ $(K^2)$, as shown in detail in (Capper et al, 1973). Note that the poles in $1/(d-2)$ have disappeared from the propagator, but have moved to the vertex functions. As mentioned before, they reflect the kinematic singularities that arise in the theory as $d \to 2$ due to the Gauss-Bonnet identity. As an illustration, Fig. 1.2 shows the lowest order diagrams contributing to the static potential between two massive spinless sources (Hamber and Liu, 1995).

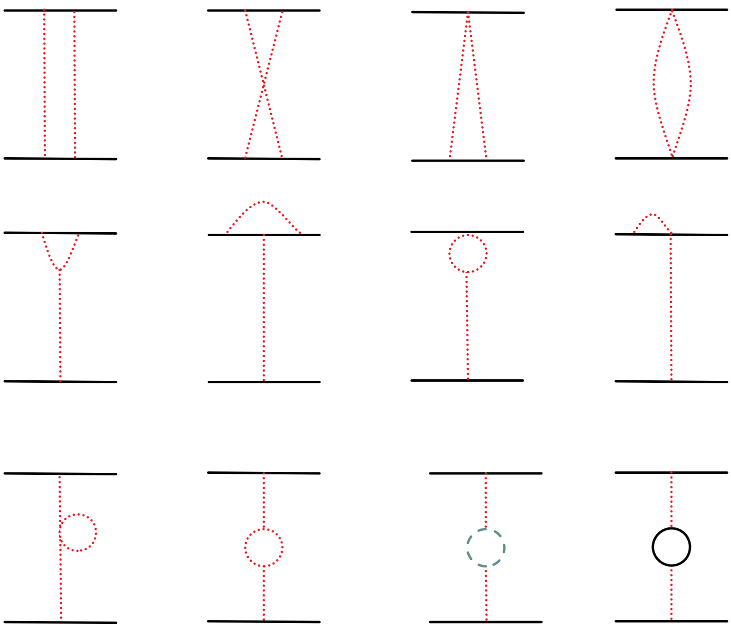


**Fig. 1.2** Lowest order diagrams illustrating modifications to the classical gravitational potential due to graviton exchange. *Continuous lines* denote a spinless heavy matter particle, *short dashed lines* a graviton and the *long dashed line* the ghost loop. The last diagram shows the scalar matter loop contribution.

## 1.5 One-Loop Divergences

Once the propagators and vertices have been defined, one can then proceed as in QED and Yang-Mills theories and evaluate the quantum mechanical one loop corrections. In a renormalizable theory with a dimensionless coupling, such as QED and Yang-Mills theories, one has that the radiative corrections lead to charge, mass and field re-definitions. In particular, for the pure $SU(N)$ gauge action one finds

$$I_{YM} = -\frac{1}{4g^2} \int dx \, \mathrm{tr} F_{\mu\nu}^2 \ \rightarrow \ -\frac{1}{4g_R^2} \int dx \, \mathrm{tr} F_{R\ \mu\nu}^2 \ , \tag{1.87}$$

so that the form of the action is preserved by the renormalization procedure: no new interaction terms such as $(D_\mu F^{\mu\nu})^2$ need to be introduced in order to re-absorb the divergences.

In gravity the coupling is dimensionful, $G \sim \mu^{2-d}$, and one expects trouble already on purely dimensional grounds, with divergent one loop corrections proportional to

$G\Lambda^{d-2}$ where $\Lambda$ is an ultraviolet cutoff.[2] Equivalently, one expects to lowest order bad ultraviolet behavior for the running Newton's constant at large momenta,

$$\frac{G(k^2)}{G} \sim 1 + \text{const.}\ G k^{d-2} + O(G^2) \ . \tag{1.88}$$

These considerations also suggest that perhaps ordinary Einstein gravity is perturbatively renormalizable in the traditional sense in two dimensions, an issue to which we will return later in Sect. 3.5.

A more general argument goes as follows. The gravitational action contains the scalar curvature $R$ which involves two derivatives of the metric. Thus the graviton propagator in momentum space will go like $1/k^2$, and the vertex functions like $k^2$. In $d$ dimensions each loop integral with involve a momentum integration $d^d k$, so that the superficial degree of divergence $\mathscr{D}$ of a Feynman diagram with $V$ vertices, $I$ internal lines and $L$ loops will be given by

$$\mathscr{D} = d\,L + 2V - 2I \ . \tag{1.89}$$

The topological relation involving $V$, $I$ and $L$

$$L = 1 + I - V \ , \tag{1.90}$$

is true for any diagram, and yields

[2] Indeed it was noticed very early on in the development of renormalization theory that perturbatively non-renormalizible theories would involve couplings with negative mass dimensions, and for which cross-sections would grow rapidly with energy (Sakata, Umezawa and Kamefuchi, 1952). It had originally been suggested by Heisenberg (Heisenberg, 1938) that the relevant mass scale appearing in such interactions with dimensionful coupling constants should be used to set an upper energy limit on the physical applicability of such theories.

$$\mathscr{D} = 2 + (d-2)\,L \ , \tag{1.91}$$

which is independent of the number of external lines. One concludes therefore that for $d > 2$ the degree of divergence increases with increasing loop order $L$.

The most convenient tool to determine the structure of the divergent one-loop corrections to Einstein gravity is the background field method (DeWitt, 1967 a,b,c; 't Hooft and Veltman, 1974) combined with dimensional regularization, wherein ultraviolet divergences appear as poles in $\varepsilon = d - 4$.[3] In non-Abelian gauge theories the background field method greatly simplifies the calculation of renormalization factors, while at the same time maintaining explicit gauge invariance.

The essence of the method is easy to describe: one replaces the original field appearing in the classical action by $A + Q$, where $A$ is a classical background field and $Q$ the quantum fluctuation. A suitable gauge condition is chosen (the background gauge), such that manifest gauge invariance is preserved for the background $A$ field. After expanding out the action to quadratic order in the $Q$ field, the functional integration over $Q$ is performed, leading to an effective action for the background $A$ field. From the structure of this effective action the renormalization of the couplings, as well as possible additional counterterms, can then be read off. In the case of gravity it is in fact sufficient to look at the structure of those terms appearing in the effective action which are quadratic in the background field $A$. A very readable introduction to the background field method as applied to gauge theories can be found in (Abbot, 1982).

Unfortunately perturbative calculations in gravity are rather cumbersome due to the large number of indices and contractions, so the rest of this section is only intended more as a general outline, with the scope of hopefully providing some of the flavor of the original calculations. The first step consists in the replacement

$$g_{\mu\nu} \ \rightarrow \ \bar{g}_{\mu\nu} = g_{\mu\nu} + h_{\mu\nu} \ , \tag{1.92}$$

where now $g_{\mu\nu}(x)$ is the classical background field and $h_{\mu\nu}$ the quantum field, to be integrated over. To determine the structure of one loop divergences it will often be sufficient to consider at the very end just the case of a flat background metric, $g_{\mu\nu} = \eta_{\mu\nu}$, or a small deviation from it.

After a somewhat tedious calculation one finds for the bare action

$$\mathscr{L} = \sqrt{g}\,[\,c_0 + c_1 R\,] \ , \tag{1.93}$$

expanded out to quadratic order in $h$

$$\mathscr{L} = \ \sqrt{g}\,[c_0\{1 + \tfrac{1}{2}h^{\alpha}{}_{\alpha} - \tfrac{1}{4}h^{\alpha}{}_{\beta}h^{\beta}{}_{\alpha} + \tfrac{1}{8}h^{\alpha}{}_{\alpha}h^{\beta}{}_{\beta}\}$$

[3] The second reference uses a complex time (Euclidean) $x_0 = ict$ notation that differs from the one used here.

$$
\begin{aligned}
&+ c_1\{R - \tfrac{1}{2}h^{\alpha}{}_{\alpha}R + h^{\alpha}{}_{\beta}R^{\beta}{}_{\alpha} - \tfrac{1}{8}Rh^{\alpha}{}_{\alpha}h^{\beta}{}_{\beta} + \tfrac{1}{4}Rh^{\alpha}{}_{\beta}h^{\beta}{}_{\alpha} \\
&- h^{\nu}{}_{\beta}h^{\beta}{}_{\alpha}R^{\alpha}{}_{\nu} + \tfrac{1}{2}h^{\alpha}{}_{\alpha}h^{\nu}{}_{\beta}R^{\beta}{}_{\nu} - \tfrac{1}{4}\nabla_{\nu}h^{\alpha}{}_{\beta}\nabla^{\nu}h^{\beta}{}_{\alpha} \\
&+ \nabla_{\nu}h^{\alpha}{}_{\alpha}\nabla^{\nu}h^{\beta}{}_{\beta} - \tfrac{1}{2}\nabla_{\beta}h^{\alpha}{}_{\alpha}\nabla^{\mu}h^{\beta}{}_{\mu} + \tfrac{1}{2}\nabla^{\alpha}h^{\nu}{}_{\beta}\nabla_{\nu}h^{\beta}{}_{\alpha}\}] \ ,
\end{aligned}
\tag{1.94}
$$

up to total derivatives. Here $\nabla_\mu$ denotes a covariant derivative with respect to the metric $g_{\mu\nu}$. For $g_{\mu\nu} = \eta_{\mu\nu}$ the above expression coincides with the weak field Lagrangian contained in Eqs. (1.7) and (1.67), with a cosmological constant term added, as given in Eq. (1.55).

To this expression one needs to add the gauge fixing and ghost contributions, as was done in Eq. (1.67). The background gauge fixing term used is

$$
-\tfrac{1}{2}C_\mu^2 = -\tfrac{1}{2}\sqrt{g}\,(\nabla_\alpha h^{\alpha}{}_{\mu} - \tfrac{1}{2}\nabla_\mu h^{\alpha}{}_{\alpha})(\nabla_\beta h^{\beta\mu} - \tfrac{1}{2}\nabla^\mu h^{\beta}{}_{\beta}) \ , \tag{1.95}
$$

with a corresponding ghost Lagrangian

$$
\mathscr{L}_{\text{ghost}} = \sqrt{g}\,\bar{\eta}_\mu(\partial_\alpha\partial^\alpha\eta^\mu - R^{\mu}{}_{\alpha}\eta^\alpha) \ . \tag{1.96}
$$

The integration over the $h_{\mu\nu}$ field can then be performed with the aid of the standard Gaussian integral formula

$$
\begin{aligned}
\ln\int[dh_{\mu\nu}] \quad & \exp\{-\tfrac{1}{2}h\cdot M(g)\cdot h - N(g)\cdot h\} \\
&= \tfrac{1}{2}N(g)\cdot M^{-1}(g)\cdot N(g) - \tfrac{1}{2}\,\mathrm{tr}\ln M(g) + \text{const.} \ ,
\end{aligned}
\tag{1.97}
$$

leading to an effective action for the $g_{\mu\nu}$ field. In practice one is only interested in the divergent part, which can be shown to be local. Specific details of the functional measure over metrics $[dg_{\mu\nu}]$ are not deemed to be essential at this stage, as in perturbation theory one is only doing Gaussian integrals, with $h_{\mu\nu}$ ranging from $-\infty$ to $+\infty$. In particular when using dimensional regularization one uses the formal rule

$$
\int d^d k = (2\pi)^d \delta^{(d)}(0) = 0 \ , \tag{1.98}
$$

which leads to some technical simplifications but obscures the role of the measure.

In the flat background field case $g_{\mu\nu} = \eta_{\mu\nu}$, the functional integration over the $h_{\mu\nu}$ fields would have been particularly simple, since then one would be using

$$
h_{\mu\nu}(x)\,h_{\alpha\beta}(x') \ \rightarrow \ <h_{\mu\nu}(x)\,h_{\alpha\beta}(x')> = \ G_{\mu\nu\alpha\beta}(x,x') \ , \tag{1.99}
$$

with the graviton propagator $G(k)$ given in Eq. (1.77). In practice, one can use the expected generally covariant structure of the one-loop divergent part

$$
\Delta\mathscr{L}_g \propto \sqrt{g}\left(\alpha R^2 + \beta R_{\mu\nu}R^{\mu\nu}\right) \ , \tag{1.100}
$$

with $\alpha$ and $\beta$ some real parameters, as well as its weak field form, obtained from

$$
\begin{aligned}
R^2 &= \partial^2 h^\mu{}_\mu \partial^2 h^\alpha{}_\alpha - 2\partial^2 h^\mu{}_\mu \partial_\alpha \partial_\beta h^{\alpha\beta} + \partial_\alpha \partial_\beta h^{\alpha\beta} \partial_\mu \partial_\nu h^{\mu\nu} \\
R_{\alpha\beta} R^{\alpha\beta} &= \tfrac{1}{4}(\partial^2 h^\mu{}_\mu \partial^2 h^\alpha{}_\alpha + \partial^2 h_{\mu\alpha} \partial^2 h^{\mu\alpha} - 2\partial^2 h^\mu{}_\mu \partial_\alpha \partial_\beta h^{\alpha\beta} \\
&\quad -2\partial_\alpha \partial_\nu h^\nu{}_\mu \partial_\alpha \partial_\beta h^{\mu\beta} + 2\partial_\mu \partial_\nu h^{\mu\nu} \partial_\alpha \partial_\beta h^{\alpha\beta}) \ ,
\end{aligned}
\tag{1.101}
$$

[compare with Eq. (1.45)], combined with some suitable special choices for the background metric, such as $g_{\mu\nu}(x) = \eta_{\mu\nu} f(x)$, to further simplify the calculation. This eventually determines the required one-loop counterterm for pure gravity to be

$$
\Delta\mathscr{L}_g = \frac{\sqrt{g}}{8\pi^2(d-4)} \left( \frac{1}{120} R^2 + \frac{7}{20} R_{\mu\nu} R^{\mu\nu} \right) \ . \tag{1.102}
$$

For the simpler case of classical gravity coupled invariantly to a single real quantum scalar field one finds

$$
\Delta\mathscr{L}_g = \frac{\sqrt{g}}{8\pi^2(d-4)} \frac{1}{120} \left( \tfrac{1}{2} R^2 + R_{\mu\nu} R^{\mu\nu} \right) \ . \tag{1.103}
$$

The complete set of one-loop divergences, computed using the alternate method of the heat kernel expansion and zeta function regularization[4] close to four dimensions, can be found in the comprehensive review (Hawking, 1977) and further references therein. In any case one is led to conclude that pure quantum gravity in four dimensions is not perturbatively renormalizable: the one-loop divergent part contains local operators which were not present in the original Lagrangian. It would seem therefore that these operators would have to be added to the bare $\mathscr{L}$, so that a consistent perturbative renormalization program can be developed in four dimensions. As a result, the field equations become significantly more complicated due to the presence of the curvature squared terms (Barth and Christensen, 1983).

There are two interesting, and interrelated, aspects of the result of Eq. (1.102). The first one is that for pure gravity the divergent part vanishes when one imposes the tree-level equations of motion $R_{\mu\nu} = 0$: the one-loop divergence vanishes on-shell. The second interesting aspect is that the specific structure of the one-loop divergence is such that its effect can actually be re-absorbed into a field redefinition,

$$
\begin{aligned}
g_{\mu\nu} &\to g_{\mu\nu} + \delta g_{\mu\nu} \\
\delta g_{\mu\nu} &\propto \frac{7}{20} R_{\mu\nu} - \frac{11}{60} R g_{\mu\nu} \ ,
\end{aligned}
\tag{1.104}
$$

which renders the one-loop amplitudes finite for pure gravity. Unfortunately this hoped-for mechanism does not seem to work to two loops, and no additional miraculous cancellations seem to occur there. At two loops one expects on general grounds

[4] The zeta-function regularization (Ray and Singer, 1971; Dowker and Critchley, 1976; Hawking, 1977) involves studying the behavior of the function $\zeta(s) = \sum_{n=0}^{\infty} (\lambda_n)^{-s}$, where the $\lambda_n$'s are the eigenvalues of the second order differential operator $M$ in question. The series will converge for $s > 2$, and can be used for an analytic continuation to $s = 0$, which then leads to the formal result $\log(\det M) = \log \prod_{n=0}^{\infty} \lambda_n = -\zeta'(0)$.

terms of the type $\nabla^4 R$, $R\nabla^2 R$ and $R^3$. It can be shown that the first class of terms reduce to total derivatives, and that the second class of terms can also be made to vanish on shell by using the Bianchi identity. Out of the last set of terms, the $R^3$ ones, one can show ('t Hooft, 2002) that there are potentially 20 distinct contributions, of which 19 vanish on shell (i.e. by using the tree level field equations $R_{\mu\nu} = 0$). An explicit calculation then shows that a new non-removable on-shell $R^3$-type divergence arises in pure gravity at two loops (Goroff and Sagnotti, 1985; van de Ven, 1992) from the only possible surviving non-vanishing counterterm, namely

$$\Delta \mathscr{L}^{(2)} = \frac{\sqrt{g}}{(16\pi^2)^2(d-4)} \frac{209}{2880} R_{\mu\nu}{}^{\rho\sigma} R_{\rho\sigma}{}^{\kappa\lambda} R_{\kappa\lambda}{}^{\mu\nu} \ . \tag{1.105}$$

To summarize, radiative corrections to pure Einstein gravity without a cosmological constant term induce one-loop $R^2$-type divergences of the form

$$\Gamma_{div}^{(1)} = \frac{1}{d-4} \frac{\hbar}{16\pi^2} \int d^4x \sqrt{g} \left( \frac{7}{20} R_{\mu\nu} R^{\mu\nu} + \frac{1}{120} R^2 \right) \ , \tag{1.106}$$

and a two-loop non-removable on-shell $R^3$-type divergence of the type

$$\Gamma_{div}^{(2)} = \frac{1}{d-4} \frac{209}{2880} \frac{\hbar^2 G}{(16\pi^2)^2} \int d^4x \sqrt{g} R_{\mu\nu}{}^{\rho\sigma} R_{\rho\sigma}{}^{\kappa\lambda} R_{\kappa\lambda}{}^{\mu\nu} \ , \tag{1.107}$$

which present an almost insurmountable obstacle to the traditional perturbative renormalization procedure in four dimensions.

$$\int d^4x \sqrt{g} R_{\mu\nu\alpha\beta} R^{\alpha\beta\rho\sigma} R_{\rho\sigma\kappa\lambda} R^{\kappa\lambda\mu\nu} \ . \tag{1.108}$$

Again on-shell all other invariants can be shown to be proportional to this one. One can therefore attempt to summarize the situation so far as follows:

- In principle perturbation theory in $G$ in provides a clear, covariant framework in which radiative corrections to gravity can be computed in a systematic loop expansion. The effects of a possibly non-trivial gravitational measure do not show up at any order in the weak field expansion, and radiative corrections affecting the renormalization of the cosmological constant, proportional to $\delta^d(0)$, are set to zero in dimensional regularization.
- At the same time at every order in the loop expansion new invariant terms involving higher derivatives of the metric are generated, whose effects cannot simply be absorbed into a re-definition of the original couplings. As expected on the basis of power-counting arguments, the theory is not perturbatively renormalizable in the traditional sense in four dimensions (although it seems to fail this test by a small measure in lowest order perturbation theory).
- The standard approach based on a perturbative expansion of the pure Einstein theory in four dimensions is therefore not convergent (it is in fact badly divergent), and represents therefore a temporary dead end.

## 1.6 Gravity in *d* Dimensions

In view of some of the discussion that will appear later on, it will be useful to recall here some rather interesting features of gravity that appear in dimensions not equal to four. The tensor field equations for general relativity with $\lambda = 0$

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \ , \tag{1.109}$$

seem at first to make sense in any dimension. But on closer examination one notices that some rather special circumstances arise in four dimensions.

As an exercise one can start for example by counting the number of algebraically independent components of the Riemann curvature tensor $R_{\mu\rho\sigma}$ which is $d^2(d^2-1)/12$ in $d$ dimensions. For the Ricci tensor one has instead $\frac{1}{2}d(d+1)$ independent components in $d > 1$ (there is no notion of intrinsic curvature in $d = 1$), for the Einstein tensor $G_{\mu\nu}$ one has also $\frac{1}{2}d(d+1)$ independent components in $d > 2$ (the Einstein tensor vanishes identically in $d = 2$), whereas the Ricci scalar has clearly only one component in any dimension $d > 1$.

In $d = 1$ space is flat, and all tensors $R_{\mu\rho\sigma}$, $G_{\mu\nu}$, $R_{\mu\nu}$ as well as $R$ vanish identically. In $d > 2$ the Ricci and Einstein tensor are related to each other

$$\begin{aligned} G_{\mu\nu} &= R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} & \qquad G &= \tfrac{1}{2}(2-d) R \\ R_{\mu\nu} &= G_{\mu\nu} - \frac{1}{d-2} g_{\mu\nu} R & \qquad R &= \frac{2}{2-d} G \ . \end{aligned} \tag{1.110}$$

For $d = 2$ both the Riemann tensor and the Ricci tensor only have one component, which is related to the scalar curvature,

$$R_{\mu\nu\rho\sigma} = \tfrac{1}{2} R (g_{\mu\sigma} g_{\nu\rho} - g_{\mu\rho} g_{\nu\sigma}) \qquad R_{\mu\nu} = \tfrac{1}{2} R g_{\mu\nu} \ . \tag{1.111}$$

Because of the last identity, in two dimensions the Einstein tensor vanishes identically, $G_{\mu\nu} = 0$, and there is no Einstein gravity in $d = 2$.

Therefore the first interesting integer dimension is clearly $d = 3$. Now, in four dimensions the Riemann tensor has 20 algebraically independent components, while the Ricci and Einstein tensor have ten (as the metric). But in three dimensions the Riemann, Ricci and Einstein tensor all have the same number of algebraically independent components (6), and are related to each other by

$$R_{\mu\nu\rho\sigma} = g_{\mu\rho} G_{\nu\sigma} + g_{\nu\sigma} G_{\mu\rho} - g_{\mu\sigma} G_{\nu\rho} - g_{\nu\rho} G_{\mu\sigma} + G (g_{\mu\sigma} g_{\nu\rho} - g_{\mu\rho} g_{\nu\sigma}) \ . \tag{1.112}$$

The field equations $G_{\mu\nu} = 8\pi G T_{\mu\nu}$ and their trace $G = 8\pi G T$ then imply, using Eq. (1.112), that the Riemann tensor is completely determined by the matter distribution implicit in $T_{\mu\nu}$,

$$R_{\mu\nu\rho\sigma} = 8\pi G \left[ g_{\mu\rho} T_{\nu\sigma} + g_{\nu\sigma} T_{\mu\rho} + g_{\mu\sigma} T_{\nu\rho} - g_{\nu\rho} T_{\mu\sigma} + T \left( g_{\mu\sigma} g_{\nu\rho} - g_{\mu\rho} g_{\nu\sigma} \right) \right] \ . \tag{1.113}$$

Furthermore, in empty space $T_{\mu\nu} = 0$, which then implies the vanishing of Riemann there

$$R_{\mu\nu\rho\sigma} = 0 \tag{1.114}$$

As a result in three dimensions classical spacetime is locally flat everywhere outside a source, gravitational fields do not propagate outside matter, and two bodies cannot experience any gravitational force: they move uniformly on straight lines.

There cannot be any gravitational waves either: the Weyl tensor which carries information about gravitational fields not determined locally by matter, vanishes identically in three dimensions. One further surprising (and disappointing) conclusion from the previous arguments is that black holes cannot exist in three dimensions, as spacetime is flat outside matter, which always allows light emitted from a star to escape to infinity. One interesting case though is three-dimensional anti-DeSitter space, with a scaled cosmological constant $\lambda = -1/\xi^2$. There one can show that objects which could be described as black holes exist, with a black hole horizon in the non-rotating case at $r_0 = \xi\sqrt{MG}$, where $M$ is the mass of the collapsed object (Banados Teitelboim and Zanelli, 1992). Note that in three dimensions $G$ has dimensions of a length, so that the product $MG$ ends up being dimensionless. The scale of the horizon is therefore supplied by the scale $\xi$.

What seems rather puzzling at first is that Newtonian theory seems to make perfect sense in $d = 3$. The Newtonian potential is non-vanishing and grows logarithmically with distance,

$$V(r) \propto G \int d^2\mathbf{k}\, e^{i\mathbf{x}\cdot\mathbf{k}}/\mathbf{k}^2 \sim G \log r \ . \tag{1.115}$$

This can only mean that the Newtonian theory is *not* recovered in the weak field limit of the relativistic theory (Deser, Jackiw and Templeton, 1982; Giddings, Abbott and Kuchar, 1984).

To see this explicitly, it is sufficient to consider the trace-reversed form of the field equations,

$$R_{\mu\nu} = 8\pi G \left( T_{\mu\nu} - \frac{1}{d-2}\, g_{\mu\nu} T \right) \ , \tag{1.116}$$

with $T = T^{\lambda}_{\ \lambda}$, in the weak field limit. In the linearized theory, with $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$, and in the Hilbert-DeDonder gauge

$$\nabla_\lambda h^{\lambda}_{\ \mu} - \tfrac{1}{2}\nabla_\mu h^{\lambda}_{\ \lambda} = 0 \ , \tag{1.117}$$

one obtains the wave equation

$$\Box\, h_{\mu\nu} = -16\pi G \left( \tau_{\mu\nu} - \frac{1}{d-2}\, \eta_{\mu\nu}\, \tau \right) \ , \tag{1.118}$$

with $\tau_{\mu\nu}$ the linearized stress tensor. After neglecting the spatial components of $\tau_{\mu\nu}$ in comparison to the mass density $\tau_{00}$, and assuming that the fields are quasi-static, one obtains a Poisson equation for $h_{00}$,

$$\nabla^2 h_{00} = -16\pi G \, \frac{d-3}{d-2} \, \tau_{00} \ . \tag{1.119}$$

In four dimension this is equivalent to Poisson's equation for the Newtonian theory

$$\nabla^2 \phi = 4\pi G \rho \ , \tag{1.120}$$

when one identifies the metric with the Newtonian field $\phi$ in the usual way via $h_{00} = -2\phi$ (an identification which follows independently of the previous arguments, from the weak field limit of the geodesic equation). But in three dimensions such a correspondence is prevented by the fact that, due to the result in Eq. (1.119) for the non-relativistic Newtonian coupling appearing in Poisson's equation, Eq. (1.120),

$$G_{\text{Newton}} = \frac{d-3}{2\,(d-2)} \, G \ , \tag{1.121}$$

the mass density $\tau_{00}$ decouples from the gravitational field $h_{00}$. As a result, the linearized theory in three dimensions fails to reproduce the Newtonian theory.

One can show by an explicit construction in the linearized theory that gravitational waves are not possible in three dimensions. To this purpose consider the wave equation in the absence of sources,

$$\Box \, h_{\mu\nu} = 0 \ . \tag{1.122}$$

It implies the existence of plane wave solutions of the type $h_{\mu\nu}(x) = \varepsilon_{\mu\nu}(k)\, e^{\pm ik\cdot x}$ with $k^2 = 0$ and polarization vector $\varepsilon_{\mu\nu}(k)$. The transverse-traceless gauge conditions

$$\partial_\lambda \, h^\lambda{}_\nu = 0 \qquad h^\lambda{}_\lambda = 0 \qquad h_{\mu 0} = 0 \ , \tag{1.123}$$

then imply for the remaining wave equation

$$\Box \, h_{ij} = 0 \ , \tag{1.124}$$

the transversality $k_i \, \varepsilon^i{}_j = 0$ and trace $\varepsilon^i{}_i = 0$ conditions, with $i, j$=1,2. The only possible solution is then the trivial one $\varepsilon_{ij}(k) = 0$ and therefore $h_{ij}(x) = 0$. One concludes therefore that three-dimensional gravity cannot sustain gravitational waves. The explicit derivation also makes apparent the issue that the absence of gravitational waves and the lack of a non-trivial Newtonian limit are not connected to each other in the three-dimensional theory.

One can count explicitly the number of physical degrees of freedom in the $d$-dimensional theory. The metric $g_{\mu\nu}$ has $\frac{1}{2}d(d+1)$ independent components, the Bianchi identity and the coordinate conditions reduce this number to $\frac{1}{2}d(d+1) - d - d = \frac{1}{2}d(d-3)$, which gives indeed the correct number of physical degrees of freedom (two) corresponding to a massless spin two particle in $d = 4$, and no physical degrees of freedom in $d = 3$ (it also gives minus one degree of freedom in $d = 2$, which will be discussed later). One can compare this counting of degrees of freedom to electromagnetism, where one has $d$ degrees of freedom for the field $A_\mu(x)$

in $d$ dimensions. The Coulomb gauge conditions$A_0 = 0$ and $\nabla_i A^i = 0$ then reduce the number of physical degrees of freedom to $d-2$.

Investigations of quantum two dimensional gravity have uncovered the fact though that there can be surviving scalar degrees of freedom in the quantum theory, at least in two dimensions. The usual treatment of two-dimensional gravity (Polyakov, 1981a,b) starts from the metric in the conformal gauge $g_{\mu\nu}(x) = e^{\phi}(x)\tilde{g}_{\mu\nu}$, where $\tilde{g}_{\mu\nu}$ is a reference metric, usually taken to be the flat one. The conformal gauge fixing then implies a non-trivial Faddeev Popov determinant, which, when exponentiated, results in the Liouville action,

$$I[\phi] \; = \; \frac{13}{24\pi} \int d^2x \left[ \tfrac{1}{2}(\partial_\mu \phi)^2 + \mu^2 e^{\phi} \right] \; , \tag{1.125}$$

where the $\mu$-term amounts to a renormalization of the bare cosmological constant. In the language of the conformal gauge, where $\sqrt{g} = e^{\phi}$ and $R = e^{-\phi}\partial^2\phi$, the preceding action can be re-written in arbitrary coordinate as a nonlocal contribution. When $D$ scalar fields are minimally coupled to two-dimensional gravity and integrated out, the conformal anomaly contribution modifies the Liouville effective action to

$$I[\phi] \; = \; \frac{26-D}{48\pi} \int d^2x \left[ \tfrac{1}{2}(\partial_\mu \phi)^2 + \mu^2 e^{\phi} \right] \; , \tag{1.126}$$

which establishes a connection to bosonic string theories (and random surfaces), whose critical embedding dimension is known to be $D = 26$. One would therefore conclude from this example that gravity in the functional integral representation needs a careful treatment of the conformal degree of freedom, since in general its dynamics cannot be assumed to be trivial. Of particular interest is of course the three-dimensional (2+1) case, since there again one has no progagating spin-two degrees of freedom. One would expect in this case that the conformal mode, if it has any residual non-trivial local dynamics, should be described by the same universality class as an interacting scalar field, equivalent to the 3d Ising model.[5]

## 1.7 Higher Derivative Terms

In the previous section it was shown that quantum corrections to the Einstein theory generate in perturbation theory $R^2$-type terms in four dimensions. It seems therefore that, for the consistency of the perturbative renormalization group approach in four dimension, these terms would have to be included from the start, at the level of the bare microscopic action. Thus the main motivation for studying gravity with higher derivative terms is that it might cure the problem of ordinary quantum gravity, namely its perturbative non-renormalizabilty in four dimensions. This is indeed the

[5] In the $2+\varepsilon$ expansion for gravity one finds in $d=3$ for the universal correlation length exponent $\nu = 1/(1+3/5+\cdots) \simeq 0.63$, which is not inconsistent with the Ising model value $\nu \simeq 0.629$.

case, in fact one can prove that higher derivative gravity (to be defined below) is perturbatively renormalizable to all orders in four dimensions.

At the same time new issues arise, which will be detailed below. The first set of problems has to do with the fact that, quite generally higher derivative theories with terms of the type $\phi\,\partial^4\phi$ suffer from potential unitarity problems, which can lead to physically unacceptable negative probabilities. But since these are genuinely dynamical issues, it will be difficult to answer them satisfactorily in perturbation theory. In non-Abelian gauge theories one can use higher derivative terms, instead of the more traditional dimensional continuation, to regulate ultraviolet divergences (Slavnov, 1972), and higher derivative terms have been used successfully for some time in lattice regulated field theories (Symanzik, 1983a,b). In these approaches the coefficient of the higher derivative terms is taken to zero at the end. The second set of issues is connected with the fact that the theory is asymptotically free in the higher derivative couplings, implying an infrared growth which renders the perturbative estimates unreliable at low energies, in the regime of perhaps greatest physical interest. Note that higher derivative terms arise in string theory as well (Forger, Ovrut, Theisen and Waldram, 1996).

Let us first discuss the general formulation. In four dimensions possible terms quadratic in the curvature are

$$
\begin{aligned}
&\int d^4x\,\sqrt{g}\,R^2 \\
&\int d^4x\,\sqrt{g}\,R_{\mu\nu}R^{\mu\nu} \\
&\int d^4x\,\sqrt{g}\,R_{\mu\nu\lambda\sigma}R^{\mu\nu\lambda\sigma} \\
&\int d^4x\,\sqrt{g}\,C_{\mu\nu\lambda\sigma}C^{\mu\nu\lambda\sigma} \\
&\int d^4x\,\sqrt{g}\,\varepsilon^{\mu\nu\kappa\lambda}\,\varepsilon^{\rho\sigma\omega\tau}\,R_{\mu\nu\rho\sigma}R^{\kappa\lambda\omega\tau}=128\pi^2\,\chi \\
&\int d^4x\,\sqrt{g}\,\varepsilon^{\rho\sigma\kappa\lambda}\,R_{\mu\nu\rho\sigma}R^{\mu\nu}{}_{\kappa\lambda}=96\pi^2\,\tau\ ,
\end{aligned}
\tag{1.127}
$$

where $\chi$ is the Euler characteristic and $\tau$ the Hirzebruch signature. It will be shown below that these quantities are not all independent. The Weyl conformal tensor is defined in $d$ dimensions as

$$
\begin{aligned}
C_{\mu\nu\lambda\sigma} = {} & R_{\mu\nu\lambda\sigma}-\tfrac{2}{d-2}(g_{\mu[\lambda}\,R_{\sigma]\nu}-g_{\nu[\lambda}\,R_{\sigma]\mu}) \\
& +\tfrac{2}{(d-1)(d-2)}\,R\,g_{\mu[\lambda}g_{\sigma]\nu}\ ,
\end{aligned}
\tag{1.128}
$$

where square brackets denote antisymmetrization. It is called conformal because it can be shown to be invariant under conformal transformations of the metric, $g_{\mu\nu}(x)\to\Omega^2(x)\,g_{\mu\nu}(x)$. In four dimensions one has

$$
C_{\mu\nu\lambda\sigma}=R_{\mu\nu\lambda\sigma}-R_{\lambda[\mu}g_{\nu]\sigma}-R_{\sigma[\mu}g_{\nu]\lambda}+\tfrac{1}{3}R\,g_{\lambda[\mu}g_{\nu]\sigma}\ . \tag{1.129}
$$

The Weyl tensor can be regarded as the traceless part of the Riemann curvature tensor,

$$g^{\lambda\sigma}C_{\lambda\mu\sigma\nu} = g^{\mu\nu}g^{\lambda\sigma}C_{\mu\lambda\nu\sigma} = 0 \ , \tag{1.130}$$

and on-shell the Riemann tensor in fact coincides with the Weyl tensor. From the definition of the Weyl tensor one infers in four dimensions the following curvature-squared identity

$$R_{\mu\nu\lambda\sigma}R^{\mu\nu\lambda\sigma} = C_{\mu\nu\lambda\sigma}C^{\mu\nu\lambda\sigma} + 2R_{\mu\nu}R^{\mu\nu} - \tfrac{1}{3}R^2 \ . \tag{1.131}$$

Some of these results are specific to four dimensions. For example, in three dimensions the Weyl tensor vanishes identically and one has

$$R_{\mu\nu\lambda\sigma}R^{\mu\nu\lambda\sigma} - 4R_{\mu\nu}R^{\mu\nu} - 3R^2 \ = \ 0 \qquad C_{\mu\nu\lambda\sigma}C^{\mu\nu\lambda\sigma} = 0 \ . \tag{1.132}$$

In four dimensions the expression for the Euler characteristic can be written equivalently as

$$\chi = \frac{1}{32\pi^2}\int d^4x\sqrt{g}\left[R_{\mu\nu\lambda\sigma}R^{\mu\nu\lambda\sigma} - 4R_{\mu\nu}R^{\mu\nu} + R^2\right] \ . \tag{1.133}$$

The last result is the four-dimensional analogue of the two-dimensional Gauss-Bonnet formula

$$\chi = \frac{1}{2\pi}\int d^2x\sqrt{g}\,R \ , \tag{1.134}$$

where $\chi = 2(g-1)$ and $g$ is the genus of the surface (the number of handles). For a manifold of fixed topology one can therefore use in four dimensions

$$R_{\mu\nu\lambda\sigma}R^{\mu\nu\lambda\sigma} = 4R_{\mu\nu}R^{\mu\nu} - R^2 + \text{const.} \tag{1.135}$$

and

$$C_{\mu\nu\lambda\sigma}C^{\mu\nu\lambda\sigma} = 2\,(R_{\mu\nu}R^{\mu\nu} - \tfrac{1}{3}R^2) + \text{const.} \tag{1.136}$$

Thus only *two* curvature squared terms for the gravitational action are independent in four dimensions (Lanczos, 1938), which can be chosen, for example, to be $R^2$ and $R^2_{\mu\nu}$. Consequently the most general curvature squared action in four dimensions can be written as

$$I = \int d^4x\sqrt{g}\left[\lambda_0 + kR + aR_{\mu\nu}R^{\mu\nu} - \tfrac{1}{3}(b+a)R^2\right] \ , \tag{1.137}$$

with $k = 1/16\pi G$, and up to boundary terms. The case $b = 0$ corresponds, by virtue of Eq. (1.136), to the conformally invariant, pure Weyl-squared case. If $b < 0$ then around flat space one encounters a tachyon at tree level. It will also be of some interest later that in the Euclidean case (signature $++++$) the full gravitational action of Eq. (1.137) is positive for $a > 0$, $b < 0$ and $\lambda_0 > -3/4b(16\pi G)^2$.

Curvature squared actions for classical gravity were originally considered in (Weyl, 1918) and (Pauli, 1958). In the sixties it was argued that the higher derivative action of Eq. (1.137) should be power counting renormalizable (Utiyama and DeWitt, 1962). Later it was proven to be renormalizable to all orders in perturbation theory (Stelle, 1977). Some special cases of higher derivative theories have been shown to be classically equivalent to scalar-tensor theories (Whitt, 1984).

One way to investigate physical properties of higher derivative theories is again via the weak field expansion. In analyzing the particle content it is useful to introduce a set of spin projection operators (Arnowitt, Deser and Misner, 1960; van Nieuwenhuizen, 1973), quite analogous to what is used in describing transverse-traceless (TT) modes in classical gravity (Misner, Thorne and Wheeler, 1973). These projection operators then show explicitly the unique decomposition of the continuum gravitational action for linearized gravity into spin two (transverse-traceless) and spin zero (conformal mode) parts. The spin-two projection operator $P^{(2)}$ is defined in $k$-space as

$$\begin{aligned} P^{(2)}_{\mu\nu\alpha\beta} = & \frac{1}{3k^2}\left(k_\mu k_\nu \eta_{\alpha\beta} + k_\alpha k_\beta \eta_{\mu\nu}\right) \\ & - \frac{1}{2k^2}\left(k_\mu k_\alpha \eta_{\nu\beta} + k_\mu k_\beta \eta_{\nu\alpha} + k_\nu k_\alpha \eta_{\mu\beta} + k_\nu k_\beta \eta_{\mu\alpha}\right) \\ & + \frac{2}{3k^4} k_\mu k_\nu k_\alpha k_\beta + \frac{1}{2}\left(\eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha}\right) - \frac{1}{3}\eta_{\mu\nu}\eta_{\alpha\beta} \ , \end{aligned} \tag{1.138}$$

the spin-one projection operator $P^{(1)}$ as

$$\begin{aligned} P^{(1)}_{\mu\nu\alpha\beta} = & \frac{1}{2k^2}\left(k_\mu k_\alpha \eta_{\nu\beta} + k_\mu k_\beta \eta_{\nu\alpha} + k_\nu k_\alpha \eta_{\mu\beta} + k_\nu k_\beta \eta_{\mu\alpha}\right) \\ & - \frac{1}{k^4} k_\mu k_\nu k_\alpha k_\beta \ , \end{aligned} \tag{1.139}$$

and the spin-zero projection operator $P^{(0)}$ as

$$\begin{aligned} P^{(0)}_{\mu\nu\alpha\beta} = & -\frac{1}{3k^2}\left(k_\mu k_\nu \eta_{\alpha\beta} + k_\alpha k_\beta \eta_{\mu\nu}\right) \\ & + \frac{1}{3}\eta_{\mu\nu}\eta_{\alpha\beta} + \frac{1}{3k^4} k_\mu k_\nu k_\alpha k_\beta \ . \end{aligned} \tag{1.140}$$

It is easy to check that the sum of the three spin projection operators adds up to unity

$$P^{(2)}_{\mu\nu\alpha\beta} + P^{(1)}_{\mu\nu\alpha\beta} + P^{(0)}_{\mu\nu\alpha\beta} = \frac{1}{2}\left(\eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha}\right) \ . \tag{1.141}$$

These projection operators then allow a decomposition of the gravitational field $h_{\mu\nu}$ into three independent modes. The spin two or transverse-traceless part

$$h^{TT}_{\mu\nu} = P^{\alpha}_{\ \mu} P^{\beta}_{\ \nu} h_{\alpha\beta} - \tfrac{1}{3} P_{\mu\nu} P^{\alpha\beta} h_{\beta\alpha} \ , \tag{1.142}$$

the spin one or longitudinal part

$$h^{L}_{\mu\nu} = h_{\mu\nu} - P^{\alpha}_{\ \mu} P^{\beta}_{\ \nu} h_{\alpha\beta} \ , \tag{1.143}$$

and the spin zero or trace part

$$h^T_{\mu\nu} \;=\; \tfrac{1}{3} P_{\mu\nu} P^{\alpha\beta} h_{\alpha\beta} \;\; , \tag{1.144}$$

are such that their sum gives the original field $h$

$$h \;=\; h^{TT} + h^L + h^T \;\; , \tag{1.145}$$

with the quantity $P_{\mu\nu}$ defined as

$$P_{\mu\nu} \;=\; \eta_{\mu\nu} - \frac{1}{\partial^2}\, \partial_\mu \partial_\nu \;\; , \tag{1.146}$$

or, equivalently, in $k$-space $P_{\mu\nu} = \eta_{\mu\nu} - k_\mu k_\nu / k^2$.

One can learn a number of useful aspects of the theory by looking at the linearized form of the equations of motion. As before, the linearized form of the action is obtained by setting $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ and expanding in $h$. Besides the expressions given in Eq. (1.45), one needs

$$\begin{aligned} \sqrt{g}\, R^2 &= (\partial^2 h^\lambda{}_\lambda - \partial_\lambda \partial_\kappa h^{\lambda\kappa})^2 + O(h^3) \\ \sqrt{g}\, R_{\lambda\mu\nu\kappa} R^{\lambda\mu\nu\kappa} &= \tfrac{1}{4} (\partial_\mu \partial_\kappa h_{\nu\lambda} + \partial_\lambda \partial_\nu h_{\mu\kappa} - \partial_\lambda \partial_\kappa h_{\mu\nu} - \partial_\mu \partial_\nu h_{\kappa\lambda})^2 + O(h^3) \;\; , \end{aligned} \tag{1.147}$$

from which one can then obtain, for example from Eq. (1.135), an expression for $\sqrt{g}\,(R_{\mu\nu})^2$,

$$\begin{aligned} \sqrt{g}\, R_{\alpha\beta} R^{\alpha\beta} \;=\; & \tfrac{1}{4}(\partial^2 h^\mu{}_\mu \partial^2 h^\alpha{}_\alpha + \partial^2 h_{\mu\alpha} \partial^2 h^{\mu\alpha} - 2\partial^2 h^\mu{}_\mu \partial_\alpha \partial_\beta h^{\alpha\beta} \\ & -2\partial_\alpha \partial_\nu h^\nu{}_\mu \partial_\alpha \partial_\beta h^{\mu\beta} + 2\partial_\mu \partial_\nu h^{\mu\nu} \partial_\alpha \partial_\beta h^{\alpha\beta}) + O(h^3) \;\; . \end{aligned} \tag{1.148}$$

Using the three spin projection operators defined previously, the action for linearized gravity without a cosmological constant term, Eq. (1.7), can then be re-expressed as

$$I_{\rm lin} \;=\; \tfrac{1}{4}\, k \int dx\, h^{\mu\nu} \, [P^{(2)} - 2P^{(0)}]_{\mu\nu\alpha\beta}\, \partial^2 h^{\alpha\beta} \;\; . \tag{1.149}$$

Only the $P^{(2)}$ and $P^{(0)}$ projection operators for the spin-two and spin-zero modes, respectively, appear in the action for the linearized gravitational field; the spin-one gauge mode does not enter the linearized action. Note also that the spin-zero mode enters with the wrong sign (in the linearized action it appears as a ghost contribution), but to this order it can be removed by a suitable choice of gauge in which the trace mode is made to vanish, as can be seen, for example, from Eq. (1.13).

It is often stated that higher derivative theories suffer from unitarity problems. This is seen as follows. When the higher derivative terms are included, the corresponding linearized expression for the gravitational action becomes

$$I_{\text{lin}} = \frac{1}{2} \int dx \; \{ \; h^{\mu\nu} \, [\tfrac{1}{2} k + \tfrac{1}{2} a \, (-\partial^2)] (-\partial^2) \, P^{(2)}_{\mu\nu\rho\sigma} \, h^{\rho\sigma}$$
$$+ \, h^{\mu\nu} \, [-k - 2b \, (-\partial^2)] (-\partial^2) \, P^{(0)}_{\mu\nu\rho\sigma} \, h^{\rho\sigma} \, \} \; . \tag{1.150}$$

Then the potential problems with unitarity and ghosts at ultrahigh energies, say comparable to the Planck mass $q \sim 1/G$, can be seen by examining the graviton propagator (Salam and Strathdee, 1978). In momentum space the free graviton propagator for higher derivative gravity and $\lambda_0 = 0$ can be written as

$$k < h_{\mu\nu}(q) \, h_{\rho\sigma}(-q) > = \frac{2P^{(2)}_{\mu\nu\rho\sigma}}{q^2 + \frac{a}{k} q^4} + \frac{P^{(0)}_{\mu\nu\rho\sigma}}{-q^2 - \frac{2b}{k} q^4} + \text{gauge terms} \; . \tag{1.151}$$

The first two terms on the r.h.s. can be decomposed as

$$2P^{(2)}_{\mu\nu\rho\sigma} \left[ \frac{1}{q^2} - \frac{1}{q^2 + \frac{k}{a}} \right] + P^{(0)}_{\mu\nu\rho\sigma} \left[ -\frac{1}{q^2} + \frac{1}{q^2 + \frac{k}{2b}} \right] \; . \tag{1.152}$$

One can see that, on the one hand, the higher derivative terms improve the ultraviolet behavior of the theory, since the propagator now falls of as $1/q^4$ for large $q^2$. At the same time, the theory appears to contain a spin-two ghost of mass $m_2 = \mu/\sqrt{a}$ and a spin-zero particle of mass $m_0 = \mu/\sqrt{2b}$. Here we have set $\mu = 1/\sqrt{(16\pi G)}$, which is of the order of the Planck mass $(1/\sqrt{G/\hbar c} = 1.2209 \times 10^{19} GeV/c^2)$. For $b < 0$ one finds a tachyon pole, which seems, for the time being, to justify the original choice of $b > 0$ in Eq. (1.137).

Higher derivative gravity theories also lead to modifications to the standard Newtonian potential, even though such deviations only become visible at very short distances, comparable to the Planck length $l_P = \sqrt{\hbar G/c^3} = 1.61624 \times 10^{-33} cm$. In some special cases they can be shown to be classically equivalent to scalar-tensor theories without higher derivative terms (Whitt, 1984). The presence of massive states in the tree level graviton propagator indicates short distance deviations from the static Newtonian potential of the form

$$h_{00} \sim \frac{1}{r} - \frac{4}{3} \frac{e^{-m_2 r}}{r} + \frac{1}{3} \frac{e^{-m_0 r}}{r} \; . \tag{1.153}$$

Moreover in the extreme case corresponding to the absence of the Einstein term $(k = 0)$ the potential is linear in $r$; but in this limit the theory is strongly infrared divergent, and it is not at all clear whether weak coupling perturbation theory is of any relevance.

In the quantum theory perturbation theory is usually performed around flat space, which requires $\lambda_0 = 0$, or around some fixed classical background. One sets again $g_{\mu\nu} \rightarrow \bar{g}_{\mu\nu} = g_{\mu\nu} + h_{\mu\nu}$ and expands the higher derivative action in powers of $h_{\mu\nu}$. If $\lambda_0$ is nonzero, one has to expand around a solution of the classical equations of motion for higher derivative gravity with a $\lambda$-term (Barth and Christensen, 1983), and the solution will no longer be constant over space-time. The above expansion is consistent with the assumption that the two higher derivative couplings $a$ and $b$ are large, since in such a limit one is close to flat space. One-loop radiative corrections then show that the theory is asymptotically free in the higher derivative couplings $a$ and $b$ (Julve and Tonin, 1978; Fradkin and Tseytlin, 1981; Avramidy and Barvinsky, 1985).

The calculation of one-loop quantum fluctuation effects proceeds in a way that is similar to the pure Einstein gravity case. One first decomposes the metric field as a classical background part $g_{\mu\nu}(x)$ and a quantum fluctuation part $h_{\mu\nu}(x)$ as in Eq. (1.92), and then expands the classical action to quadratic order in $h_{\mu\nu}$, with gauge fixing and ghost contributions added, similar to those in Eqs. (1.95) and (1.96), respectively. The first order variation of the action of Eq. (1.137) gives the field equations for higher derivative gravity in the absence of sources,

$$\begin{aligned}\frac{\partial I}{\partial g^{\mu\nu}} &= \frac{1}{\kappa^2}\sqrt{g}\,(R^{\mu\nu}-\tfrac{1}{2}g^{\mu\nu}R)+\tfrac{1}{2}\lambda_0\sqrt{g}\,g^{\mu\nu}\\ &\quad +a\sqrt{g}\,[\tfrac{2}{3}(1+\omega)R\,(R^{\mu\nu}-\tfrac{1}{4}g^{\mu\nu}R)\\ &\quad +\tfrac{1}{2}g^{\mu\nu}R_{\alpha\beta}R^{\alpha\beta}-2R^{\mu\alpha\nu\beta}R_{\alpha\beta}+\tfrac{1}{3}(1-2\omega)\nabla^\mu\nabla^\nu R\\ &\quad -\Box R^{\mu\nu}+\tfrac{1}{6}(1+4\omega)\,g^{\mu\nu}\,\Box R]=0\ ,\end{aligned} \tag{1.154}$$

where we have set for the ratio of the two higher derivative couplings $\omega = b/a$.

The second order variation is done similarly. It then allows the Gaussian integral over the quantum fields to be performed using the formula of Eq. (1.97). One then finds that the one-loop effective action, which depends on $g_{\mu\nu}$ only, can be expressed as

$$\Gamma = \tfrac{1}{2}\,\mathrm{tr}\ln F_{mn} - \mathrm{tr}\ln Q_{\alpha\beta} - \tfrac{1}{2}\,\mathrm{tr}\ln c^{\alpha\beta}\ , \tag{1.155}$$

with the quantities $F_{nm}$ and $Q_{\alpha\beta}$ defined by

$$\begin{aligned}F_{nm} &= \frac{\delta^2 I}{\delta g^m\,\delta g^n}+\frac{\delta\chi_\alpha}{\delta g^m}\,c^{\alpha\beta}\,\frac{\delta\chi_\beta}{\delta g^n}\\ Q_{\alpha\beta} &= \frac{\delta\chi_\alpha}{\delta g^m}\,\nabla^m_\beta\ .\end{aligned} \tag{1.156}$$

A shorthand notation is used here, where spacetime and internal indices are grouped together so that $g^m = g_{\mu\nu}(x)$. $\chi_\alpha$ are a set of gauge conditions, $c^{\alpha\beta}$ is a nonsingular functional matrix fixing the gauge, and the $\nabla^i_\alpha$ are the local generators of the group of general coordinate transformations, $\partial^i_\alpha f^\alpha = 2g_{\alpha(\mu}\nabla_{\nu)}f^\alpha(x)$.

Ultimately one is only interested in the divergent part of the effective one-loop action. The method of extracting the divergent part out of the determinant (or trace) expression in Eq. (1.155) is similar to what is done, for example, in QED to evaluate the contribution of the fermion vacuum polarization loop to the effective action. There, after integrating out the fermions, one obtains a functional determinant of the massless Dirac operator $\not{D}(A)$ in an external $A_\mu$ field,

$$\mathrm{tr}\ln \not{D}(A) - \mathrm{tr}\ln \not{\partial} = \frac{c}{\varepsilon} \int d^4x A^\mu \left(\eta_{\mu\nu} \partial^2 - \partial_\mu \partial_\nu\right) A^\nu + \dots \tag{1.157}$$

with $c$ a calculable numerical constant. The trace needs to be regulated, and one way of doing it is via the integral representation

$$\tfrac{1}{2}\,\mathrm{tr}\ln \not{D}^2(A) = -\tfrac{1}{2} \int_\eta^\infty \frac{dt}{t}\, \mathrm{tr}\exp\left[-t\, \not{D}^2(A)\right] \ , \tag{1.158}$$

with $\eta$ a cutoff that is sent to zero at the end of the calculation. For gauge theories a more detailed discussion can be found for example in (Rothe, 2005), and references therein. In the gravity case further discussions and more results can be found in (DeWitt, 1967; 't Hooft and Veltman, 1974; Gilkey, 1975; Christensen and Duff, 1978) and references therein.

In the end, by a calculation similar to the one done in the pure Einstein gravity case, one finds that the one-loop contribution to the effective action contains for $d \to 4$ a divergent term of the form

$$\Delta \mathscr{L} = \frac{\sqrt{g}}{16\pi^2(4-d)} \left\{ \beta_2 \left(R_{\mu\nu}^2 - \tfrac{1}{3} R^2\right) + \beta_3 R^2 + \beta_4 R + \beta_5 \right\} \ , \tag{1.159}$$

with the coefficients for the divergent parts given by

$$\begin{aligned}
\beta_2 &= \frac{133}{10} \\
\beta_3 &= \frac{10}{9}\,\omega^2 + \frac{5}{3}\,\omega + \frac{5}{36} \\
\beta_4 &= \frac{1}{a\kappa^2} \left( \frac{10}{3}\,\omega - \frac{13}{6} - \frac{1}{4\omega} \right) \\
\beta_5 &= \frac{1}{a^2\kappa^4} \left( \frac{5}{2} + \frac{1}{8\omega^2} \right) + \frac{\tilde{\lambda}}{a\kappa^4} \left( \frac{56}{3} + \frac{2}{3\omega} \right) \ .
\end{aligned} \tag{1.160}$$

Here $\omega \equiv b/a$ and $\tilde{\lambda}$ is the dimensionless combination of the cosmological and Newton's constant $\tilde{\lambda} \equiv \frac{1}{2}\lambda_0 \kappa^4$ with $\kappa^2 = 16\pi G$. A divergence proportional to the topological invariant $\chi$ with coefficient $\beta_1$ has not been included, as it only adds a field-independent constant to the action for a manifold of fixed topology. Also $\delta^{(4)}(0)$-type divergences possibly originating from a non-trivial functional measure over the $g_{\mu\nu}$'s have been set to zero.

The structure of the ultraviolet divergences (which for an explicit momentum cutoff $\Lambda$ would have appeared as $1/\varepsilon \leftrightarrow \ln\Lambda$) allow one to read off immediately the renormalization group $\beta$-functions for the various couplings. To this order, the renormalization group equations for the two higher derivative couplings $a$ and $b$ and the dimensionless ratio of cosmological and Newton's constant $\tilde{\lambda}$ are

$$\frac{\partial a}{\partial t} = \beta_2 + \dots$$

$$
\begin{aligned}
\frac{\partial \omega}{\partial t} &= -\frac{1}{a}\,(3\beta_3 + \omega\beta_2) + \dots \\
\frac{\partial \tilde{\lambda}}{\partial t} &= \tfrac{1}{2}\kappa^4\beta_5 + 2\tilde{\lambda}\,\kappa^2\beta_4 + \dots
\end{aligned}
\tag{1.161}
$$

with the dots indicating higher loop corrections. Here $t$ is the logarithm of the relevant energy scale, $t = (4\pi)^{-2}\ln(\mu/\mu_0)$, with $\mu$ a momentum scale $q^2 \approx \mu^2$, and $\mu_0$ some fixed reference scale. It is argued furthermore by the quoted authors that only the quantities $\beta_2$, $\beta_3$ and the combination $\kappa^4\beta_5 + 4\tilde{\lambda}\,\kappa^2\beta_4$ are gauge independent, the latter combination appearing in the renormalization group equation for $\tilde{\lambda}(t)$ (this is a point to which we shall return later, as it follows quite generally from the properties of the gravitational action, and therefore from the gravitational functional integral, under a field rescaling, see Sect. 3.5).

The perturbative scale dependence of the couplings $a(\mu)$, $b(\mu)$ and $\tilde{\lambda}(\mu)$ follows from integrating the three differential equations in Eq. (1.161). The first renormalization group equation is easily integrated, and shows the existence of an ultraviolet fixed point at $a^{-1} = 0$; the one-loop result for the running coupling $a$ is simply given by $a(t) = a(0) + \beta_2\, t$, or

$$
a^{-1}(\mu) \underset{\mu\to\infty}{\sim} \frac{16\pi^2}{\beta_2 \ln(\mu/\mu_0)} \ , \tag{1.162}
$$

with $\mu_0$ a reference scale. It suggests that the effective higher derivative coupling $a(\mu)$ increases at short distances, but decreases in the infrared regime $\mu \to 0$. But one should keep in mind that the one loop results are reliable at best only at very short distances, or large energy scales, $t \to \infty$. At the same time these results seem physically reasonable, as one would expect curvature squared terms to play less of a role at larger distances, as in the classical theory.

The scale dependence of the other couplings is a bit more complicated. The equation for $\omega(t)$ exhibits two fixed points at $\omega_{uv} \approx -0.0229$ and $\omega_{ir} \approx -5.4671$; in either case this would correspond to a higher derivative action with a positive $R^2$ term. It would also give rise to rapid short distance oscillations in the static potential, as can be seen for example from Eq. (1.153) and the definition of $m_0 = \mu/\sqrt{2b}$. The equation for $\tilde{\lambda}(t)$ gives a solution to one-loop order $\tilde{\lambda}(t) \sim \text{const.}\, t^q$ with $q \approx 0.91$, suggesting that the effective gravitational constant, in units of the cosmologial constant, decreases at large distances. The experimental value for Newton's constant $\hbar G/c^3 = (1.61624 \times 10^{-33} cm)^2$ and for the scaled cosmological constant $G\lambda_0 \sim 1/(10^{28} cm)^2$ is such that the observed dimensionless ratio between the two is very small, $G^2\lambda_0 \sim 10^{-120}$. In the present model is seems entirely unclear how such a small ratio could arise from perturbation theory alone.

At short distances the dimensionless coupling $\tilde{\lambda} \sim \lambda_0 G^2$ seems to increases rapidly, thus partially invalidating the conclusions of a weak field expansion around flat space, which are based generally on the assumption of small $G$ and $\lambda_0$. At the same time, the fact that the higher derivative coupling $a$ grows more rapidly in the ultraviolet than the coupling $\tilde{\lambda}$ can be used retroactively at least as a partial justification for the flat space expansion, in which the cosmological and Einstein terms are treated perturbatively. Ultimately the resolution of such delicate and complex issues would presumably require the development of the perturbative expansion not around flat space, but more appropriately around the de Sitter metric, for which $R = 2\lambda_0/\kappa^2$. Even then one would have to confront such genuinely non-perturbative issues, such as what happens to the spin-zero ghost mass, whether the ghost poles gets shifted away from the real axis by quantum effects, and what the true ground state of the theory looks like in the long distance, strong fluctuation regime not accessible by perturbation theory.

What is also a bit surprising is that higher derivative gravity, to one-loop order, does not exhibit a nontrivial ultraviolet point in $G$, even though such a fixed point is clearly present in the $2+\varepsilon$ expansion (to be discussed later) at the one- and two-loop order, as well as in the lattice regularized theory in four dimensions (also to be discussed later). But this could just reflect a limitation of the one-loop calculation; to properly estimate the uncertainties of the perturbative results in higher derivative gravity and their potential physical implications a two-loop calculation is needed, which hopefully will be performed in the near future.

To summarize, higher derivative gravity theories based on $R^2$-type terms are perturbatively renormalizable, but exhibit some short-distance oddities in the tree-level spectrum, associated with either ghosts or tachyons. Their perturbative (weak field) treatment suggest that the higher derivative couplings are only relevant at short distances, comparable to the Planck length, but the general evolution of the couplings away from a regime where perturbation theory is reliable remains an open question, which perhaps will never be answered satisfactorily in perturbation theory, if non-Abelian gauge theories, which are also asymptotically free, are taken as a guide.

## 1.8 Supersymmetry

An alternative approach to the vexing problem of ultraviolet divergences in perturbative quantum gravity (and for that matter, in any field theory) is to build in some additional degree of symmetry between bosons and fermions, such that loop effects acquire reduced divergence properties, or even become finite. One such proposal, based on the invariance under local supersymmetry transformation, adds to the Einstein gravity Lagrangian a spin-$3/2$ gravitino field, whose purpose is to exactly cancel the loop divergences in the ordinary gravitational contribution. This last result comes from the well known fact that fermion loops in quantum field theory carry an additional factor of minus one, thus potentially reducing, or even canceling out entirely, a whole class of divergent diagrams. The issue then is to specify the nature of such a supersymmetry transformation, and from it deduce an extension of pure gravity which includes such a symmetry in an exact way. Since ordinary gravity has a local gauge invariance under the diffeomorphism group, one would expect its supersymmetric extension to have some sort of local supersymmetry.

The first step towards defining a theory of supergravity is therefore to introduce the concept of global supersymmetry. Quantum field theory in flat space is invariant under the Poincaré group, whose generators include $P_\mu$ for space translations, and $\Sigma_{\mu\nu} = -\Sigma_{\nu\mu}$ for Lorentz transformations. Their algebra

$$\begin{aligned}
[P^\mu, P^\nu] &= 0 \\
[\Sigma^{\mu\nu}, P^\lambda] &= P^\mu \eta^{\nu\lambda} - P^\nu \eta^{\mu\lambda} \\
[\Sigma^{\mu\nu}, \Sigma^{\lambda\sigma}] &= \Sigma^{\mu\sigma} \eta^{\nu\lambda} - \Sigma^{\nu\sigma} \eta^{\mu\lambda} - \Sigma^{\mu\lambda} \eta^{\nu\sigma} + \Sigma^{\nu\lambda} \eta^{\mu\sigma},
\end{aligned} \tag{1.163}$$

with infinitesimal group element

$$U(\omega, \varepsilon) = 1 + \tfrac{1}{2} \Sigma_{\mu\nu} \omega^{\mu\nu} + i P_\mu \varepsilon^\mu \;, \tag{1.164}$$

and infinitesimal spacetime transformation

$$x^\mu \to x^\mu + \omega^{\mu\nu} x_\nu + \varepsilon^\mu \;, \tag{1.165}$$

has therefore the general structure

$$[P,P] = 0 \qquad [P,\Sigma] \simeq P \qquad [\Sigma,\Sigma] \simeq \Sigma \;. \tag{1.166}$$

The first relationship implies that translations commute with each other, the second one that translations transform under the Lorentz group as four-vectors, and the third one that the Lorentz generators transform under the Lorentz group as antisymmetric tensors.

Supersymmetry generalizes the Poincaré group by adding Grassman-valued (fermionic) generators $Q_\alpha$, whose most important property is to transform bosons into fermions

$$\begin{aligned}
Q\,|\text{boson}\rangle &= |\text{fermion}\rangle \\
Q\,|\text{fermion}\rangle &= |\text{boson}\rangle \;.
\end{aligned} \tag{1.167}$$

The new fermionic operators are such that their anti-commutator is an operator proportional to the Hamiltonian, so that it automatically commutes with it; but actually in a Lorentz-invariant theory the anticommutator should be proportional to the total energy-momentum $P^\mu$. Consequently the new fermion operators need to satisfy a set of mixed commutation and anti-commutation relations of the type

$$\begin{aligned}
[P^\mu, Q_\alpha] &= [P^\mu, \bar{Q}_\alpha] = 0 \\
\{Q_\alpha, Q_\beta\} &= \{\bar{Q}_\alpha, \bar{Q}_\beta\} = 0 \\
\{Q_\alpha, \bar{Q}_\beta\} &= 2\, \gamma^\mu_{\alpha\beta} P_\mu \\
[\Sigma_{\mu\nu}, Q_\alpha] &= \frac{i}{2} \left(\sigma_{\mu\nu}\right)_{\alpha\beta} Q_\beta \;,
\end{aligned} \tag{1.168}$$

with the Dirac spin matrix $\sigma_{\mu\nu} \equiv \frac{1}{2i}[\gamma_\mu, \gamma_\nu]$; for a more complete discussion see for example (Fayet and Ferrara, 1977; Ferrara, 1984). The superalgebra of the $P$'s, $Q$'s and $\Sigma$'s has therefore the general structure

$$\{Q,Q\} = 0 \qquad [P,Q] = 0 \qquad [Q,\Sigma] \simeq Q \qquad \{Q,\bar{Q}\} \simeq P \ . \tag{1.169}$$

The first relationship implies that $Q$ is Grassmann-valued, the second that $Q$ commutes with spacetime translations (including therefore the generator of time translations $H$), the third one that $Q$ transforms under Lorentz transformations as a two-component Weyl spinor, and finally the last one that two supersymmetry transformations together give a spacetime translation. Physically, the first and second identities imply that there are pairs of fermion-boson states which are degenerate, while the last equality implies that the supersymmetry charge $Q$ can in some sense be considered as the "square root" of the Hamiltonian operator $P^0 \equiv H$. The fact that the supersymmetry generator $Q$ is tied with the translation generator $P$ causes some problems when trying to implement supersymmetry on a lattice, since the latter is generally only translationally invariant by an integer multiple of the lattice spacing (although one can find ways around it, as shown in Curci and Veneziano, 1987).

One of the remarkable properties of supersymmetry is that it predicts that every bosonic state be paired with a fermionic state of the same energy, and vice versa. Furthermore the supersymmetric vacuum has zero energy, since zero momentum implies $P^i \, |0\rangle >= 0$, while supersymmetry gives

$$Q_\alpha \, |0\rangle > = \bar{Q}_\beta \, |0\rangle > = 0 \ , \tag{1.170}$$

and therefore from Eq. (1.168)

$$\langle 0 | H | 0 \rangle > = 0 \ . \tag{1.171}$$

The last result is particularly interesting in the case of gravity, since it would tend to generally imply, for unbroken supersymmetry, a vanishing vacuum energy, and therefore a vanishing cosmological constant (until recently it was in fact believed that observationally the cosmological constant was consistent with zero, and therefore in good agreement with the predictions from supersymmetry; this has changed in the last decade since the distant supernovae surveys find that the cosmological constant is non-zero and positive). One more consequence of supersymmetry is that in a relativistic quantum field theory mass renormalization effects are expected to be identical for particles belonging to the same supersymmetric pair. A more general feature of supersymmentric theories is that they give rise to what are often referred to as non-renormalization theorems: if a particular type of bare coupling is omitted, it cannot be generated by radiative corrections. But since no supersymmetric partners of the standard model particles have been observed so far, supersymmetry must be rather far from an exact symmetry of the real world, at least at ordinary energies.

In the original formulation of supersymmetry there is only one fermionic generator which is a Majorana spinor. But it is in fact possible to have more than one supersymmetric charge, so that a more general form for the anti-commutation relation for the $Q$'s is of the type

$$\{Q^i_\alpha, \bar{Q}^j_\beta\} \; = \; 2\,\delta^{ij}\,\gamma^\mu_{\alpha\beta}\,P_\mu \;\; , \tag{1.172}$$

with $i, j = 1 \ldots \mathscr{N}$. The original case of Eq. (1.168) then corresponds to the simplest choice, $\mathscr{N} = 1$ supersymmetry, whereas $\mathscr{N} > 1$ is referred to as extended supersymmetry. So different supersymmetric theories can be labeled by the number $\mathscr{N}$ of supersymmetric charges, but it turns out that this number is highly constrained, as in any given spacetime dimensions only certain values of $\mathscr{N}$ are possible. As shown above, in four dimensions $\mathscr{N} = 1$ supersymmetry has a complex pair $Q, \bar{Q}$ of supersymmetry charges, which are each two-component Weyl spinors, thus giving a total of four real supercharges. On the other hand, still in four dimensions, $\mathscr{N} = 4$ supersymmetry has four complex pairs $Q, \bar{Q}$ of supersymmetry charges, again with each a two-component Weyl spinors, thus giving now a total of sixteen real supercharges. Accordingly the renormalization properties of supersymmetric field theories vary dramatically, depending on which type of supersysmmetry is actually being implemented. For example, for $\mathscr{N} = 2$ supersymmetry the vanishing of the $\beta$-function at leading order implies that it will vanish to all orders. For $\mathscr{N} = 4$ supersymmetry the situation is even more remarkable, since there one has $\beta(g) = 0$ to all orders in $g$ without any need to fine-tune the interaction. The latter provides an example of a theory with no ultraviolet divergences, and truly constant coupling constant. Ultimately whether any of these theories are just ingenious elaborate mathematical recreations, or appear instead as parts of physical theories realized in nature in some form or another remains so far still an open question (for a recent survey of phenomenological opportunities for supersymmetric theories see, for example Zumino and Gaillard, 2008). After all QED or QCD are *not* finite theories, and still lead to perfectly acceptable, non-trivial and experimentally verifiable predictions once the problem of ultraviolet divergences is treated correctly via the renormalization procedure. The danger in the case of supersymmetric theories is that after all the elaborate work done to construct such theories one might be left with an empty shell: a trivial theory and a complicated way of re-writing an essentially non-interacting, Gaussian theory.

Of great phenomenological interests are supersymmetric Yang-Mills theories in four dimensions. The simplest corresponds to an $SU(N_c)$ pure gauge theory with $\mathscr{N} = 1$ supersymmetry. The theory contains gauge bosons $A^a_\mu$ (the ordinary gluons, with $a = 1 \ldots N_c^2 - 1$) and a single 4-component Majorana spinor $\lambda^a$, the gluino, satisfying the Majorana condition $\bar{\lambda}^a = \lambda^{aT} C$. The gluino is the supersymmetric partner of the gluon, and, like the gluon itself, transforms under the adjoint representation of the group (thus in the case of $SU(3)$ both the gluon and the gluino are in a color octet representation). The susy-Yang-Mills Lagrangian is

$$\mathscr{L} \; = \; -\tfrac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} + \tfrac{1}{2}\,\bar{\lambda}\,\gamma^\mu\,D_\mu(A)\,\lambda \;\; , \tag{1.173}$$

with $F^a_{\mu\nu}$ the usual Yang-Mills field strength tensor, and $D_\mu(A)$ the usual gauge covariant derivative acting on $\lambda^a$. The action is locally invariant under supersymmetry transformations

$$\begin{aligned}
\delta A_\mu &= -2g\,\bar{\lambda}^a T^a \gamma_\mu\,\varepsilon \\
\delta\lambda^a &= -\frac{i}{g}\,\sigma^{\mu\nu} F^a_{\mu\nu}\,\varepsilon \\
\delta\bar{\lambda}^a &= \frac{i}{g}\,\bar{\varepsilon}\,\sigma^{\mu\nu} F^a_{\mu\nu} \ ,
\end{aligned} \tag{1.174}$$

with $\sigma^{\mu\nu} = \frac{1}{2i}[\gamma^\mu, \gamma^\nu]$, and $\varepsilon(x)$ an arbitrary Grassmann parameter with Majorana properties. The supersymmetry of Eq. (1.174) leaves the action locally invariant, and at the same time relates fermions to bosons. The corresponding Noether current is

$$S^\mu \ = \ -F^a_{\rho\sigma}\,\sigma^{\rho\sigma}\,\gamma^\mu\,\lambda^a \ , \tag{1.175}$$

and satisfies $\partial_\mu S^\mu = 0$ after using the field equations, as well as $\gamma^\mu S_\mu = 0$. Furthermore it is known that in this theory gluino condensation occurs non-perturbatively, giving rise to a non-vanishing condensate $\langle\bar{\lambda}\lambda\rangle \neq 0$.

## 1.9 Supergravity

When supersymmetry is promoted to a local invariance of the theory one obtains supergravity: supergravity can therefore rightfully be considered as *the* gauge theory of supersymmetry. In the simplest model one adds to the Einstein gravity Lagrangian a spin-$3/2$ gravitino field, whose purpose is to exactly cancel loop divergences from the Einstein contributions. The enhanced symmetry is built into the action so as to ensure that such a cancellation does not just occur at one loop order, but propagates to every order of the loop expansion. In these theories the gravitino therefore emerges as the natural fermionic partner of the graviton, with zero mass for unbroken supersymmetry. The intent of this section is more to provide the general flavor of such an approach, and illustrate supergravity theories by a few specific examples of suitable actions, without getting into elegant technical aspect such as superfields and superpropagators. The reader is then referred to the vast literature on the subject for further examples, as well as contemporary leading candidate theories.

As stated, in the simplest scenario, one adds to gravity a spin-$\frac{3}{2}$ fermion field with suitable symmetry properties. A generally covariant action describing the interaction of vierbein fields $e^a_\mu(x)$ (with the metric field given by $g_{\mu\nu} = e^a_\mu e_{a\nu}$) and Rarita-Schwinger spin-$\frac{3}{2}$ fields $\psi_\mu(x)$, subject to the Majorana constraints $\psi_\rho = C\bar{\psi}^T_\rho$, was originally given in (Ferrara, Freedman and van Nieuwenhuizen, 1976). In the second order formulation it contains three contributions

$$I = \int d^4x\,(\mathscr{L}_2 + \mathscr{L}_{3/2} + \mathscr{L}_4) \ , \tag{1.176}$$

with the usual Einstein term

$$\mathscr{L}_2 = \frac{1}{4\kappa^2}\sqrt{g}R \ , \tag{1.177}$$

the gravitino contribution

$$\mathscr{L}_{3/2} = -\tfrac{1}{2}\, \varepsilon^{\lambda\rho\mu\nu}\, \bar{\psi}_\lambda \gamma_5 \gamma_\mu D_\nu \psi_\rho \;\; , \tag{1.178}$$

and a quartic fermion self-interaction

$$\begin{aligned} \mathscr{L}_4 = -\, \tfrac{1}{8\kappa^2 \sqrt{g}} \;& (\varepsilon^{\tau\alpha\beta\nu} \varepsilon_\tau{}^{\gamma\delta\mu} + \varepsilon^{\tau\alpha\mu\nu} \varepsilon_\tau{}^{\gamma\delta\beta} - \varepsilon^{\tau\beta\mu\nu} \varepsilon_\tau{}^{\gamma\delta\alpha}) \\ \times \quad & (\bar{\psi}_\alpha \gamma_\mu \psi_\beta)(\bar{\psi}_\gamma \gamma_\nu \psi_\delta) \;\; , \end{aligned} \tag{1.179}$$

with the Rarita-Schwinger fields subject to the Majorana constraint $\psi_\mu = C \bar{\psi}_\mu(x)^T$. The covariant derivative defined as

$$D_\nu \psi_\rho = \partial_\nu \psi_\rho - \Gamma^\sigma_{\nu\rho}\, \psi_\sigma + \tfrac{1}{2}\, \omega_{\nu ab}\, \sigma^{ab}\, \psi_\rho \;\; , \tag{1.180}$$

involves the standard affine connection $\Gamma^\sigma_{\nu\rho}$, as well as the vierbein connection

$$\begin{aligned} \omega_{\nu\, ab} = \; & \tfrac{1}{2}\; [e_a{}^\mu (\partial_\nu e_{b\mu} - \partial_\mu e_{b\nu}) + e_a{}^\rho e_b{}^\sigma (\partial_\sigma e_{c\rho}) e^c{}_\nu] \\ & -\, (a \leftrightarrow b) \;\; , \end{aligned} \tag{1.181}$$

with Dirac spin matrices $\sigma_{ab} = \frac{1}{2i}[\gamma_a, \gamma_b]$. One can show that the combined Lagrangian is invariant, up to terms of order $(\psi)^5$, under the simultaneous transformations

$$\begin{aligned} \delta e^a{}_\mu &= i\kappa\, \bar{\varepsilon}\, \gamma^a\, \psi_\mu \\ \delta g_{\mu\nu} &= i\kappa\, \bar{\varepsilon}\, [\gamma_\mu\, \psi_\nu + \gamma_\nu\, \psi_\mu] \\ \delta \psi_\mu &= \kappa^{-1} D_\mu\, \varepsilon + \tfrac{1}{4}\, i\kappa\, (2\, \bar{\psi}_\mu \gamma_a \psi_b + \bar{\psi}_a \gamma_\mu \psi_b)\, \sigma^{ab}\, \varepsilon \;\; , \end{aligned} \tag{1.182}$$

where $\varepsilon(x)$ in an arbitrary Majorana spinor.

The action of Eq. (1.176) can be written equivalently in first order form (Deser and Zumino, 1976) as

$$I = \int d^4x \left( \tfrac{1}{4} \kappa^{-2}\, e R - \tfrac{1}{2}\, \varepsilon^{\lambda\rho\mu\nu}\, \bar{\psi}_\lambda\, \gamma_5\, \gamma_\mu\, D_\nu\, \psi_\rho \right) \;\; , \tag{1.183}$$

with $e_{a\mu}$ the vierbein with $e_{a\mu} e^a_\nu = g_{\mu\nu}$, and

$$e = \det e_{a\mu} \;, \qquad R = e_a{}^\mu e_b{}^\nu R_{\mu\nu}{}^{ab} \;\; . \tag{1.184}$$

The covariant derivative $D_\mu$ on $\psi_\nu$ is defined in terms of its spin-$\frac{1}{2}$ part only

$$D_\mu = \partial_\mu - \tfrac{1}{2}\, \omega_{\mu\, ab}\, \sigma^{ab} \;\; , \tag{1.185}$$

and is related to the curvature tensor via the commutator identity

$$[D_\mu, D_\nu] = -\tfrac{1}{2}\, R_{\mu\nu\, ab}\, \sigma^{ab} \;\; . \tag{1.186}$$

The first order action in Eq. (1.183) is invariant under

$$\begin{aligned}
\delta e^{a}{}_{\mu} &= i\kappa\,\bar{\varepsilon}(x)\,\gamma^{a}\,\psi_{\mu} \\
\delta\psi_{\mu} &= \kappa^{-1}\,D_{\mu}\,\varepsilon \\
\delta\omega_{\mu}{}^{ab} &= B_{\mu}{}^{ab} - \tfrac{1}{2}\,e_{\mu}{}^{b}\,B_{c}{}^{ab} + \tfrac{1}{2}\,e_{\mu}{}^{a}\,B_{c}{}^{bc} \quad ,
\end{aligned} \tag{1.187}$$

with the quantity $B$ defined as

$$B_{a}^{\lambda\mu} = i\,\varepsilon^{\lambda\mu\nu\rho}\,\bar{\varepsilon}\,\gamma_{5}\,\gamma_{a}\,D_{\nu}\,\psi_{\rho} \quad , \tag{1.188}$$

and $\varepsilon(x)$ in an arbitary local Majorana spinor. In the first order formulation the vierbeins $e_{a\mu}(x)$, the connections $\omega_{\mu}{}^{ab}(x)$ and the Majorana vector-spinors $\psi_{\mu}(x)$ are supposed to be varied independently.

The original motivation for the supergravity action of Eqs. (1.176) or (1.183) was that, just like ordinary source-free gravity is ultraviolet finite on-shell because of the identity relating the invariant $(R_{\mu\nu\rho\sigma})^2$ to $(R_{\mu\nu})^2$ and $R^2$, identities among invariants constructed out of $\psi_{\mu}$ and the strong constraints of supersymmetry would ensure one-loop, and higher, renormalizabilty of supergravity. There are reasons to believe that the triviality results found originally in globally supersymmetric theories (Nicolai, 1984) will not carry over into theories with local supersymmetry.

It was shown originally in (Grisaru, van Nieuwenhuizen and Vermaseren, 1976) and (Grisaru, 1977) that the original supergravity theory is finite to at least two loops, but most likely it fails to be finite at three loops. As a consequence, more complex theories were devised to avoid the three-loop catastrophe. A new formulation, $\mathcal{N}=4$ extended supergravity based on an $SO(4)$ symmetry, was suggested in (Das, 1977; Cremmer and Scherk, 1977; Nicolai and Townsend, 1981). This theory now contains vector, spinor and scalar particle in addition to the gravitino and the graviton. Specifically, the theory contains a vierbein field $e_{a\mu}$, four spin-$\frac{3}{2}$ Majorana fields $\psi_{\mu}^{i}$, four spin-$\frac{1}{2}$ Majorana fields $\xi^{i}$, six vector fields $A_{\mu}^{ij}$, a scalar field $A$ and a pseudoscalar field $B$, all massless, for a grand total of 53 independent terms in the Lagrangian. Subsequently $\mathcal{N}=8$ supergravity was proposed, based on the even larger group $SO(8)$ (Cremmer and Julia, 1978). The enlarged theory now contains one graviton, 8 gravitinos, 28 vector fields, 56 Majorana spin-$\frac{1}{2}$ fields and 70 scalar fields, all massless. In general, $SO(\mathcal{N})$ supergravity contains $\mathcal{N}$ gravitinos, $\frac{1}{2}\mathcal{N}(\mathcal{N}-1)$ gauge fields, as well as several spin-$\frac{1}{2}$ Majorana fermions and complex scalars. The $SO(\mathcal{N})$ symmetry here is one which rotates, for example, the $\mathcal{N}$ gravitinos into each other. In (Christensen, Duff, Gibbons and Roček, 1980) it was shown that in general such theories are finite at one loop order for $\mathcal{N}>4$. For $\mathcal{N}>8$ these theories become less viable since one then has more than one graviton, which leads to paradoxes, as well as particles with spin $j>2$.

It is beyond our scope here to go any more deeply in the issue of the origin of such intriguing ultraviolet cancellations (for a broad overview see for example Ferrara, 1984; Wess and Bagger, 1992). But, as perhaps the simplest and most elementary motivation, one can use the Nielsen-Hughes formula (Nielsen, 1981; Hughes, 1981) for the one-loop $\beta$-function contribution from a particle of spin $s$

$$\beta_0 = -(-1)^{2s}\,[(2s)^2 - \tfrac{1}{3}] \ , \tag{1.189}$$

to verify, by virtue of the particle multiplicities given above, that for example for $\mathcal{N} = 4$ the lowest order divergences cancel

$$\beta_0 = -\tfrac{47}{3}\cdot 1 + \tfrac{26}{3}\cdot 4 - \tfrac{11}{3}\cdot 6 + \tfrac{2}{3}\cdot 4 + \tfrac{1}{3}\cdot 1 = 0 \ . \tag{1.190}$$

For $\mathcal{N} = 8$ one has a similar complete cancellation

$$\beta_0 = -\tfrac{47}{3}\cdot 1 + \tfrac{26}{3}\cdot 8 - \tfrac{11}{3}\cdot 28 + \tfrac{2}{3}\cdot 56 + \tfrac{1}{3}\cdot 35 = 0 \ . \tag{1.191}$$

Still, the issue of perturbative ultraviolet finiteness of these theories remains largely an open question, in part due to the daunting complexity of higher loop calculations, even though one believes that the high level of symmetry should ensure the cancellation of ultraviolet divergences to a very high order (perhaps up to seven loops). Recently it was even suggested, based on the correspondence between $\mathcal{N} = 8$ supergravity and $\mathcal{N} = 4$ super Yang-Mills theory and the cancellations which arise at one and higher loops, that supergravity theories might be finite to all orders in the loop expansion (Bern et al, 2007).

One undoubtedly very attractive feature of supergravity theories is that they lead naturally to a small, or even vanishing, renormalized cosmological constant $\lambda_0$. Due to the high level of symmetry, quartic and quadratic divergences in this quantity are expected to cancel exactly between bosonic and fermionic contributions, leaving a finite or even zero result. The hope is that some of these desirable features will survive supersymmetry breaking, a mechanism eventually required in order to remove, or shift to a high mass, the so far unobserved supersymmetric partners of the standard model particles. Another possibility discussed separately later on is that supergravity theories, such as the $\mathcal{N} = 1$ one, might exhibit a non-trivial ultraviolet fixed point in $2+\varepsilon$, and perhaps even four dimension. This would cause a new phase to appear at sufficiently strong coupling, as in the non-linear sigma model.

## 1.10 String Theory

String theory postulates that the fundamental constituents of matter are not point particles, as in usual quantum field theory where fields $\phi(x)$ are defined at the point $x$, but tiny one-dimensional strings. As a result, the short distance behavior is expected to be changed drastically, as one is ultimately dealing with intrinsically extended objects. String theory's ultimate ambition goes in fact beyond the problem of regulating ultraviolet divergences in quantum field theory and quantum gravity: part of the general program is to provide a truly fundamental unified theory that contains all known interactions and particle species, including of course the graviton. One might argue that supergravity theories have had similar ambitions, but in string theory one would expect even better ultraviolet behavior due to the delocalized nature of the string, provided of course one understands how to formulate the theory in a consistent and calculable way.

The concept of a relativistic string (see Fig. 1.3) originated in the late sixties in the context of hadron physics, where it provided a very useful phenomenological description of certain peculiarities of strong interaction amplitudes. Later on three main motivations for studying string theories emerged: the search for a description of quark confinement in terms of gluon strings, a model of grand unification based on superstrings (Schwarz, 1982), and a description of the three-dimensional Ising model in terms of some sort of fermionic string (Polyakov, 1979).

One of the first concrete field-theoretic models of a string was given in (Nambu, 1969; Goto, 1971). The usual description introduces world-sheet coordinates $\sigma$ and $\tau$, defined on the two-dimensional surface swept out by the time evolution of the string. When the string is embedded in a $d$-dimensional space, points on the string world sheet are assigned coordinates $X^\mu(\sigma,\tau)$ with $\mu = 1 \dots d$. In analogy with the action for a relativistic point particle, which is proportional to the proper time, and therefore to the length of the world line, $I = \int_{\tau_i}^{\tau_f} d\tau$, the simplest action for such a string is the total area of the world sheet, $I = \int dA$. One can re-write this quantity by introducing an induced metric $g_{ab}$ on the worldsheet,

$$g_{ab}(\sigma,\tau) \; = \; \eta_{\mu\nu} \, \frac{\partial X^\mu}{\partial \sigma^a} \frac{\partial X^\nu}{\partial \sigma^b} \;, \tag{1.192}$$

with $\sigma^1 \equiv \tau$ and $\sigma^2 \equiv \sigma$, and $\eta_{\mu\nu}$ the flat metric in $d$ dimensions. Then the two-dimensional volume element is $dA = \sqrt{g}\, d^2\sigma$ with $g = -\det(g_{ab})$, and one has

$$I \; = \; \int_S d^2\sigma \, \sqrt{g} \;. \tag{1.193}$$

In a gravity language what one has so far is essentially a cosmological constant term. In terms of the variables $\dot{X} = \partial X / \partial \tau$ and $X' = \partial X / \partial \sigma$ one has

$$I[X] \; = \; \frac{1}{2\pi\alpha'} \int_S d^2\sigma \, \sqrt{(\dot{X} \cdot X')^2 - (\dot{X})^2 \, (X')^2} \;, \tag{1.194}$$

where a coupling constant $\alpha'$ has been introduced, having dimensions of an area, or of an inverse mass squared. The quantity $T_0 = 1/2\pi\alpha'$ is often referred to for obvious reasons as the string tension: a string of spatial size $R$ and time extent $T$ will have an energy per unit length of value $T_0$. In the following it will be convenient to re-absorb such a coupling into a re-definition of the $X$ variables.

One clear distinction that appears early on in this picture is between open (describing geometrically, in their time evolution, a sheet) and closed strings (described by a tube). Another important property of the string action is its invariance under reparametrizations of the world sheet coordinates, $X^\mu(\sigma,\tau) \to X^\mu[f(\sigma,\tau)]$. These can be considered as two dimensional diffeomorphism acting within the surface; they express an invariance of the area action under $\sigma^a$ coordinate redefinitions.

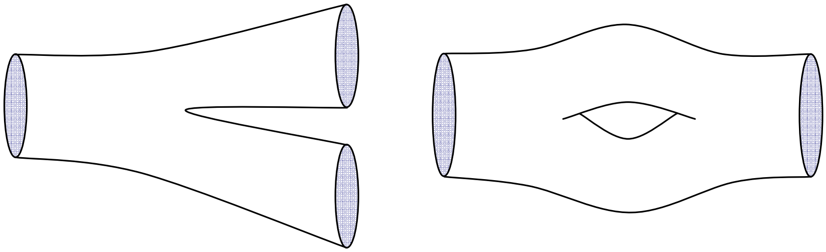


**Fig. 1.3** A string vertex (*left*) and a closed string loop (*right*).

The Nambu-Goto string is not easy to work with due to the square root. One can write down an action (Polyakov, 1981a,b) which is classically equivalent to it, but does not involve the square root of the $X$ variables, if one introduces the metric $g_{ab}$ as a Lagrange multiplier. The Euclidean action

$$I[g,X] = \tfrac{1}{2} \int d^2\sigma \sqrt{g}\, g^{ab}\, \partial_a X^\mu\, \partial_b X^\mu \ . \tag{1.195}$$

Variation of this action with respect to $X^\mu$ gives Laplace's equation for $X^\mu$, while the variation with respect to $g_{ab}$ gives Eq. (1.192). One noteworthy feature of the action in Eq. (1.195) is its large invariance group. It is invariant under world sheet $(a,b)$ diffeomorphisms, spacetime $(\mu,\nu)$ Lorentz invariance, and invariance under Weyl or conformal transformations

$$g_{ab}(\sigma,\tau) \to e^{2\,\omega(\sigma,\tau)}\, g_{ab}(\sigma,\tau) \ , \tag{1.196}$$

which implies that one is dealing with a two-dimensional conformal field theory. Note also that the string so far is embedded in flat space, but one could consider a more general embedding, with a suitable change in the spacetime metric $\eta_{\mu\nu} \to G_{\mu\nu}(X)$, and possibly additional terms in the action such as curvature $[{}^2R(X)]$ contributions.

Note that in the gauge $g_{ab} = \eta_{ab}$ the field $X$ satisfies the wave equation

$$\Box X^\mu \equiv \left( \frac{\partial^2}{\partial\sigma^2} - \frac{\partial^2}{\partial\tau^2} \right) X^\mu = 0 \ , \tag{1.197}$$

supplemented by the constraint equations $T_{ab} = 0$ with

$$T_{10} = T_{01} = \dot{X} \cdot X' = 0 \tag{1.198}$$

and

$$T_{00} = T_{11} = \tfrac{1}{2}\left( \dot{X}^2 + X'^2 \right) = 0 \ . \tag{1.199}$$

Solutions to the massless wave equation in Eq. (1.197) can be obtained in the usual way, by setting

$$X^\mu(\sigma,\tau) \;=\; X_R^\mu(\sigma^-) + X_L^\mu(\sigma^+) \;, \tag{1.200}$$

with $\sigma^\pm \equiv \tau \pm \sigma$. Then $X_R^\mu(\sigma^-)$ is the right-moving mode, while $X_L^\mu(\sigma^+)$ is the left-moving mode. The boundary conditions depend on whether one has a closed or open string. For closed strings the boundary condition is a periodicity in the $\sigma$ coordinate,

$$X^\mu(\sigma,\tau) \;=\; X^\mu(\sigma+\pi,\tau) \;. \tag{1.201}$$

For open strings one requires the vanishing of the normal derivative $X'^\mu = 0$ at $\sigma = 0, \pi$.

One can then expand the solutions to the wave equation of Eq. (1.197) in Fourier amplitudes $\alpha_n^\mu$

$$X_R^\mu \;=\; \tfrac{1}{2} x^\mu + \tfrac{1}{2} l^2 p^\mu (\tau-\sigma) + \tfrac{1}{2} i l \sum_{n\neq 0} \frac{1}{n} \alpha_n^\mu \, e^{-2in(\tau-\sigma)} \;, \tag{1.202}$$

and similarly for $X_L^\mu$ in terms of $\tilde{\alpha}_n^\mu$. Only an outline of the method will be provided here, the reader is referred for more detail for example to the recent monograph (Becker, Becker and Schwarz, 2007). Here $l$ is the fundamental string length $l = \sqrt{2\alpha'}$, and $x^\mu$ and $p^\mu$ are the center of mass coordinate and momentum of the string. The reality condition on $X^\mu$ implies

$$\alpha_{-n}^\mu \;=\; (\alpha_n^\mu)^\dagger \;, \tag{1.203}$$

and a similar requirement for the $\tilde{\alpha}_n^\mu$'s. To get the correct commutation relations for the $\alpha_n^\mu$'s and $\tilde{\alpha}_n^\mu$'s one needs the Poisson brackets for the $X^\mu$ variables. The only non-vanishing one is

$$\{\dot{X}^\mu(\sigma), X^\nu(\sigma')\} \;=\; T_0^{-1}\, \delta(\sigma-\sigma')\, \eta^{\mu\nu} \;, \tag{1.204}$$

which implies the commutation relations (via the usual replacement of the Poisson bracket with the commutator $\{\ldots\} \to -i[\ldots]$) for the $\alpha_n^\mu$'s

$$[\alpha_m^\mu, \alpha_n^\nu] \;=\; m\, \delta_{m+n}\, \eta^{\mu\nu} \;, \tag{1.205}$$

a similar expression for $\tilde{\alpha}_n^\mu$, and all other commutators equal to zero. Up to a factor of $\sqrt{m}$ difference in normalization, these are quite similar to the usual harmonic oscillator operators. But note that due to the appearance of the $\eta_{\mu\nu}$ on the r.h.s. the Hilbert space built up from the oscillator operators $\alpha_n^\mu$ is not positive definite.

The classical Hamiltonian for the two-dimensional closed string is given by

$$H \;=\; \tfrac{1}{2} T \int_0^\pi d\sigma \left( \dot{X}^2 + X'^2 \right) \;=\; \tfrac{1}{2} \sum_{n=-\infty}^{+\infty} \alpha_{-n}\, \alpha_n \;. \tag{1.206}$$

Similarly one can expand the constraint $T_{ab}$ in Fourier modes; it is more convenient to write these constraints as $\dot{X}_R^2 \equiv T_{--} = 0$ and $\dot{X}_L^2 \equiv T_{++} = 0$. One defines

$$L_m \equiv \int_0^\pi d\sigma\, T_{--} = \tfrac{1}{2} \sum_{n=-\infty}^{+\infty} \alpha_{m-n} \cdot \alpha_n \quad , \tag{1.207}$$

and similarly for $\tilde{L}_m$ in terms of $\dot{X}_L^2$ and therefore $\tilde{\alpha}_n^\mu$. A little algebra then gives the classical Poisson bracket

$$\{L_m, L_n\} = i\,(m-n)\,L_{m+n} \quad , \tag{1.208}$$

and an analogous expression for $\tilde{L}_m$. A simple interpretation for the occurrence of the above algebra in the closed string case is that it is obeyed by the generators $D_n$ of the infinitesimal "diffeomorphisms" on the unit circle $S^1$

$$D_n = i e^{in\theta} \frac{d}{d\theta} \quad . \tag{1.209}$$

In a quantum mechanical treatment for the operators $\alpha_n^\mu$ and $\tilde{\alpha}_n^\mu$ one has to be careful about ordering ambiguities, which were not taken into account when deriving the classical result of Eq. (1.208). These do not affect the above result unless $m+n=0$, in which case a new term can arise, the so-called central extension of the Virasoro algebra. In particular one needs to be careful to restrict the physical Hilbert space through the conditions

$$\begin{aligned} L_m\,|\phi\rangle &= 0 \quad (m>0) \\ (L_0 - a)\,|\phi\rangle &= 0 \quad , \end{aligned} \tag{1.210}$$

and the normalization requirement

$$\langle 0|\,[L_2, L_{-2}]\,|0\rangle = \frac{d}{2} \quad , \tag{1.211}$$

where $a$ is an arbitrary parameter (it will turn out to be $a=1$). Thus by a careful treatment of the operator ordering problem and a suitable physically motivated choice of the oscillator ground state one finds that the quantum-mechanical version of Eq. (1.208) is

$$[L_m, L_n] = (m-n)\,L_{m+n} + \tfrac{1}{12}\,d\,(m^3 - m)\,\delta_{m+n,0} \quad . \tag{1.212}$$

The origin of the central term proportional to $\delta_{m+n}$ is a requirement that the operator $L_0$ be normal ordered so as to obtain a finite matrix element.

Of great interest is of course the ground state of the bosonic string. From the form of the string Hamiltonian the mass $M$ of the closed string excitations (for $\alpha' = \frac{1}{2}$) is given by

$$M^2 \;=\; -8a + 8\sum_{n=1}^{\infty} \alpha_{-n}\cdot\alpha_n \;\;, \tag{1.213}$$

and for the open string

$$M^2 \;=\; -2a + 2\sum_{n=1}^{\infty} \alpha_{-n}\cdot\alpha_n \;\;, \tag{1.214}$$

with $a$ a constant to be determined from consistency conditions (absence of ghosts). Thus the mass squared for the for the ground state of closed strings is four times that for the open strings. In the first case on can show that the vector particle is massless, but the scalar ground state is a tachyon with $m^2 < 0$. Further analysis reveals though that the absence of ghosts (or negative norm states), which in the case of the bosonic string are associated with the timelike mode of $X^\mu(\sigma,\tau)$, implies that either $d = 26$ and $a = 1$, or $d \leq 25$ and $a \leq 1$. But the theory can be shown to be truly Lorentz invariant only in $d = 26$, which implies that for the bosonic string quantization and regularization are only consistent at $d = 26$.

Once the commutator algebra has been specified (as we have seen it is strongly restricted by the Lorentz and Weyl invariance of the theory) one can start enumerating the lowest excitations. The structure of the spectrum for closed strings is

$$\begin{aligned} |\,0\,\rangle &\rightarrow \text{tachyon} \\ \alpha_1^{\dagger\,\mu}\tilde{\alpha}_1^{\dagger\,\nu}\,|\,0\,\rangle &\rightarrow \text{massless tensor} \;\;, \end{aligned} \tag{1.215}$$

while for open strings one finds the following spectrum

$$\begin{aligned} |\,0\,\rangle &\rightarrow \text{tachyon} \\ \alpha_1^{\dagger\,\mu}\,|\,0\,\rangle &\rightarrow \text{massless vector} \\ \alpha_2^{\dagger\,\mu}\,|\,0\,\rangle &\rightarrow \text{massive vector} \\ \alpha_1^{\dagger\,\mu}\alpha_1^{\dagger\,\nu}\,|\,0\,\rangle &\rightarrow \text{massive tensor} \;\;, \end{aligned} \tag{1.216}$$

with the mass squared increasing linearly with spin (linear Regge trajectories).

Thus open bosonic string theory contains a massless spin two particle, described by a traceless symmetric tensor, whose low energy limit should be identified with the action for general relativity plus higher order corrections. In the original string theory framework this was regarded as a major disappointment, as no such particle appeared in the known hadron spectrum (Scherk and Schwartz, 1974).

There is one big problem that remains with the bosonic string discussed so far, namely that the ground state corresponds to a tachyon, a particle of mass $m^2 < 0$, which suggests some sort of fundamental instability of the theory. It is possible that the tachyon is just an artifact due to the expansion around the wrong string theory vacuum, but this remains a largely unsolved questions since it is not easy to treat the bosonic string non-perturbatively. But one possible approach to this problem is offered by the covariant Euclidean Feynman path integral.

It is possible at least in principle to write down a path integral for strings, in analogy to what one does for a free particle, and therefore provide a manifestly covariant quantization framework, without possible artifacts that might arise from the choice of conformal gauge. There one has for the Euclidean propagator

$$G(x,x') \;=\; \sum_{\text{paths x}\rightarrow \text{x}'} \exp[-m\,L(x,x')] \ , \tag{1.217}$$

with $L(x,x')$ the length of the path connecting $x$ and $x'$. In the case of strings one would write an object of the type

$$Z_{\text{string}}(C_i,C_f) \;=\; \int_{C_i}^{C_f} [dX]\,[dg]\,e^{-I[X,g]} \ , \tag{1.218}$$

where the sum is over all possible configurations of the string, including all allowed intervening topologies, and matching the configuration of loops at some initial ($C_i$) and final ($C_f$) time. So in principle there will be a number of preassigned boundary terms in the string amplitude, corresponding to some specified initial and final configurations for one or more strings. In string perturbation theory one would then consider breaking up the sum into terms involving increasingly complex topologies, starting from the trivial one, say a two-sphere or a tree-level type diagram in the case of an amplitude describing string interactions. It is clear that a quantity of central importance in the first quantized bosonic string is in fact the N-point scattering amplitude, which is computed by treating the incoming and outgoing strings essentially as point-like entities (which in the bosonic string theory are actually tachyons with some momentum $k$) connected to string world sheet surface on specific surface points.

The quantum string of Eq. (1.195) is therefore more generally defined by an action for the surface $S$ given by

$$I\,[g,X] \;=\; \lambda \int_S d^2\sigma\,\sqrt{g} \,+\, \tfrac{1}{2}\int d^2\sigma\,\sqrt{g}\,g^{ab}\partial_a X^\mu\,\partial_b X^\mu \,+\, k\,\chi \ , \tag{1.219}$$

where $\lambda$ is a (so far arbitrary) cosmological term, required later by renormalization effects, and $k\chi$ a curvature term, related by the Gauss-Bonnet theorem to the Euler characteristic $\chi$ of the surface,

$$\chi \;=\; \frac{1}{4\pi}\int d^2\sigma\,\sqrt{g}\,{}^2R + \frac{1}{2\pi}\int_{\partial S} ds\,K \ . \tag{1.220}$$

Here ${}^2R(g)$ is the scalar curvature on the two-dimensional surface, and $K$ the curvature on the boundary $\partial S$. The path integral is then a sum over all such surfaces,

$$Z \;=\; \int [dg_{ab}]\,[dX^\mu]\,\exp(\,-I[g,X]\,) \ , \tag{1.221}$$

with $[dg_{ab}]$ an invariant functional measure over the two-dimensional metrics. The Euler characteristic term is a constant for a surface of given topology, but adds a relative weight $e^{-k\chi}$ for string contributions with different topology. Eventually one would like to sum over all topologies for the string (i.e. sum over surfaces of arbitrary genus $h$ with $\chi = 2[1-h)]$, but what could stand in the way is that bosonic string perturbation theory seems badly divergent and is not Borel summable, $Z \sim \sum_h g^h h!$, contrary to what happens for example in *QED*, (Gross and Periwal, 1988). Such a result is usually taken as an indication of a serious vacuum instability and large non-perturbative contributions (Parisi, 1979) (originally this estimate was in fact viewed as a welcome feature, as there were many aspects of perturbative string theory that were not shared by the real world!).

The usual treatment of quantum two-dimensional gravity then proceeds to set the metric in the conformal gauge $g_{ab}(x) = e^{\phi}(x)\,\tilde{g}_{ab}$, where $\tilde{g}_{ab}$ is a reference metric, usually taken to be the flat one, $\delta_{ab}$. The conformal gauge fixing for the metric then implies a non-trivial Faddeev Popov determinant, which when exponentiated results in the Liouville action for two-dimensional pure gravity in the conformal gauge

$$I[\phi] = \frac{13}{24\pi} \int d^2\sigma \left[ \tfrac{1}{2}(\partial_a \phi)^2 + \mu^2 e^{\phi} \right] \,, \tag{1.222}$$

with the $\mu$-term amounting to a renormalization of the bare cosmological constant. In the language of the conformal gauge, where $\sqrt{g} = e^{\phi}$ and $R = e^{-\phi}\,\partial^2\phi$, the preceding action can in fact be re-written in arbitrary coordinate as a nonlocal contribution.

When the $d$ scalar $X$ fields are coupled to the two-dimensional gravity and integrated out (since they appear quadratically in the action), the conformal anomaly contribution modifies the Liouville effective action to

$$I[\phi] = \frac{26-d}{48\pi} \int d^2\sigma \left[ \tfrac{1}{2}(\partial_a \phi)^2 + \mu^2 e^{\phi} \right] \,, \tag{1.223}$$

which suggests that the bosonic string only makes sense for embedding dimensions $d < 26$, i.e. less than the critical dimension known already from dual models.

So far fluctuations in the string are essentially unconstrained when viewed from embedding space. But it is possible to add some rigidity to the string by considering extrinsic curvature terms (Polyakov, 1786). There are a number of equivalent ways of writing such contributions, the simplest one being of the form

$$I_{\text{ex}}[g,X] = \frac{1}{2\alpha} \int_S d^2\sigma \sqrt{g}\, g^{ab}\, \partial_a t_{\mu\nu}\, \partial_b t_{\mu\nu} \tag{1.224}$$

with

$$t_{\mu\nu} = \varepsilon^{ab} \frac{1}{\sqrt{g}}\, \partial_a X^{\mu}\, \partial_b X^{\nu} \,, \tag{1.225}$$

and $\alpha$ a dimensionless coupling. Note that the new term involves higher derivatives of the "matter field" $X^{\mu}$. A one-loop calculation shows that the renormalization group behavior for $\alpha$ is $\alpha^{-1}(\mu) = \alpha^{-1}(\Lambda) + (D/4\pi)\log(\mu/\Lambda)$ with $\Lambda$ the ultraviolet cutoff. This implies that the coupling $\alpha$ is asymptotically free, and therefore grows with distance. Since it is $1/\alpha$ that appears in the action, one would still recover the Nambu-Goto string in the continuum limit, unless some new unexpected fixed points emerge to higher order.

Finally it is possible to generalize the bosonic string action of Eq. (1.195) by including a background $X$ that is not flat (Callan, Friedan, Martinec and Perry, 1985)

$$I_{\mathrm{nlsm}}[g,X,b,\phi] = \frac{1}{4\pi\alpha'} \int d^2\sigma \sqrt{g} \, [\sqrt{g} \, g^{ab} \, g_{\mu\nu}(X) \, \partial_a X^\mu \, \partial_b X^\mu + g^{-1/2} \varepsilon^{ab} A_{\mu\nu}(X) \, \partial_a X^\mu \, \partial_b X^\mu - \tfrac{1}{2} \alpha'^2 R \, \phi(X)] \ , \tag{1.226}$$

where the new fields include the graviton $g_{\mu\nu}(X)$, an antisymmetric tensor field $A_{\mu\nu}(X)$, and the dilaton $\phi(X)$. These models are usually referred to, for historical reasons, as non-linear sigma models for strings. Note, from the structure of the last term in Eq. (1.226) involving ${}^2R$, that the dilaton field is related to the string coupling "constant", with the $n$-loop amplitude involving a factor $e^{-2(1-n)\phi}$ (at least for a slowly varying dilaton field).

In general these theories will no longer be conformally invariant unless one imposes conditions on the "beta functions for each field", such as $R_{\mu\nu}(X) + \ldots = 0$, where the ellipsis refers to contributions from the other two fields $\phi(X)$ and $A_{\mu\nu}(X)$. It can be shown that these string consistency equations can be derived from a $d$-dimensional action with a rather simple form

$$I_{\mathrm{dil}} = -\frac{1}{16\pi G} \int d^d X \sqrt{G} e^{\phi} \left\{ R + 4(\nabla\phi)^2 - \tfrac{1}{12} F^2_{\mu\nu\sigma} + \ldots \right\} \ , \tag{1.227}$$

with $F_{\mu\nu\sigma}$ the curl of $A_{\mu\nu}$. After rescaling the $d$-dimensional metric $G_{\mu\nu} \to e^{4\phi/(d-2)} G_{\mu\nu}$ one obtains the more familiar form

$$I_{\mathrm{dil}} = -\frac{1}{16\pi G} \int d^d X \sqrt{G} \left\{ R - \frac{4}{d-2} (\nabla\phi)^2 - \tfrac{1}{12} e^{-8\phi(d-2)} F^2_{\mu\nu\sigma} + \ldots \right\} \ , \tag{1.228}$$

which shows the general feature of strings coupled to background gravity: they generally involve dilaton corrections to Einstein gravity.

## 1.11 Supersymmetric Strings

From the preceding discussion it appears that there are three main problems with the bosonic string, the first one being that the ground state is a tachyon, a particle of mass $m^2 < 0$. The second problem is that the bosonic string is only consistently defined in $d = 26$ spacetime dimensions, and the third problem is that it does not contain fermions which are after all an essential component of ordinary matter.

One way of introducing fermionic degrees of freedom is to look for a supersymmetric extension of the bosonic string action of Eq. (1.195) (Brink and Schwarz, 1977),

$$I[g,\chi,X,\psi] = \frac{1}{2}\int d^2\sigma\, e\,[\, g^{ab}\, \partial_a X^\mu\, \partial_b X^\mu + \bar{\psi}^\mu\, i\gamma^a\, \partial_a\, \psi^\mu$$
$$+\; \bar{\chi}_a \gamma^b \gamma^a (\partial_b X^\mu + \tfrac{1}{2}\chi_b\, \psi^\mu)\, \psi^\mu\,] \; . \quad (1.229)$$

Here again $X^\mu(\sigma,\tau)$ ($\mu = 1\ldots d$) parametrizes the surface, $\psi^\mu(\sigma,\tau)$ is a two-component Majorana spinor, $\chi_a(\sigma,\tau)$ a spin-$\frac{3}{2}$ gravitino field (a two-component Majorana spinor and a world-sheet vector), and $e^\alpha_a(\sigma,\tau)$ a zweibein for the metric $g_{ab}$, such that $\sqrt{g} = e$. In order to ensure local supersymmetry, $g_{ab}$ and $\chi_a$ have to be treated as independent variables. The action of Eq. (1.229) now has a much larger invariance, which consists of the local supersymmetry transformations

$$\begin{aligned} \delta X^\mu &= \bar{\varepsilon}\, \psi^\mu & \delta\psi^\mu &= -i\gamma^a\, \varepsilon\, (\partial_a X^\mu - \bar{\psi}^\mu\, \chi_a) \\ \delta e^\alpha_a &= -2i\bar{\varepsilon}\, \gamma^\alpha\, \chi_a & \delta\chi_a &= \nabla_a\, \varepsilon \;, \end{aligned} \quad (1.230)$$

with $\varepsilon(x)$ an arbitrary fermionic function. In addition there is the local Weyl (or conformal) symmetry, already present in the bosonic string,

$$\begin{aligned} \delta X^\mu &= 0 & \delta\psi^\mu &= -\tfrac{1}{2}\Lambda\, \psi^\mu \\ \delta e^\alpha_a &= \Lambda\, e^\alpha_a & \delta\chi_a &= \tfrac{1}{2}\Lambda\, \chi_a \;, \end{aligned} \quad (1.231)$$

with $\Lambda(x)$ a real function, as well as the purely fermionic local symmetry

$$\begin{aligned} \delta X^\mu &= 0 & \delta\psi^\mu &= 0 \\ \delta e^\alpha_a &= 0 & \delta\chi_a &= i\gamma_a\, \eta \;, \end{aligned} \quad (1.232)$$

with $\eta(x)$ an arbitrary Majorana spinor. The resulting invariance under $\varepsilon$, $\Lambda$ and $\eta$ transformations is denoted as superconformal.

Just as conformal invariance of the bosonic string restricted the Virasoro algebra for the quantities $L_m$, here the corresponding superconformal symmetry will restrict the structure of the commutation relations for the quantities $L_m$, $F_m$ and $G_r$. One now finds that the theory is ghost-free provides $d = 10$ and $a = 1/2$ in the bosonic sector, and $a = 0$ in the fermionic one.

Alternatively one can treat the theory using covariant functional integral methods. First one needs to fix the superconformal gauge by the choice

$$g_{ab}(\sigma,\tau) = e^2(\sigma,\tau)\, \delta_{ab} \qquad \chi_a(\sigma,\tau) = \gamma_a\, \chi(\sigma,\tau) \;, \quad (1.233)$$

after which one can integrate out the $\psi$ and $X$ fields (Polyakov, 1981a,b) as was done in the bosonic case. This gives an effective action for the $e$ and $\chi$ fields just defined,

$$e^{-S(e,\chi)} = \int [d\psi]\, [dX]\, e^{-I[g,\chi,X,\psi]} \; . \quad (1.234)$$

The superconformal gauge fixing implies a Faddeev-Popov determinant contribution to the functional measure; after this contribution is properly taken into account one finds an effective action given in terms of the direct supersymmetric extension of the Liouville action,

$$I[\phi,\chi] = \frac{10-d}{8\pi}\int d^2\sigma\left[\tfrac{1}{2}(\partial_a\phi)^2 + \tfrac{1}{2}\mu^2 e^{2\phi} + \tfrac{1}{2}i\bar{\chi}(\gamma\cdot\partial)\chi + \tfrac{1}{2}\mu(\bar{\chi}\gamma_5\chi)e^{\phi}\right] , \tag{1.235}$$

where $d$ again is the number of components of the original $X$ field, or, equivalently, the embedding dimension of the supersymmetric string. The theory then describes a two-dimensional renormalizable field theory, which is intended to reproduce a sum over random surfaces with fermionic structure. The result of Eq. (1.235) implies that the supersymmetric string only makes sense in ten dimensions, just like the bosonic string was only consistent in ten dimensions. There is one important difference though: the ground state of the supersymmetric string can be chosen so as to avoid the tachyon (the requirement to achieve this are known as the GSO conditions). Also, the previous discussion dealt with an extension of the original bosonic string which included $\mathcal{N}=1$ world-sheet supersymmetry. It is possible though to consider a wider range of string theories which have $\mathcal{N}=2$ (with a local $SO(2)$ invariance) and $\mathcal{N}=4$ world-sheet supersymmetry (with a local $SU(2)$ invariance).

But there is one important physical aspect that is still missing in the ten-dimensional supersymmetric theory, and that is the presence of non-abelian gauge bosons, which are necessary, at least for a grand unified theory, in order to eventually make contact with the real world. There are two approaches one can follow in introducing gauge interactions in $d=10$, which will be outlined here.

In the first approach the new internal symmetry charges are placed at the ends of the string. In type $I$ superstring has one supersymmetry ($\mathcal{N}=1$) in the ten-dimensional sense, and therefore 16 supercharges. The unique feature of this theory is that it is based on unoriented open and closed strings; since only type $I$ superstring theories contain open strings, only there this approach is possible. In type $I$ strings the symmetry group is $SO(32)$.

The remaining string theories are based on oriented closed strings. The type II string has $\mathcal{N}=2$ supersymmetry and cannot be consistently coupled to open strings, which only allow at most $\mathcal{N}=1$ supersymmetry. Here one has two supersymmetries in the ten-dimensional sense, giving 32 supercharges. There are in fact two kinds of type *II* strings, the *IIA* and the *IIB* type. The main difference is in the fact that the *IIA* theory massless fermion modes are such that the theory is non-chiral and thus parity conserving, while the *IIB* theory is chiral and therefore parity violating.

In the so-called heterotic string one again has supersymmetry and closed strings only, but with separate right- and left-moving string modes. These models are based therefore on a peculiar hybrid of the type $I$ superstring and a bosonic string. There are two kinds of heterotic strings which differ in their ten-dimensional gauge groups, which can be either $SO(32)$ or $E_8\times E_8$ (Gross, Harvey, Martinec and Rohm, 1985). In these theories the charges are distributed on closed strings (as stated before, open strings are not possible in the heterotic string scenario). Since it is a general feature of closed-string theories (at least in flat backgrounds) that the left- and right-moving modes are decoupled, one has the freedom of allowing left-moving modes to be of one type, and right-moving modes to be of a different type. The hybridization implied by the name heterotic string comes about because the right moving modes are chosen to be superstring modes (thus avoiding the tachyon of the bosonic string), whereas the left-moving ones are taken to be representations of a suitable current algebra describing gauge degrees of freedom. One surprising aspect that was discovered later is the fact type $I$ superstrings are in fact "dual" to the heterotic $SO(32)$ superstring theory, and therefore more closely related to it than would appear on the surface.

A theory of this type is described by the action

$$I[g,\chi,X,\psi] = \frac{1}{2}\int d^2\sigma \left( g^{ab}\,\partial_a X^\mu\,\partial_b X^\mu - 2i\,\bar{\psi}_-^\mu\,\partial_+\,\psi_-^\mu - 2i\sum_{A=1}^{n}\lambda_+^A\,\partial_-\lambda_+^A \right) . \tag{1.236}$$

Here $\psi^\mu$, with $\mu = 1\dots 10$, is a Majorana-Weyl fermion transforming as the vector representation of the Lorentz group, and $\lambda^A$, with $A = 1\dots n$, a set of Majorana-Weyl fermions which are Lorentz singlets, possibly endowed with some internal quantum numbers. For a ten-dimensional supersymmetric theory one has a 32-component Majorana spinor. This can be decomposed into a pair of 16-component Majorana-Weyl (chiral) spinors. The right-moving modes are the $\psi^\mu$ and the right-moving part of $X^\mu$, whereas the left-moving modes are the left-moving part of $X^\mu$ and the $\lambda^A$ fields. Note that the absence of a "left-moving" supersymmetry. For $n = 32$ one obtains the heterotic $SO(32)$ theory which is anomaly free. Another possibility is a theory based on the exceptional group $E_8 \times E_8$.

The anomaly issue in ten-dimensional string theories arises from the fact that chiral gauge theories can be inconsistent due to anomalies. This happens already in ordinary non-abelian gauge theories when certain one-loop Feynman diagrams cause a quantum mechanical breakdown, induced by radiative corrections, of the gauge symmetry, which leads to a non-conservation of gauge currents. In string theory the anomalies were canceled out via the Green-Schwarz mechanism, which severely restricts the gauge group structure of ten dimensional strings.

String theories can in principle contain additional rather exotic objects, such as $D$-Branes; these are membrane-like configuration in the ten-dimensional superstring theory, which can occur as a result of a Kaluza-Klein compactification of an eleven-dimensional $M$-theory (matrix theory) containing membranes.

Of course one key issue that ten-dimensional supersymmetric string theories have to face eventually is how to come down from ten to four dimensions, which raises the key problem of string compactification, a largely unsolved problem to this day. One usually assumes that the ten-dimensional string compactifies to four dimension through the choice of some suitable manifold (often of the Calabi-Yau type) so that the main phenomenologically desirable feature of the original ten-dimensional theory (such as chiral fermions) survive in four dimensions. There is no clear input from the ten-dimensional string dynamics on what choices might be dynamically allowed or favored, the manifold is generally chosen so that desirable features are achieved for the four-dimensional theory; indeed recent work suggests that there might be infinitely many choices for the vacua in ten dimensions. Generally in these compactification scenarios the additional unwanted six space dimensions are curled up in tiny tubes whose spatial extent is of the order of the string scale, although several alternative scenarios are possible (for a recent review see Font and Theisen, 2005). One of the more speculative points of view asserts that string theory not only predicts extra dimensions, but that the strength of four-dimensional gravity is in fact affected by the presence of these extra dimensions by making it stronger in these higher dimensional directions, an effect which be detectable by future high precision experiments.

There is one generic feature of all string-theoretic models of gravity, and that is the appearance of an extra scalar particle called the dilaton. All perturbative string theories (type I, type II and heterotic) already start out with a dilaton in ten dimensions.

Although different in the details, the low energy effective action for the gravitational degrees of freedom is similar to the bosonic dilaton actions of Eqs. (1.227) and (1.228). In particular for the heterotic string one finds (Fradkin and Tseytlin, 1985)

$$I_{\mathrm{dil}}[G,\phi,A_{\mu\nu}] \;=\; -\frac{1}{16\pi G}\int d^dX\,\sqrt{G}\,e^{-2\phi}\,\left\{R+4\,(\partial_\mu\phi)^2-\tfrac{1}{12}F^2_{\mu\nu\sigma}+\ldots\right\}\;, \tag{1.237}$$

with $R$ the scalar curvature for the metric $G_{\mu\nu}$, and

$$\sqrt{16\pi G}\;\sim\; g\,\alpha'^{(d-2)/4}\;, \tag{1.238}$$

with $g$ the dimensionless string coupling constant. Away from ten dimensions one also has a cosmological constant contribution which is order one in units of $\alpha'$. Here and in the following it will be assumed that compactification has occurred by now, so that the effective low energy theory resides in the physical dimension $d=4$. Furthermore it is a general feature of the string that it contains both massless and massive modes. In the low energy effective field theory description only the massless modes are retained, but the effect of the massive modes can in many ways be regarded as being equivalent to having a cutoff at the string mass scale $\Lambda_s=(\alpha')^{-1/2}$.

After the Weyl rescaling $G_{\mu\nu}\to G_{\mu\nu}\exp[4\pi/(d-2)]$ the action coincides with the corresponding part of $\mathcal{N}=2$ $d=10$ supergravity,

$$I_{\mathrm{dil}} \;=\; -\frac{1}{16\pi G}\int d^dX\,\sqrt{G}\,\left\{R-\tfrac{1}{2}(\partial_\mu\phi)^2-\tfrac{1}{12}F^2_{\mu\nu\sigma}\,e^{-\phi}+\ldots\right\}\;. \tag{1.239}$$

The last term involving the gauge fields contains a dilaton-field dependent gauge coupling constant, so that the main modification to the matter sector is in the form

$$I_{\text{matt}} = \int d^d X \sqrt{G} \left\{ -\bar{\psi} D \psi - \frac{1}{4 g^2(\phi)} F^2_{\mu\nu\sigma} + \dots \right\} \,, \tag{1.240}$$

with $g^{-2} \equiv e^{-\phi}/12$, with the vacuum expectation value of the dilaton field, related to the original string coupling constant $g_{st}$ by $g^2_{\text{st}} = \langle e^{2\phi} \rangle$. Therefore the coupling constant $g_{\text{st}}$ is a dynamical variable in string theory, unlike the case of quantum field theory where it is usually considered a true constant, up to renormalization group effects. As long as supersymmetry is unbroken, scalar fields like $\phi$ can take arbitrary values (they are referred to as moduli). However, supersymmetry breaking could create a potential for these scalar fields, whose minima could then in principle be calculable. But since the non-perturbative potential for the dilaton field is not known, there is no theory for the dilaton mass either.

Nevertheless there are two main, in principle observable, effects of the dilaton in low energy phenomena (Damour and Polyakov, 1994; Damour, Piazza and Veneziano, 2002). The first one is that the dilaton field enters the gravitational action and modifies it, as shown in Eq. (1.239). It can therefore lead to violations of the Equivalence Principle, which could be large if the dilaton mass is small. Therefore high precision tests of the Equivalence Principle such as the universality of free fall, could be viewed as possible windows on string-scale physics. Similarly a small dilaton mass could affect high precision tests of the inverse square law for gravity on sub-millimeter scales, although, by the nature of its couplings, a light dilaton is not likely to play an important role in cosmological evolution.

The second main effect of the dilaton is its influence on the gauge coupling constant, through the gluon field strength in Eq. (1.240) (Taylor and Veneziano, 1988; Kaplan and Wise, 2000). Again specific investigation of the effects associated with the dilaton require some educated guess on its mass, which in some scenarios, based on specific mechanisms for supersymmetry breaking, is assumed to be of the order $m \sim \Lambda^2_{\text{susy}}/\mu_p$, giving for $\Lambda_{\text{susy}} \sim 1TeV$ a dilaton Compton wavelength of a few millimeters. From these numbers on can then make appropriate estimates on the modifications of matter couplings.

Finally one interesting aspect of string theories is how they relate to the physics of black holes. In the field theory (supergravity) description of strings the $D$-branes that appear in the perturbative string picture re-emerge in the supergravity framework as so-called black-branes. In the supergravity framework it is in fact more natural to look at charged black holes in anti-DeSitter space, since these spaces are supergravity solutions with maximal supersymmetry. It is believed that in string theory black hole evaporation then arise through the emission of closed strings from excited $D$-branes.

In conclusion superstring theory provides a fascinating alternative to the traditional field theoretic approach to quantum gravity. Yet it still has to confront some very basic issues: there is no known non-perturbative formulation of strings that would allow the investigation and selection of superstring vacua in ten dimensions. Furthermore, the dynamical mechanism for compactification is not understood, instead the usual avenue for compactification is the selection of a class of manifolds which appear to have desirable properties in four dimensions. Indeed there is to this date no compelling argument for why one should end up in four dimensions instead of three or five. And finally there is the remaining core issue of what the mechanisms for supersymmetry breaking might be, for which string theory so far has not provided an indication for a possible scenario.

# Chapter 2
# Feynman Path Integral Formulation

## 2.1 The Path Integral

So far the discussion of quantum gravity has focused almost entirely on perturbative scenarios, where the gravitational coupling $G$ is assumed to be weak, and the weak field expansion based on $\bar{g}_{\mu\nu} = g_{\mu\nu} + h_{\mu\nu}$ can be performed with some degree of reliability. At every order in the loop expansion the problem then reduces to the systematic evaluation of an increasingly complex sequence of Gaussian integrals over the small quantum fluctuation $h_{\mu\nu}$.

But there are reasons to expect that non-perturbative effects play an important role in quantum gravity. Then an improved formulation of the quantum theory is required, which does not rely exclusively on the framework of a perturbative expansion. Indeed already classically a black hole solution can hardly be considered as a small perturbation of flat space. Furthermore, the fluctuating metric field $g_{\mu\nu}$ is dimensionless and carries therefore no natural scale. For the simpler cases of a scalar field and non-Abelian gauge theories a consistent non-perturbative formulation based on the Feynman path integral has been known for some time and is by now well developed. Combined with the lattice approach, it provides an effective and powerful tool for systematically investigating non-trivial strong coupling behavior, such as confinement and chiral symmetry breaking. These phenomena are known to be generally inaccessible in weak coupling perturbation theory. Furthermore, the Feynman path integral approach provides a manifestly covariant formulation of the quantum theory, without the need for an artificial $3+1$ split required by the more traditional canonical approach, and the ambiguities that may follow from it. In fact, as will be seen later, in its non-perturbative lattice formulation no gauge fixing of any type is required.

In a nutshell, the Feynman path integral formulation for pure quantum gravitation can be expressed in the functional integral formula

$$Z = \int_{\text{geometries}} e^{\frac{i}{\hbar} I_{\text{geometry}}} \, , \tag{2.1}$$

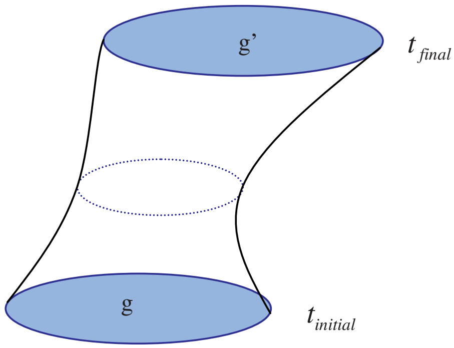



**Fig. 2.1** Quantum mechanical amplitude of transitioning from an initial three-geometry described by $g$ at time $t_{initial}$ to a final three-geometry described by $g'$ at a later time $t_{final}$. The full amplitude is a sum over all intervening metrics connecting the two bounding three-surfaces, weighted by $\exp(iI/\hbar)$ where $I$ is a suitably defined gravitational action.

(for an illustration see Fig. 2.1), just like the Feynman path integral for a non-relativistic quantum mechanical particle (Feynman, 1948; 1950; Feynman and Hibbs, 1965) expresses quantum-mechanical amplitudes in terms of sums over paths

$$A(i \rightarrow f) = \int_{\text{paths}} e^{\frac{i}{\hbar} I_{\text{path}}} \quad . \tag{2.2}$$

What is the precise meaning of the expression in Eq. (2.1)? The remainder of this section will be devoted to discussing attempts at a proper definition of the gravitational path integral of Eq. (2.1). A modern rigorous discussion of path integrals in quantum mechanics and (Euclidean) quantum field theory can be found, for example, in (Albeverio and Hoegh-Krohn, 1976), (Glimm and Jaffe, 1981) and (Zinn-Justin, 2002).

## 2.2 Sum over Paths

Already for a non-relativistic particle the path integral needs to be defined quite carefully, by discretizing the time coordinate and introducing a short distance cutoff. The standard procedure starts from the quantum-mechanical transition amplitude

$$A(q_i, t_i \rightarrow q_f, t_f) \; = < q_f \,|\, e^{-\frac{i}{\hbar} H(t_f - t_i)} \,|\, q_i > \ , \tag{2.3}$$

and subdivides the time interval into $n+1$ segments of size $\varepsilon$ with $t_f = (n+1)\varepsilon + t_i$. Using completeness of the coordinate basis $|q_j>$ at all intermediate times, one obtains the textbook result, here for a non-relativistic particle described by a Hamiltonian $H(p,q) = p^2/(2m) + V(q)$,

$$A(q_i,t_i \to q_f,t_f) = \lim_{n\to\infty} \int_{-\infty}^{\infty} \prod_{j=1}^{n} \frac{dq_j}{\sqrt{2\pi i\hbar\varepsilon/m}} \times \exp\left\{ \frac{i}{\hbar} \sum_{j=1}^{n+1} \varepsilon \left[ \frac{1}{2} m \left( \frac{q_j - q_{j-1}}{\varepsilon} \right)^2 - V\left( \frac{q_j + q_{j-1}}{2} \right) \right] \right\} . \tag{2.4}$$

The expression in the exponent is easily recognized as a discretized form of the classical action. The above quantum-mechanical amplitude $A$ is then usually written in shorthand as

$$A(q_i,t_i \to q_f,t_f) = \int_{q_i(t_i)}^{q_f(t_f)} [dq] \exp\left\{ \frac{i}{\hbar} \int_{t_i}^{t_f} dt\, L(q,\dot{q}) \right\} , \tag{2.5}$$

with $L = \frac{1}{2} m\dot{q}^2 - V(q)$ the Lagrangian for the particle. What appears therefore in the exponent is the classical action

$$I = \int_{t_i}^{t_f} dt\, L(q,\dot{q}) , \tag{2.6}$$

associated with a given trajectory $q(t)$, connecting the initial coordinate $q_i(t_i)$ with the final one $q_f(t_f)$. Then the quantity $[dq]$ is the functional measure over paths $q(t)$, as spelled out explicitly in the precise lattice definition of Eq. (2.4). One advantage associated with having the classical action appear in the quantum mechanical amplitude is that all the symmetries of the theory are manifest in the Lagrangian form. The symmetries of the Lagrangian then have direct implications for the study of quantum mechanical amplitudes. A stationary phase approximation to the path integral, valid in the limit $\hbar \to 0$, leads to the least action principle of classical mechanics

$$\delta I = 0 . \tag{2.7}$$

In the above derivation it is not necessary to use a uniform lattice spacing $\varepsilon$; one could have used as well a non-uniform spacing $\varepsilon_i = t_i - t_{i-1}$ but the result would have been the same in the limit $n \to \infty$ (in analogy with the definition of the Riemann sum for ordinary integrals). Since quantum mechanical paths have a zig-zag nature and are nowhere differentiable, the mathematically correct definition should be taken from the finite sum in Eq. (2.4). In fact it can be shown that differentiable paths have zero measure in the Feynman path integral: already for the non-relativistic particle most of the contributions to the path integral come from paths that are far from smooth on all scales (Feynman and Hibbs, 1965), the so-called Wiener paths, in turn related to Brownian motion. In particular, the derivative $\dot{q}(t)$ is not always defined, and the correct definition for the path integral is the one given in Eq. (2.4). A very complete and contemporary reference to the many applications of path integrals to non-relativistic quantum systems and statistical physics can be found in two recent monographs (Zinn-Justin, 2005; Kleinert, 2006).

As a next step, one can generalized the Feynman path integral construction to $N$ particles with coordinates $q_i(t)$ $(i = 1, N)$, and finally to the limiting case of continuous fields $\phi(x)$. If the field theory is defined from the start on a lattice, then the quantum fields are defined on suitable lattice points as $\phi_i$.

## 2.3 Eulidean Rotation

In the case of quantum fields, one is generally interested in the vacuum-to-vacuum amplitude, which requires $t_i \to -\infty$ and $t_f \to +\infty$. Then the functional integral with sources is of the form

$$Z[J] = \int [d\phi] \exp\left\{ i \int d^4x [\mathscr{L}(x) + J(x)\phi(x)] \right\} \ , \tag{2.8}$$

where $[d\phi] = \prod_x d\phi(x)$, and $\mathscr{L}$ the usual Lagrangian density for the scalar field,

$$\mathscr{L} = -\tfrac{1}{2} [(\partial_\mu \phi)^2 - \mu^2 \phi^2 - i\varepsilon \phi^2] - V(\phi) \ . \tag{2.9}$$

However even with an underlying lattice discretization, the integral in Eq. (2.8) is in general ill-defined without a damping factor, due to the $i$ in the exponent (Zinn-Justin, 2003).

Advances in axiomatic field theory (Osterwalder and Schrader, 1972; 1973; 1975; Glimm and Jaffe, 1974; Glimm and Jaffe, 1981) indicate that if one is able to construct a well defined field theory in Euclidean space $x = (\mathbf{x}, \tau)$ obeying certain axioms, then there is a corresponding field theory in Minkowski space $(\mathbf{x}, t)$ with

$$t = -i\,\tau \ , \tag{2.10}$$

defined as an analytic continuation of the Euclidean theory, such that it obeys the Wightmann axioms (Streater and Wightman, 2000). The latter is known as the *Euclidicity Postulate*, which states that the Minkowski Green's functions are obtained by analytic continuation of the Green's function derived from the Euclidean functional. One of the earliest discussion of the connection between Euclidean and Minkowski filed theory can be found in (Symanzik, 1969). In cases where the Minkowski theory appears pathological, the situation generally does not improve by rotating to Euclidean space. Conversely, if the Euclidean theory is pathological, the problems are generally not removed by considering the Lorentzian case. From a constructive field theory point of view it seems difficult for example to make sense, for either signature, out of one of the simplest cases: a scalar field theory where the kinetic term has the wrong sign (Gallavotti, 1985).

Then the Euclidean functional integral with sources is defined as

$$Z_E[J] = \int [d\phi] \exp\left\{ - \int d^4x [\mathscr{L}_E(x) + J(x)\phi(x)] \right\} \ , \tag{2.11}$$

with $\int \mathscr{L}_E$ the Euclidean action, and

$$\mathscr{L}_E = \tfrac{1}{2}(\partial_\mu \phi)^2 + \tfrac{1}{2}\mu^2\phi^2 + V(\phi) \ , \tag{2.12}$$

with now $(\partial_\mu \phi)^2 = (\nabla\phi)^2 + (\partial\phi/\partial\tau)^2$. If the potential $V(\phi)$ is bounded from below, then the integral in Eq. (2.11) is expected to be convergent. In addition, the Euclidicity Postulate determines the correct boundary conditions to be imposed on the propagator (the Feynman $i\varepsilon$ prescription). Euclidean field theory has a close and deep connection with statistical field theory and critical phenomena, whose foundations are surveyed for example in the comprehensive monographs of (Parisi, 1981) and (Cardy, 1997).

Turning to the case of gravity, it should be clear that to all orders in the weak field expansion there is really no difference of substance between the Lorentzian (or pseudo-Riemannian) and the Euclidean (or Riemannian) formulation. Indeed most, if not all, of the perturbative calculations in the preceding sections could have been carried out with the Riemannian weak field expansion about flat Euclidean space

$$g_{\mu\nu} = \delta_{\mu\nu} + h_{\mu\nu} \ , \tag{2.13}$$

with signature $++++$, or about some suitable classical Riemannian background manifold, without any change of substance in the results. The structure of the divergences would have been identical, and the renormalization group properties of the coupling the same (up to the trivial replacement of say the Minkowski momentum $q^2$ by its Euclidean expression $q^2 = q_0^2 + \mathbf{q}^2$ etc.). Starting from the Euclidean result, the analytic continuation of results such as Eq. (1.161) to the pseudo-Riemannian case would have been trivial.

## 2.4 Gravitational Functional Measure

It is still true in function space that one needs a metric before one can define a volume element. Therefore, following DeWitt (DeWitt, 1962; 1964), one needs first to define an invariant norm for metric deformations

$$\|\delta g\|^2 = \int d^d x \, \delta g_{\mu\nu}(x) \, G^{\mu\nu,\alpha\beta}\big(g(x)\big) \, \delta g_{\alpha\beta}(x) \ , \tag{2.14}$$

with the supermetric $G$ given by the ultra-local expression

$$G^{\mu\nu,\alpha\beta}\big(g(x)\big) = \tfrac{1}{2}\sqrt{g(x)} \left[ g^{\mu\alpha}(x) g^{\nu\beta}(x) + g^{\mu\beta}(x) g^{\nu\alpha}(x) + \lambda \, g^{\mu\nu}(x) g^{\alpha\beta}(x) \right] \ , \tag{2.15}$$

with $\lambda$ a real parameter, $\lambda \neq -2/d$. The DeWitt supermetric then defines a suitable volume element $\sqrt{G}$ in function space, such that the functional measure over the $g_{\mu\nu}$'s taken on the form

$$\int [d\, g_{\mu\nu}] \equiv \int \prod_x \Big[ \det G[g(x)] \Big]^{1/2} \prod_{\mu \geq \nu} d g_{\mu\nu}(x) \ . \tag{2.16}$$

The assumed locality of the supermetric $G^{\mu\nu,\alpha\beta}[g(x)]$ implies that its determinant is a local function of $x$ as well. By a scaling argument given below one finds that, up to an inessential multiplicative constant, the determinant of the supermetric is given by

$$\det G[g(x)] \propto (1 + \tfrac{1}{2} d\, \lambda) \left[ g(x) \right]^{(d-4)(d+1)/4} \ , \tag{2.17}$$

which shows that one needs to impose the condition $\lambda \neq -2/d$ in order to avoid the vanishing of $\det G$. Thus the local measure for the Feynman path integral for pure gravity is given by

$$\int \prod_x [g(x)]^{(d-4)(d+1)/8} \prod_{\mu \geq \nu} d g_{\mu\nu}(x) \ . \tag{2.18}$$

In four dimensions this becomes simply

$$\int [d\, g_{\mu\nu}] = \int \prod_x \prod_{\mu \geq \nu} d g_{\mu\nu}(x) \ . \tag{2.19}$$

However it is not obvious that the above construction is unique. One could have defined, instead of Eq. (2.15), $G$ to be almost the same, but without the $\sqrt{g}$ factor in front,

$$G^{\mu\nu,\alpha\beta}\left[g(x)\right] = \tfrac{1}{2} \left[ g^{\mu\alpha}(x) g^{\nu\beta}(x) + g^{\mu\beta}(x) g^{\nu\alpha}(x) + \lambda\, g^{\mu\nu}(x) g^{\alpha\beta}(x) \right] \ . \tag{2.20}$$

Then one would have obtained

$$\det G[g(x)] \propto (1 + \tfrac{1}{2} d\, \lambda) \left[ g(x) \right]^{-(d+1)} \ , \tag{2.21}$$

and the local measure for the path integral for gravity would have been given now by

$$\int \prod_x [g(x)]^{-(d+1)/2} \prod_{\mu \geq \nu} d g_{\mu\nu}(x) \ . \tag{2.22}$$

In four dimensions this becomes

$$\int [d\, g_{\mu\nu}] = \int \prod_x [g(x)]^{-5/2} \prod_{\mu \geq \nu} d g_{\mu\nu}(x) \ , \tag{2.23}$$

which was originally suggested in (Misner, 1957).

One can find in the original reference an argument suggesting that the last measure is unique, provided the product $\prod_x$ is interpreted over "physical" points, and invariance is imposed at one and the same "physical" point. Furthermore since there are $d(d+1)/2$ independent components of the metric in $d$ dimensions, the Misner measure is seen to be invariant under a re-scaling $g_{\mu\nu} \to \Omega^2 g_{\mu\nu}$ of the metric for any $d$, but as a result is also found to be singular at small $g$. Indeed the DeWitt measure of Eq. (2.18) and the Misner scale invariant measure of Eqs. (2.22) and (2.23) could be just as well regarded as two special cases of a slightly more general supermetric $G$ with prefactor $\sqrt{g}^{(1-\omega)}$, with $\omega = 0$ and $\omega = 1$ corresponding to the original DeWitt and Misner measures, respectively.

The power in Eqs. (2.17) and (2.18) can be found for example as follows. In the Misner case, Eq. (2.22), the scale invariance of the functional measure follows directly from the original form of the supermetric $G(g)$ in Eq. (2.20), and the fact that the metric $g_{\mu\nu}$ has $\frac{1}{2}d(d+1)$ independent components in $d$ dimensions. In the DeWitt case one rescales the matrix $G(g)$ by a factor $\sqrt{g}$. Since $G(g)$ is a $\frac{1}{2}d(d+1) \times \frac{1}{2}d(d+1)$ matrix, its determinant is modified by an overall factor of $g^{d(d+1)/4}$. So the required power in the functional measure is $-\frac{1}{2}(d+1)+\frac{1}{8}d(d+1) = \frac{1}{8}(d-4)(d+1)$, in agreement with Eq. (2.18).

Furthermore, one can show that if one introduces an $n$-component scalar field $\phi(x)$ in the functional integral, it leads to further changes in the gravitational measure. First, in complete analogy to the gravitational case, one has for the scalar field deformation

$$\|\delta\phi\|^2 = \int d^d x \sqrt{g(x)} \left(\delta\phi(x)\right)^2 \ , \tag{2.24}$$

and therefore for the functional measure over $\phi$ one has the expression

$$\int [d\phi] = \int \prod_x \left[\sqrt{g(x)}\right]^{n/2} \prod_x d\phi(x) \ . \tag{2.25}$$

The first factor clearly represents an additional contribution to the gravitational measure. One can indeed verify that one just followed the correct procedure, by evaluating for example the scalar functional integral in the large mass limit,

$$\int \prod_x \left[\sqrt{g(x)}\right]^{n/2} \prod_x d\phi(x) \exp\left(-\tfrac{1}{2}m^2 \int \sqrt{g}\,\phi^2\right) = \left(\frac{2\pi}{m^2}\right)^{nV/2} = \text{const.} \tag{2.26}$$

so that, as expected, for a large scalar mass $m$ the field $\phi$ completely decouples, leaving the dynamics of pure gravity unaffected.

These arguments would lead one to suspect that the volume factor $g^{\sigma/2}$, when included in a slightly more general gravitational functional measure of the form

$$\int [d\, g_{\mu\nu}] = \prod_x [g(x)]^{\sigma/2} \prod_{\mu \geq \nu} dg_{\mu\nu}(x) \ , \tag{2.27}$$

perhaps does not play much of a role after all, at least as far as physical properties are concerned. Furthermore, in $d$ dimensions the $\sqrt{g}$ volume factors are entirely absent ($\sigma = 0$) if one chooses $\omega = 1 - 4/d$, which would certainly seem the simplest choice from a practical point of view.

When considering a Hamiltonian approach to quantum gravity, one finds a rather different form for the functional measure (Leutwyler, 1964), which now includes non-covariant terms. This is not entirely surprising, as the introduction of a Hamiltonian requires the definition of a time variable and therefore a preferred direction, and a specific choice of gauge. The full invariance properties of the original action are no longer manifest in this approach, which is further reflected in the use of a rigid lattice to properly define and regulate the Hamiltonian path integral, allowing subsequent formal manipulations to have a well defined meaning. In the covariant approach one can regard formally the measure contribution as effectively a modification of the Lagrangian, leading to an $L_{eff}$. The additional terms, if treated consistently will result in a modification of the Hamiltonian, which therefore in general will not be of the form one would have naively guessed from the canonical rules (Abers, 2004). One can see therefore that the possible original measure ambiguity found in the covariant approach is still present in the canonical formulation. One new aspect of the Hamiltonian approach is though that conservation of probability, which implies the unitarity of the scattering matrix, can further restrict the form of the measure, if such a requirement is pushed down all the way to the cutoff scale (in a simplicial lattice context, the latter would be equivalent to the requirement of Osterwalder-Schrader reflection positivity at the cutoff scale). Whether such a requirement is physical and meaningful in a geometry that is strongly fluctuating at short distances, and for which a notion of time and orthogonal space-like hypersurfaces is not necessarily well defined, remains an open question, and perhaps mainly an academic one. When an ultraviolet cutoff is introduced (without which the theory would not be well defined), one is after all concerned in the end only with distance scales which are much larger than this short distance cutoff.

Along these lines, the following argument supporting the possible irrelevance of the measure parameter $\sigma$ can be given (Faddeev and Popov, 1973; Fradkin and Vilkovisky, 1973). Namely, one can show that the gravitational functional measure of Eq. (2.27) is invariant under infinitesimal general coordinate transformations, irrespective of the value of $\sigma$. Under an infinitesimal change of coordinates $x'^{\mu} = x^{\mu} + \varepsilon^{\mu}(x)$ one has

$$\prod_x [g(x)]^{\sigma/2} \prod_{\mu \geq \nu} dg_{\mu\nu}(x) \;\to\; \prod_x \left( \det \frac{\partial x'^{\beta}}{\partial x^{\alpha}} \right)^{\gamma} [g(x)]^{\sigma/2} \prod_{\mu \geq \nu} dg_{\mu\nu}(x) \;, \tag{2.28}$$

with $\gamma$ a power that depends on $\sigma$ and the dimension. But for an infinitesimal coordinate transformations the additional factor is equal to one,

$$\prod_x \left( \det \frac{\partial x'^{\beta}}{\partial x^{\alpha}} \right)^{\gamma} = \prod_x \left[ \det( \delta_{\alpha}{}^{\beta} + \partial_{\alpha} \varepsilon^{\beta} ) \right]^{\gamma} \;=\; \exp \left\{ \gamma \delta^d(0) \int d^d x \, \partial_{\alpha} \varepsilon^{\alpha} \right\} = 1 \;, \tag{2.29}$$

and we have used

$$a^d \sum_x \to \int d^d x \;, \tag{2.30}$$

with lattice spacing $a = \pi/\Lambda$ and momentum cutoff $\Lambda$ [see Eq. (1.98)]. So in some respects it appears that $\sigma$ can be compared to a gauge parameter.

In conclusion, there is no clear a priori way of deciding between the various choices for $\sigma$, and the evidence so far suggests that it may very well turn out to be an irrelevant parameter. The only constraint seems that the regularized gravitational path integral should be well defined, which would seem to rule out singular measures, which need additional regularizations at small volumes. It is noteworthy though that the $g^{\sigma/2}$ volume term in the measure is completely local and contains no derivatives. Thus in perturbation theory it cannot affect the propagation properties of gravitons, and only contributes ultralocal $\delta^d(0)$ terms to the effective action, as can be seen from

$$\prod_x [g(x)]^{\sigma/2} = \exp\left\{ \tfrac{1}{2}\, \sigma\, \delta^d(0) \int d^d x \ln g(x) \right\} \tag{2.31}$$

with

$$\ln g(x) = \tfrac{1}{2}\, h_\mu^{\ \mu} - \tfrac{1}{4}\, h_{\mu\nu} h^{\mu\nu} + O(h^3) \ , \tag{2.32}$$

which follows from the general formal expansion formula for an operator $\mathbf{M} \equiv 1 + \mathbf{K}$

$$\operatorname{tr} \ln(1 + \mathbf{K}) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \operatorname{tr} \mathbf{K}^n \ , \tag{2.33}$$

which is valid provided the traces of all powers of $\mathbf{K}$ exist. On a spacetime lattice one can interpret the delta function as an ultraviolet cutoff term, $\delta^d(0) \approx \Lambda^d$. Then the first term shifts the vacuum solution and the second one modifies the bare cosmological constant. To some extent these type of contributions can be regarded as similar to the effects arising from a renormalization of the cosmological constant, ultimately affecting only the distribution of local volumes. So far numerical studies of the lattice models to be discussed later show no evidence of any sensitivity of the critical exponents to the measure parameter $\sigma$.

Later in this review (Sect. 6.9) we will again return to the issue of the functional measure for gravity in possibly the only context where it can be posed, and to some extent answered, satisfactorily: in a lattice regularized version of quantum gravity, going back to the spirit of the original definition of Eq. (2.4).

In conclusion, the Euclidean Feynman path integral for pure Einstein gravity with a cosmological constant term is given by

$$Z_{cont} \ = \ \int [d\, g_{\mu\nu}] \ \exp\left\{ -\lambda_0 \int dx \sqrt{g} + \frac{1}{16\pi G} \int dx \sqrt{g} R \right\} \ . \tag{2.34}$$

It involves a functional integration over all metrics, with measure given by a suitably regularized form of

$$\int [d\, g_{\mu\nu}] \ = \ \int \prod_x \, [g(x)]^{\sigma/2} \prod_{\mu \geq \nu} d g_{\mu\nu}(x) \ , \tag{2.35}$$

as in Eqs. (2.18), (2.22) and (2.27). For geometries with boundaries, further terms should be added to the action, representing the effects of those boundaries. Then the path integral will depend in general on some specified initial and final three-geometry (Hartle and Hawking, 1977; Hawking, 1979).

## 2.5 Conformal Instability

Euclidean quantum gravity suffers potentially from a disastrous problem associated with the conformal instability: the presence of kinetic contributions to the linearized action entering with the wrong sign.

As was discussed previously in Sect. 1.7, the action for linearized gravity without a cosmological constant term, Eq. (1.7), can be conveniently written using the three spin projection operators $P^{(0)}, P^{(1)}, P^{(2)}$ as

$$I_{\rm lin} = \frac{k}{4} \int dx \, h^{\mu\nu} \, [P^{(2)} - 2P^{(0)}]_{\mu\nu\alpha\beta} \, \partial^2 \, h^{\alpha\beta} \quad , \tag{2.36}$$

so that the spin-zero mode enters with the wrong sign, or what is normally referred to as a ghost contribution. Actually to this order it can be removed by a suitable choice of gauge, in which the trace mode is made to vanish, as can be seen, for example, in Eq. (1.13). Still, if one were to integrate in the functional integral over the spin-zero mode, one would have to distort the integration contour to complex values, so as to render the functional integral convergent.

The problem is not removed by introducing higher derivative terms, as can be seen from the action for the linearized theory of Eq. (1.150),

$$\begin{aligned} I_{\rm lin} = \tfrac{1}{2} \int dx \, \{ \, & h^{\mu\nu} \, [\tfrac{1}{2} k + \tfrac{1}{2} a \, (-\partial^2)] (-\partial^2) \, P^{(2)}_{\mu\nu\rho\sigma} \, h^{\rho\sigma} \\ & + h^{\mu\nu} \, [-k - 2b \, (-\partial^2)] (-\partial^2) \, P^{(0)}_{\mu\nu\rho\sigma} \, h^{\rho\sigma} \, \} \quad , \end{aligned} \tag{2.37}$$

as the instability reappears for small momenta, where the higher derivative terms can be ignored [see for example Eq. (1.152)]. There is a slight improvement, as the instability is cured for large momenta, but it is not for small ones. If the perturbative quantum calculations can be used as a guide, then at the fixed points one has $b < 0$, corresponding to a tachyon pole in the spin-zero sector, which would indicate further perturbative instabilities. Of course in perturbation theory there never is a real problem, with or without higher derivatives, as one can just define Gaussian integrals by a suitable analytic continuation.

But the instability seen in the weak field limit is not an artifact of the weak field expansion. If one attempts to write down a path integral for pure gravity of the form

$$Z = \int [d \, g_{\mu\nu}] \, e^{-I_E} \quad , \tag{2.38}$$

with an Euclidean action

$$I_E = \lambda_0 \int dx \sqrt{g} - \frac{1}{16\pi G} \int dx \sqrt{g} R \ , \tag{2.39}$$

one realizes that it too appears ill defined due to the fact that the scalar curvature can become arbitrarily positive, or negative. In turn this can be seen as a direct consequence of the fact that while gravitational radiation has positive energy, gravitational potential energy is negative because gravity is attractive. To see more clearly that the gravitational action can be made arbitrarily negative consider the conformal transformation $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$ where $\Omega$ is some positive function. Then the Einstein action transforms into

$$I_E(\tilde{g}) = -\frac{1}{16\pi G} \int d^4x \sqrt{g} \left( \Omega^2 R + 6\, g^{\mu\nu} \partial_\mu \Omega \, \partial_\nu \Omega \right) \ , \tag{2.40}$$

which can be made arbitrarily negative by choosing a rapidly varying conformal factor $\Omega$. Indeed in the simplest case of a metric $g_{\mu\nu} = \Omega^2 \eta_{\mu\nu}$ one has

$$\sqrt{g}\,(R - 2\lambda) = 6\, g^{\mu\nu} \partial_\mu \Omega \, \partial_\nu \Omega \, - \, 2\lambda \Omega^4 \ , \tag{2.41}$$

which looks like a $\lambda\phi^4$ theory but with the wrong sign for the kinetic term. The problem is referred to as the conformal instability of the classical Euclidean gravitational action (Hawking, 1977). The gravitational action is unbounded from below, and the functional integral is possibly divergent, depending on the detailed nature of the gravitational measure contribution $[dg_{\mu\nu}]$, more specifically its behavior in the regime of strong fields and rapidly varying conformal factors.

A possible solution to the unboundedness problem has been described by Hawking, who suggests performing the integration over all metrics by first integrating over conformal factors by distorting the integration contour in the complex plane to avoid the unboundedness problem, followed by an integration over conformal equivalence classes of metrics (Gibbons and Hawking, 1977; Hawking, 1978a,b; Gibbons, Hawking and Perry, 1978; Gibbons and Perry, 1978). Explicit examples have been given where manifestly convergent Euclidean functional integrals have been formulated in terms of physical (transverse-traceless) degrees of freedom, where the weighting can be shown to arise from a manifestly positive action (Schleich, 1985; Schleich, 1987). A similar convergent procedure seem obtainable for some so-called minisuperspace models, where the full functional integration over the fluctuating metric is replace by a finite dimensional integral over a set of parameters characterizing the reduced subspace of the metric in question, see for example (Barvinsky, 2007). But it is unclear how this procedure can be applied outside perturbation theory, where it not obvious how such a split for the metric should be performed.

An alternate possibility is that the unboundedness of the classical Euclidean gravitational action (which in the general case is certainly physical, and cannot therefore be simply removed by a suitable choice of gauge) is not necessarily an obstacle to defining the quantum theory. The quantum mechanical attractive Coulomb well problem has, for zero orbital angular momentum or in the one-dimensional case, a similar type of instability, since the action there is also unbounded from below. The way the quantum mechanical treatment ultimately evades the problem is that the particle has a vanishingly small probability amplitude to fall into the infinitely deep well. In other words, the effect of quantum mechanical fluctuations in the paths (their zig-zag motion) is just as important as the fact that the action is unbounded. Not unexpectedly, the Feynman path integral solution of the Coulomb problem requires again first the introduction of a lattice, and then a very careful treatment of the behavior close to the singularity (Kleinert, 2006). For this particular problem one is of course aided by the fact that the exact solution is known from the Schrödinger theory.

In quantum gravity the question regarding the conformal instability can then be rephrased in a similar way: Will the quantum fluctuations in the metric be strong enough so that physical excitations will not fall into the conformal well? Phrased differently, what is the role of a non-trivial gravitational measure, giving rise to a density of states $n(E)$

$$Z \propto \int_0^\infty dE \, n(E) \, e^{-E} \, , \tag{2.42}$$

regarding the issue of ultimate convergence (or divergence) of the Euclidean path integral. Of course to answer such questions satisfactorily one needs a formulation which is not restricted to small fluctuations and to the weak field limit. Ultimately in the lattice theory the answer is yes, for sufficiently strong coupling $G$ (Hamber and Williams, 1984; Berg, 1985).

# Chapter 3
# Gravity in $2+\varepsilon$ Dimensions

## 3.1 Dimensional Expansion

In the previous sections it was shown that pure Einstein gravity is not perturbatively renormalizable in the traditional sense in four dimensions. To one-loop order higher derivative terms are generated, which, when included in the bare action, lead to potential unitarity problems, whose proper treatment most likely lies outside the perturbative regime. The natural question then arises: Are there any other field theories where the standard perurbative treatment fails, yet for which one can find alternative methods and from them develop consistent predictions? The answer seems unequivocally yes (Parisi, 1975; 1985). Outside of gravity, there are two notable examples of field theories, the non-linear sigma model and the self-coupled fermion model, which are not perturbatively renormalizable for $d > 2$, and yet lead to consistent and in some instances testable predictions above $d = 2$.

The key ingredient to all of these results is, as originally recognized by Wilson, the existence of a non-trivial ultraviolet fixed point, a phase transition in the statistical field theory context, with non-trivial universal scaling dimensions (Wilson, 1971a,b; Wilson and Fisher, 1972; Wilson, 1973; 1975; Gross, 1976). Furthermore, three quite different theoretical approaches are available for comparing predictions: the $2+\varepsilon$ expansion, the large-$N$ limit, and the lattice approach. Within the lattice approach, several additional techniques are available: the strong coupling expansion, the weak coupling expansion and the numerically exact evaluation of the path integral. Finally, the results for the non-linear sigma model in the scaling regime around the non-trivial ultraviolet fixed point can be compared to high accuracy satellite experiments on three-dimensional systems, and the results agree in some cases to several decimals.

The next three sections will therefore discuss these models from the perspective of those results which will have some relevance later for the gravity case. Of particular interest are predictions for universal corrections to free field behavior, for the scale dependence of couplings, and the role of the non-perturbative correlation length which arises in the strong coupling regime.

Later sections will then discuss the $2+\varepsilon$ expansion for gravity, and what can be learned from it by comparing it to the analogous expansion in the non-linear sigma model. The similarity between the two models is such that they both exhibit a non-trivial ultraviolet fixed point, a two-phase structure, non-trivial exponents and scale-dependent couplings.

## 3.2 Perturbatively Non-renormalizable Theories: The Sigma Model

The $O(N)$-symmetric non-linear $\sigma$-model provides an instructive and rich example of a theory which, above two dimensions, is not perturbatively renormalizable in the traditional sense, and yet can be studied in a controlled way in the context of Wilson's $2+\varepsilon$ expansion. Such framework provides a consistent way to calculate nontrivial scaling properties of the theory in those dimensions where it is not perturbatively renormalizable (for example $d=3$ and $d=4$), which can then be compared to non-perturbative results based on the lattice theory, as well as to experiments, since in $d=3$ the model describes either a ferromagnet or superfluid helium in the vicinity of its critical point. In addition, the model can be solved exactly in the large $N$ limit for any $d$, without any reliance on the $2+\varepsilon$ expansion. Remarkably, in all three approaches it exhibits a non-trivial ultraviolet fixed point at some coupling $g_c$ (a phase transition in statistical mechanics language), separating a weak coupling massless ordered phase from a massive strong coupling phase.

The non-linear $\sigma$-model is described by an $N$-component field $\phi_a$ satisfying a unit constraint $\phi^2(x)=1$, with functional integral given by

$$Z[J] = \int [d\phi] \prod_x \delta\left[\phi(x)\cdot\phi(x)-1\right] \times \exp\left(-\frac{\Lambda^{d-2}}{g} S(\phi) + \int d^d x\, J(x)\cdot\phi(x)\right) . \tag{3.1}$$

The action is taken to be $O(N)$-invariant

$$S(\phi) = \tfrac{1}{2}\int d^d x\, \partial_\mu\phi(x)\cdot\partial_\mu\phi(x) . \tag{3.2}$$

$\Lambda$ here is the ultraviolet cutoff and $g$ the bare dimensionless coupling at the cutoff scale $\Lambda$; in a statistical field theory context $g$ plays the role of a temperature.

In perturbation theory one can eliminate one $\phi$ field by introducing a convenient parametrization for the unit sphere, $\phi(x)=\{\sigma(x),\pi(x)\}$ where $\pi_a$ is an $N-1$-component field, and then solving locally for $\sigma(x)$

$$\sigma(x) = [1-\pi^2(x)]^{1/2} . \tag{3.3}$$

In the framework of perturbation theory in $g$ the constraint $|\pi(x)| < 1$ is not important as one is restricting the fluctuations to be small. Nevertheless the $\pi$ integrations will be extended from $-\infty$ to $+\infty$, which reduces the development of the perturbative expansion to a sequence of Gaussian integrals. Values of $\pi(x) \sim 1$ give exponentially small contributions of order $\exp(-\text{const.}/g)$ which are therefore negligible to any finite order in perturbation theory.

In term of the $\pi$ field the original action $S$ becomes

$$S(\pi) = \tfrac{1}{2} \int d^d x \left[ (\partial_\mu \pi)^2 + \frac{(\pi \cdot \partial_\mu \pi)^2}{1 - \pi^2} \right] . \tag{3.4}$$

The change of variables from $\phi(x)$ to $\pi(x)$ also gives rise to a Jacobian

$$\prod_x \left[ 1 - \pi^2 \right]^{-1/2} \sim \exp\left[ -\tfrac{1}{2}\, \delta^d(0) \int d^d x \ln(1 - \pi^2) \right] , \tag{3.5}$$

which is necessary for the cancellation of spurious tadpole divergences. The combined functional integral for the unconstrained $\pi$ field is then given by

$$Z[J] = \int [d\pi] \exp\left( - \frac{\Lambda^{d-2}}{g} S_0(\pi) + \int d^d x\, J(x) \cdot \pi(x) \right) \tag{3.6}$$

with

$$\begin{aligned} S_0(\pi) = \tfrac{1}{2} \int d^d x & \left[ (\partial_\mu \pi)^2 + \frac{(\pi \cdot \partial_\mu \pi)^2}{1 - \pi^2} \right] \\ & + \tfrac{1}{2}\, \delta^d(0) \int d^d x \, \ln(1 - \pi^2) . \end{aligned} \tag{3.7}$$

In perturbation theory the above action is expanded out in powers of $\pi$. The propagator for the $\pi$ field can be read off from the quadratic part of the action,

$$\Delta_{ab}(k^2) = \frac{\delta_{ab}}{k^2} . \tag{3.8}$$

In the weak coupling limit the $\pi$ fields correspond to the Goldstone modes of the spontaneously broken $O(N)$ symmetry, the latter being broken spontaneously by a non-vanishing vacuum expectation value $\langle \pi \rangle \neq 0$.

Since the $\pi$ field has mass dimension $\frac{1}{2}(d-2)$, and the interactions $\partial^2 \pi^{2n}$ consequently has dimension $n(d-2)+2$, one finds that the theory is renormalizable in $d = 2$ and perturbatively non-renormalizable above $d = 2$. Furthermore, in spite of the theory being non-polynomial, it can still be renormalized via the introduction of only two renormalization constant, the coupling renormalization being given by a constant $Z_g$ and the wavefunction renormalization by a second constant $Z$. Potential infrared problems due to massless propagators are handled by introducing an external $h$-field term for the original composite $\sigma$ field, which then acts as a mass term for the $\pi$ field,

$$
\begin{aligned}
h\int d^d x\,\sigma(x) &= h\int d^d x\,[\,1-\pi^2(x)\,]^{1/2} \\
&= \int d^d x\,[\,h-\tfrac{1}{2}h\,\pi^2(x)+\ldots\,]\ . \qquad (3.9)
\end{aligned}
$$

A proof can be found (David, 1982) that all $O(N)$ invariant Green's are infrared finite in the limit $h \to 0$.

One can write down the same field theory on a lattice, where it corresponds to the $O(N)$-symmetric classical Heisenberg model at a finite temperature $T \sim g$. The simplest procedure is to introduce a hypercubic lattice of spacing $a$, with sites labeled by integers $\mathbf{n} = (n_1 \ldots n_d)$, which introduces an ultraviolet cutoff $\Lambda \sim \pi/a$. On the lattice field derivatives are replaced by finite differences

$$
\partial_\mu \phi(x) \;\to\; \Delta_\mu \phi(\mathbf{n}) = \frac{\phi(\mathbf{n}+\mu)-\phi(\mathbf{n})}{a}\ , \qquad (3.10)
$$

and the discretized path integral then reads

$$
\begin{aligned}
Z[J] &= \int \prod_{\mathbf{n}} d\phi(\mathbf{n})\,\delta[\phi^2(\mathbf{n})-1] \\
&\times \exp\left[-\frac{a^{2-d}}{2g}\sum_{\mathbf{n},\mu}\big(\Delta_\mu\phi(\mathbf{n})\big)^2+\sum_{\mathbf{n}}J(x)\cdot\phi(x)\right]\ . \qquad (3.11)
\end{aligned}
$$

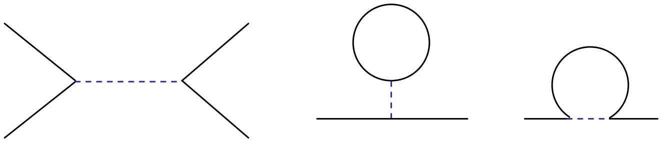
The above expression is recognized as the partition function for a ferromagnetic $O(N)$-symmetric spin system at finite temperature. Besides ferromagnets, it can be used to describe systems which are related to it by universality, such as superconductors and superfluid helium transitions, whose critical behavior is described by a complex phase, and which are therefore directly connected to the plane rotator $N=2$, or $U(1)$, model.

In addition the lattice model of Eq. (3.11) provides an explicit regularization for the continuum theory, which makes expressions like the one in Eq. (3.5) acquire a well defined meaning. It is in fact the only regularization which allows a discussion of the role of the measure in perturbation theory (Zinn-Justin, 2002). At the same time it provides an ultraviolet regularization for perturbation theory, and allows for various non-perturbative calculations, such as power series expansions in three dimensions and explicit numerical integrations of the path integral via Monte Carlo methods.

In two dimensions one can compute the renormalization of the coupling $g$ from the action of Eq. (3.6) and one finds after a short calculation (Polyakov, 1975) for small $g$

$$
\frac{1}{g(\mu)} = \frac{1}{g}+\frac{N-2}{8\pi}\ln\frac{\mu^2}{\Lambda^2}+\ldots \qquad (3.12)
$$

where $\mu$ is an arbitrary momentum scale. Physically one can view the origin of the factor of $N-2$ in the fact that there are $N-2$ directions in which the spin can



(a) (b) (c)

**Fig. 3.1** One-loop diagrams giving rise to coupling and field renormalizations in the non-linear $\sigma$-model. Group theory indices $a$ flow along the *thick lines*, *dashed lines* should be contracted to a point.

experience rapid small fluctuations perpendicular to its average slow motion on the unit sphere, and that only these fluctuations contribute to leading order.

In two dimensions the quantum correction (the second term on the r.h.s.) increases the value of the effective coupling at low momenta (large distances), unless $N = 2$ in which case the correction vanishes. In fact the quantum correction can be shown to vanish to all orders in this case; the vanishing of the $\beta$-function in two dimensions for the $O(2)$ model is true only in perturbation theory, for sufficiently strong coupling a phase transition appears, driven by the unbinding of vortex pairs (Kosterlitz and Thouless, 1973). For $N > 2$ as $g(\mu)$ flows toward increasingly strong coupling it eventually leaves the regime where perturbation theory can be considered reliable. But for bare $g \approx 0$ the quantum correction is negligible and the theory is scale invariant around the origin: the only fixed point of the renormalization group, at least in lowest order perturbation theory, is at $g = 0$. For fixed cutoff $\Lambda$, the theory is weakly coupled at short distances but strongly coupled at large distances. The results in two dimensions for $N > 2$ are qualitatively very similar to asymptotic freedom in four-dimensional $SU(N)$ Yang-Mills theories.

Above two dimensions, $d - 2 = \varepsilon > 0$ and one can redo the same type of perturbative calculation to determine the coupling renormalization. The relevant diagrams are shown in Fig. 3.1. One finds for the effective coupling $g_e$, i.e. the coupling which includes the leading radiative correction (using dimensional regularization, which is more convenient than an explicit ultraviolet cutoff $\Lambda$ for performing actual perturbative calculations),

$$\frac{1}{g_e} = \frac{\Lambda^{\varepsilon}}{g}\left[1 - \frac{1}{\varepsilon}\frac{N-2}{2\pi}g + O(g^2)\right] . \tag{3.13}$$

The requirement that the dimensionful effective coupling $g_e$ be defined independently of the scale $\Lambda$ is expressed as $\Lambda \frac{d}{d\Lambda} g_e = 0$, and gives for the Callan-Symanzik $\beta$-function (Callan, 1970; Symanzik, 1970) for $g$

$$\Lambda \frac{\partial g}{\partial \Lambda} = \beta(g) = \varepsilon g - \frac{N-2}{2\pi} g^2 + O\left(g^3, \varepsilon g^2\right) . \tag{3.14}$$

The above $\beta$-function determines the scale dependence (at least in perturbation theory) of $g$ for an arbitrary scale, which from now on will be denoted as $\mu$. Then the differential equation $\mu \frac{\partial g}{\partial \mu} = \beta[g(\mu)]$ uniquely determines how $g(\mu)$ flows as a function of momentum scale $\mu$. The scale dependence of $g(\mu)$ is such that if the initial $g$ is less than the ultraviolet fixed point value $g_c$, with

$$g_c = \frac{2\pi\varepsilon}{N-2} + \dots \tag{3.15}$$

then the coupling will flow towards the Gaussian fixed point at $g = 0$. The new phase that appears when $\varepsilon > 0$ and corresponds to a low temperature, spontaneously broken phase with finite order parameter. On the other hand if $g > g_c$ then the coupling $g(\mu)$ flows towards increasingly strong coupling, and eventually out of reach of perturbation theory. In two dimensions the $\beta$-function has no zero and only the strong coupling phase is present.

The simplest way of obtaining explicitly the renormalization group behavior of the coupling $g$ is as follows. One parametrises, for $N > 2$, the original $N$-component field $\phi$, which is constrained by $\phi^2 = 1$, in terms of two fields $\theta$ and $\tau$, defined by

$$\begin{aligned}
\phi_1 &= \sqrt{1 - \tau^2} \cos\theta \\
\phi_2 &= \sqrt{1 - \tau^2} \sin\theta \\
\phi_j &= \tau_j \qquad j = 3 \dots N .
\end{aligned} \tag{3.16}$$

In this new set of variables, the original action for the non-linear sigma model becomes

$$S = \frac{1}{2} \int d^d x \left[ (1 - \tau^2) (\partial_\mu \theta)^2 + (\partial_\mu \sqrt{1 - \tau^2})^2 + (\partial_\mu \tau)^2 \right] . \tag{3.17}$$

The $\theta$ variable now enters the action in a very simple way, through a quadratic term. Since the action now contains as unconstrained variables $\theta$ and $\tau$, one can now think, at least for sufficiently small coupling $g$ (low temperatures), of integrating first over the $\tau$ variables, after rescaling $\tau \to \sqrt{g}\,\tau$. The result is an effective action for the $\theta$ variables, and to lowest order the overall effect is to replace in the first term $\tau^2$ by its average $\langle \tau^2 \rangle$. Since $\tau$ is to lowest order a free field, one computes

$$\langle \tau_i(x) \tau_j(y) \rangle = g \Lambda^{2-d} \delta_{ij} \int \frac{d^d k}{(2\pi)^d} \frac{e^{ik\cdot(x-y)}}{k^2} . \tag{3.18}$$

After evaluating the $k$ integral, taking the limit $x \to y$, and doing the trace over the index $i = 3 \dots N$ one obtains for $d > 2$

$$\langle \tau^2 \rangle = (N-2) \frac{g}{2\pi\varepsilon} + O(g^2) . \tag{3.19}$$

This then gives the desired result for the beta function of the nonlinear sigma model, to lowest order in $g$, as in Eqs. (3.13) and (3.14).

In exactly $d = 2$ one needs to cutoff the $k$ integral, and one finds instead $\langle \tau^2 \rangle = (N-2)\frac{g}{2\pi}\log\Lambda + O(g^2)$. So in two dimensions the effective coupling is

$$\frac{1}{g_{\text{eff}}} = \frac{1}{g} - \frac{N-2}{2\pi}\log\Lambda + O(g) \quad (d=2) \ , \tag{3.20}$$

which implies the asymptotic freedom result for $d = 2$ and $N > 2$, namely

$$\beta(g) = -(N-2)\frac{g^2}{2\pi} + O(g^3) \quad (d=2) \ . \tag{3.21}$$

(see Fig. 3.2).

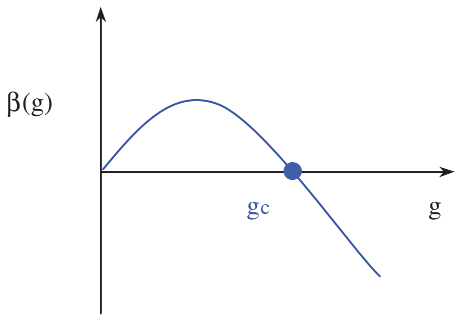



**Fig. 3.2** The $\varepsilon$-expansion result for the Callan-Symanzik $\beta$-function in the non-linear $\sigma$-model for $d > 2$. The non-trivial ultraviolet fixed point is, to lowest order in $\varepsilon$, located at $g_c = 2\pi(d-2)/(N-2)$.

The one-loop running of $g$ as a function of a sliding momentum scale $\mu = k$ and $\varepsilon > 0$ can be obtained by integrating Eq. (3.14). One finds

$$g(k^2) = \frac{g_c}{1 \pm a_0\,(m^2/k^2)^{(d-2)/2}} \ , \tag{3.22}$$

with $a_0$ a positive constant and $m$ a mass scale; the combination $a_0 m^{d-2}$ is just the integration constant for the differential equation, which we prefer to split here in a momentum scale and a dimensionless coefficient for reasons that will become clear later. The choice of $+$ or $-$ sign is determined from whether one is to the left (+), or to right (-) of $g_c$, in which case $g(k^2)$ decreases or, respectively, increases as one flows away from the ultraviolet fixed point. The renormalization group invariant mass scale $\sim m$ arises here as an arbitrary integration constant of the renormalization group equations, and cannot be determined from perturbative arguments alone. It should also be clear that multiplying both sides of Eq. (3.22) by the ultraviolet cutoff factor $\Lambda^{2-d}$ to get back the original dimensionful coupling multiplying the action $S(\phi)$ in Eq. (3.1) does not change any of the conclusions.

It is important to point out that the result of Eq. (3.22) is quite different from the naive expectation based on straight perturbation theory in $d > 2$ dimensions (where the theory is not perturbatively renormalizable)

$$\frac{g(k^2)}{g} \sim 1 + \text{const.}\; g\, k^{d-2} + O(g^2) \;, \tag{3.23}$$

which gives a much worse ultraviolet behavior. The existence of a non-trivial ultraviolet fixed point alters the naive picture and drastically improves the ultraviolet behavior.

For large $g$ one can easily see, for example from the structure of the lattice action in Eq. (3.11), that correlation functions must decay exponentially at large separations. In the strong coupling limit, spins separated by a distance $|x|$ will fluctuate in an uncorrelated fashion, unless they are connected by a minimal number of link contributions from the action. One expects therefore for the lattice connected correlation function of two $\phi$ fields, separated by a lattice distance $na$,

$$< \pi(na)\, \pi(0) >_c \underset{n\to\infty}{\sim} \left(\frac{1}{g}\right)^n \;, \tag{3.24}$$

which can be re-written in continuum language

$$< \pi(x)\, \pi(0) >_c \underset{|x|\to\infty}{\sim} \exp(-|x|/\xi) \;, \tag{3.25}$$

with $m = 1/\xi = \Lambda\,[\ln g + O(1/g)]$ and $\Lambda = 1/a$. From the requirement that the correlation length $\xi$ be a physical quantity independent of scale, and consequently a renormalization group invariant,

$$\Lambda\, \frac{d}{d\Lambda}\, m[\Lambda, g(\Lambda)] = 0 \;, \tag{3.26}$$

one obtains in the strong coupling limit

$$\beta(g) = -g \ln g + O(1/g) \;. \tag{3.27}$$

The quantity $m = 1/\xi$ is usually referred to as the mass gap of the theory, that is the energy difference between the ground state (vacuum) and the first excited state.

If in Eq. (3.22) one sets the momentum scale $k$ equal to the cutoff scale $\Lambda$ and solves for $m$ in the strong coupling phase one obtains

$$\frac{g_c}{g(\Lambda)} = 1 - a_0 \left(\frac{m^2}{\Lambda^2}\right)^{(d-2)/2} \;, \tag{3.28}$$

and therefore for $m$ in terms of the bare coupling $g \equiv g(\Lambda)$

$$m(g) = \Lambda \left(\frac{g_c}{a_0}\right)^{2/(d-2)} \left(\frac{1}{g_c} - \frac{1}{g}\right)^{1/(d-2)} \;. \tag{3.29}$$

This last result shows the following important fact: if $m$ is identified with the inverse of the correlation length $\xi$ (which can be extracted non-perturbatively, for example from the exponential decay of correlation functions, and is therefore generally unambiguous), then the calculable constant relating $m$ to $g$ in Eq. (3.29) uniquely determines the coefficient $a_0$ in Eq. (3.22). For example, in the large $N$ limit the value for $a_0$ will be given later in Eq. (3.60).

In general one can write down the complete renormalization group equations for the cutoff-dependent $n$-point functions $\Gamma^{(n)}(p_i,g,h,\Lambda)$ (Brezin and Zinn-Justin, 1976; Zinn-Justin, 2002). For this purpose one needs to define the renormalized truncated $n$-point function $\Gamma_r^{(n)}$,

$$\Gamma_r^{(n)}(p_i,g_r,h_r,\mu) = Z^{n/2}(\Lambda/\mu,g)\,\Gamma^{(n)}(p_i,g,h,\Lambda) \ , \tag{3.30}$$

where $\mu$ is a renormalization scale, and the constants $g_r$, $h_r$ and $Z$ are defined by

$$\begin{aligned} g &= (\Lambda/\mu)^{d-2} Z_g\, g_r \qquad \pi(x) = Z^{1/2}\pi_r(x) \\ h &= Z_h h_r \qquad\qquad\quad Z_h = Z_g/\sqrt{Z} \ . \end{aligned} \tag{3.31}$$

The requirement that the renormalized $n$-point function $\Gamma_r^{(n)}$ be independent of the cutoff $\Lambda$ then implies

$$\left[\Lambda\frac{\partial}{\partial\Lambda} + \beta(g)\frac{\partial}{\partial g} - \frac{n}{2}\zeta(g) + \rho(g)\,h\frac{\partial}{\partial h}\right]\Gamma^{(n)}(p_i,g,h,\Lambda) = 0 \ , \tag{3.32}$$

with the renormalization group functions $\beta(g)$, $\zeta(g)$ and $\rho(g)$ defined as

$$\begin{aligned} \Lambda\frac{\partial}{\partial\Lambda}\Big|_{\text{ren.fixed}}\, g &= \beta(g) \\ \Lambda\frac{\partial}{\partial\Lambda}\Big|_{\text{ren.fixed}}\,(-\ln Z) &= \zeta(g) \\ 2-d+\tfrac{1}{2}\zeta(g)+\frac{\beta(g)}{g} &= \rho(g) \ . \end{aligned} \tag{3.33}$$

Here the derivatives of the bare coupling $g$, of the $\pi$-field wave function renormalization constant $Z$ and of the external field $h$ with respect to the cutoff $\Lambda$ are evaluated at fixed renormalized (or effective) coupling, at the renormalization scale $\mu$.

To determine the renormalization group functions $\beta(g)$, $\zeta(g)$ and $\rho(g)$ one can in fact follow a related but equivalent procedure, in which, instead of requiring the renormalized $n$-point functions $\Gamma_r^{(n)}$ to be independent of the cutoff $\Lambda$ at fixed renormalization scale $\mu$ as in Eq. (3.33), one imposes that the *bare* $n$-point functions $\Gamma^{(n)}$ be independent of the renormalization scale $\mu$ at *fixed* cutoff $\Lambda$. One can show (Brezin, Le Guillou and Zinn-Justin, 1976) that the resulting renormalization group functions are identical to the previous ones, and that one can obtain the scale dependence of the couplings [i.e. $\beta(g)$] either way. Physically the latter way of thinking is perhaps more suited to a situation where one is dealing with a finite cutoff theory, where the ultraviolet cutoff $\Lambda$ is fixed and one wants to investigate the scale (momentum) dependence of the couplings, for example $g(k^2)$.

For our purposes it will sufficient to look, in the zero-field case $h = 0$, at the $\beta$-function of Eq. (3.14) which incorporates, as should already be clear from the result of Eq. (3.33), a tremendous amount of information about the model. Herein lies the power of the renormalization group: the knowledge of a handful of functions $[\beta(g), \zeta(g)]$ is sufficient to completely determine the momentum dependence of all $n$-point functions $\Gamma^{(n)}(p_i, g, h, \Lambda)$.

One can integrate the $\beta$-function equation in Eq. (3.14) to obtain the renormalization group invariant quantity

$$\xi^{-1}(g) = m(g) = \text{const.}\ \Lambda \exp\left(-\int^{g} \frac{dg'}{\beta(g')}\right) \ , \tag{3.34}$$

which is identified with the correlation length appearing, for example, in Eq. (3.25). The multiplicative constant in front of the expression on the r.h.s. arises as an integration constant, and cannot be determined from perturbation theory in $g$. Conversely, it is easy to verify that $\xi$ is indeed a renormalization group invariant, $\Lambda \frac{d}{d\Lambda} \xi[\Lambda, g(\Lambda)] = 0$, as stated previously in Eq. (3.26).

In the vicinity of the fixed point at $g_c$ one can do the integral in Eq. (3.34), using Eq. (3.15) and the resulting linearized expression for the $\beta$-function in the vicinity of the non-trivial ultraviolet fixed point,

$$\beta(g) \underset{g \to g_c}{\sim} \beta'(g_c)\,(g - g_c) + \ldots \tag{3.35}$$

and one finds

$$\xi^{-1}(g) = m(g) \propto \Lambda\, |\, g - g_c\,|^{\nu} \ , \tag{3.36}$$

with a correlation length exponent $\nu = -1/\beta'(g_c) \sim 1/(d-2) + \ldots$. Thus the correlation length $\xi(g)$ diverges as one approaches the fixed point at $g_c$.

In general the existence of a non-trivial ultraviolet fixed point implies that the large momentum behavior above two dimensions is not given by naive perturbation theory; it is given instead by the critical behavior of the renormalized theory. In the weak coupling, small $g$ phase the scale $m$ can be regarded as a crossover scale between the free field behavior at large distance scales and the critical behavior which sets in at large momenta.

In the non-linear $\sigma$-model another quantity of physical interest is the function $M_0(g)$,

$$M_0(g) = \exp\left[-\tfrac{1}{2}\int_0^{g} dg'\, \frac{\zeta(g')}{\beta(g')}\right] \ , \tag{3.37}$$

which is proportional to the order parameter (the magnetization) of the non-linear $\sigma$-model. To one-loop order one finds $\zeta(g) = \frac{1}{2\pi}(N-1)g + \ldots$ and therefore

$$M_0(g) = \text{const.}\,(g_c - g)^{\beta} \ , \tag{3.38}$$

with $\beta = \frac{1}{2}\nu(d-2+\eta)$ and $\eta = \zeta(g_c) - \varepsilon$. To leading order in the $\varepsilon$ expansion one has for the anomalous dimension of the $\pi$ field $\eta = \varepsilon/(N-2) + O(\varepsilon^2)$. In gauge theories, including gravity, there is no local order parameter, so this quantity has no obvious generalization there.

In general the $\varepsilon$-expansion is only expected to be asymptotic. This is already seen from the expansion for $\nu$ which has recently been computed to four loops (Hikami and Brezin, 1978; Bernreuther and Wegner, 1986; Kleinert, 2000)

$$\begin{aligned}\nu^{-1} = \; & \varepsilon + \frac{\varepsilon^2}{N-2} + \frac{\varepsilon^3}{2(N-2)} \\ & - [30 - 14N + N^2 + (54 - 18N)\zeta(3)] \\ & \times \frac{\varepsilon^4}{4(N-2)^3} + \dots \end{aligned} \tag{3.39}$$

which needs to be summed by Borel-Padé methods to obtain reliable results in three dimensions. For example, for $N = 3$ one finds in three dimensions $\nu \approx 0.799$, which can be compared to the $4 - \varepsilon$ result for the $\lambda\phi^4$ theory to five loops $\nu \simeq 0.705$, to the seven-loop perturbative expansion for the $\lambda\phi^4$ theory directly in $3d$ which gives $\nu \simeq 0.707$, with the high temperature series result $\nu \simeq 0.717$ and the Monte Carlo estimates $\nu \simeq 0.718$, as compiled for example in a recent comprehensive review (Guida and Zinn-Justin, 1998).

There exist standard methods to deal with asymptotic series such as the one in Eq. (3.39). To this purpose one considers a general series

$$f(g) = \sum_{n=0}^{\infty} f_n g^n \quad , \tag{3.40}$$

and defines its Borel transform as

$$F(b) = \sum_{n=0}^{\infty} \frac{f_n}{n!} b^n \quad . \tag{3.41}$$

One can attempt to sum the series for $F(b)$ using Padé methods and conformal transformations. The original function $f(g)$ is then recovered by performing an integral over the Borel transform variable $b$

$$f(g) = \tfrac{1}{g} \int_0^\infty db\, e^{-b/g} F(b) \quad , \tag{3.42}$$

where the familiar formula

$$\int_0^\infty dz\, z^n\, e^{-z/g} = n!\, g^{n+1} \quad , \tag{3.43}$$

has been used. Bounds on the coefficients $f_n$ suggest that in most cases $F(z)$ is analytic in a circle of radius $a$ around the origin, and that the integral will converge for $|z|$ small enough, within a sector $|\arg z| < \alpha/2$ with typically $\alpha \geq \pi$ (Le Guillou and Zinn-Justin, 1980).

The first singularity along the positive real axis is generally referred to as an infrared renormalon, and is expected to be, in the $2d$ non-linear $\sigma$-model, at $b = 1/2\beta_0$ where $\beta_0 = (N-2)/2\pi$, and gives rise to non-analytic corrections of order $\exp(-2\pi/(N-2)g)$ (David, 1982). Such non-analytic contributions presumably account for the fact (Cardy and Hamber, 1980) that the $N = 2$ model has a vanishing $\beta$-function to all orders in $d = 2$, and yet has non-trivial finite exponents in $d = 3$, in spite of the result of Eq. (3.39). Indeed the $2+\varepsilon$ expansion is not particularly useful for the special case of the $N = 2$ $\sigma$-model. Then the action is simply given by

$$S(\theta) = \frac{\Lambda^{d-2}}{2g} \int d^d x \left[ \partial_\mu \theta(x) \right]^2 \ , \tag{3.44}$$

with $\phi_1(x) = \sin\theta(x)$ and $\phi_2(x) = \cos\theta(x)$, describing the fluctuations of a planar spin in $d$ dimensions. The $\beta$-function of Eq. (3.14) then vanishes identically in $d = 2$, and the corrections to $\nu$ diverge for $d > 2$, as in Eq. (3.39). Yet this appears to be more a pathology of the perturbative expansion in $\varepsilon$, since after all the lattice model of Eq. (3.11) is still well defined, and so is the $4-\varepsilon$ expansion for the continuum linear $O(N)$ $\sigma$-model. Thus, in spite of the model being again not perturbatively renormalizable in $d = 3$, one can still develop, for these models, the full machinery of the renormalization group and compute the relevant critical exponents.

Perhaps more importantly, a recent space shuttle experiment (Lipa et al, 2003) has succeeded in measuring the specific heat exponent $\alpha = 2 - 3\nu$ of superfluid Helium (which is supposed to share the same universality class as the $N = 2$ non-linear $\sigma$-model, with the complex phase of the superfluid condensate acting as the order parameter) to very high accuracy

$$\alpha = -0.0127(3) \ . \tag{3.45}$$

Previous theoretical predictions for the $N = 2$ model include the most recent four-loop $4-\varepsilon$ continuum result $\alpha = -0.01126(10)$ (Kleinert, 2000), a recent lattice Monte Carlo estimate $\alpha = -0.0146(8)$ (Campostrini et al, 2001), and the lattice variational renormalization group prediction $\alpha = -0.0125(39)$.

Perhaps the message one gains from this rather lengthy discussion of the non-linear $\sigma$-model in $d > 2$ is that:

- The model provides a specific example of a theory which is not perturbatively renormalizable in the traditional sense, and for which the naive perturbative expansion in fixed dimension leads to uncontrollable divergences and inconsistent results;
- Yet the model can be constructed perturbatively in terms of a double expansion in $g$ and $\varepsilon = d - 2$. This new perturbative expansion, combined with the renormalization group, in the end provides explicit and detailed information about universal scaling properties of the theory in the vicinity of the non-trivial ultraviolet point at $g_c$;
- The continuum field theory predictions obtained this way generally agree, for distances much larger than the cutoff scale, with lattice results, and, perhaps more

importantly, with high precision experiments on systems belonging to the same universality class of the $O(N)$ model.

## 3.3 Non-linear Sigma Model in the Large-$N$ Limit

A rather fortunate circumstance is represented by the fact that in the large $N$ limit the non-linear $\sigma$-model can be solved exactly. This allows an independent verification of the correctness of the general ideas developed in the previous section, as well as a direct comparison of explicit results for universal quantities. The starting point is the functional integral of Eq. (3.1),

$$Z = \int [d\phi(x)] \prod_x \delta\left[\phi^2(x) - 1\right] \exp[-S(\phi)] \tag{3.46}$$

with

$$S(\phi) = \frac{1}{2T} \int d^d x\, \partial_\mu \phi(x) \cdot \partial_\mu \phi(x) \ . \tag{3.47}$$

The constraint on the $\phi$ field can be implemented via an auxiliary Lagrange multiplier field $\alpha(x)$. One writes

$$Z = \int [d\phi(x)]\,[d\alpha(x)] \exp[-S(\phi,\alpha)] \tag{3.48}$$

with

$$S(\phi,\alpha) = \frac{1}{2T} \int d^d x \left[ [\partial_\mu \phi(x)]^2 + \alpha(x)(\phi^2(x) - 1) \right] \ . \tag{3.49}$$

Since the action is now quadratic in $\phi(x)$ one can integrate over $N-1$ $\phi$-fields (denoted previously by $\pi$). The resulting determinant is then re-exponentiated, and one is left with a functional integral over the remaining first field $\phi_1(x) \equiv \sigma(x)$, as well as the Lagrange multiplier field $\alpha(x)$,

$$Z = \int [d\sigma(x)\, d\alpha(x)] \exp[-S_N(\phi,\alpha)] \tag{3.50}$$

with now

$$\begin{aligned} S_N(\phi,\alpha) = {} & \frac{1}{2T} \int d^d x \left[ (\partial_\mu \sigma)^2 + \alpha(\sigma^2 - 1) \right] \\ & + \tfrac{1}{2}(N-1)\,\mathrm{tr}\ln[-\partial^2 + \alpha] \ . \end{aligned} \tag{3.51}$$

In the large $N$ limit one can neglect, to leading order, fluctuations in the $\alpha$ and $\sigma$ fields. For a constant $\alpha$ field, $<\alpha(x)>= m^2$, the last (trace) term can be written in momentum space as

$$\tfrac{1}{2}(N-1)\int^{\Lambda}\frac{d^dk}{(2\pi)^d}\ln(k^2+m^2)\ , \tag{3.52}$$

which makes the evaluation of the trace straightforward. As should be clear from Eq. (3.49), the parameter $m$ can be interpreted as the mass of the $\phi$ field. The functional integral in Eq. (3.50) can then be evaluated by the saddle point method. It is easy to see from Eq. (3.51) that the saddle point conditions are

$$\begin{aligned}\sigma^2 &= 1-(N-1)\Omega_d(m)T\ ,\\ m^2\sigma &= 0\end{aligned} \tag{3.53}$$

with the function $\Omega_d(m)$ given by the integral

$$\Omega_d = \int^{\Lambda}\frac{d^dk}{(2\pi)^d}\frac{1}{k^2+m^2}\ . \tag{3.54}$$

The latter can be evaluated in terms of a hypergeometric function,

$$\Omega_d = \frac{1}{2^{d-1}\pi^{d/2}\Gamma(d/2)}\frac{\Lambda^d}{m^2d}\ {}_2F_1\left[1,\frac{d}{2};1+\frac{d}{2};-\frac{\Lambda^2}{m^2}\right]\ , \tag{3.55}$$

but here one only really needs it in the large cutoff limit, $m \ll \Lambda$, in which case one finds the more tractable expression

$$\Omega_d(m)-\Omega_d(0) = m^2[c_1m^{d-4}+c_2\Lambda^{d-4}+O(m^2\Lambda^{d-6})]\ , \tag{3.56}$$

with $c_1$ and $c_2$ some $d$-dependent coefficients.

From Eq. (3.53) one notices that at weak coupling and for $d>2$ a non-vanishing $\sigma$-field expectation value implies that $m$, the mass of the $\pi$ field, is zero. If one sets $(N-1)\Omega_d(0)=1/T_c$, one can then write the first expression in Eq. (3.53) as

$$\sigma(T) = \pm[1-T/T_c]^{1/2}\ , \tag{3.57}$$

which shows that $T_c$ is the critical coupling at which the order parameter $\sigma$ vanishes.

Above $T_c$ the order parameter $\sigma$ vanishes, and $m(T)$ is obtained, from Eq. (3.53), by the solution of the nonlinear gap equation

$$\frac{1}{T} = (N-1)\int^{\Lambda}\frac{d^dk}{(2\pi)^d}\frac{1}{k^2+m^2}\ . \tag{3.58}$$

Using the definition of the critical coupling $T_c$, one can now write, for $2<d<4$, for the common mass of the $\sigma$ and $\pi$ fields

$$m(T)\underset{m\ll\Lambda}{\sim}\left(\frac{1}{T_c}-\frac{1}{T}\right)^{1/(d-2)}\ , \tag{3.59}$$

which gives for the correlation length exponent the non-gaussian value $\nu = 1/(d-2)$, with the gaussian value $\nu = 1/2$ being recovered as expected at $d = 4$ (Wilson and Fisher, 1972). Note that in the large $N$ limit the constant of proportionality in Eq. (3.59) is completely determined by the explicit expression for $\Omega_d(m)$.

Perhaps one of the most striking aspects of the non-linear sigma model above two dimensions is that all particles are massless in perturbation theory, yet they all become massive in the strong coupling phase $T > T_c$, with masses proportional to the non-perturbative scale $m$.

Again one can perform a renormalization group analysis as was done in the previous section in the context of the $2+\varepsilon$ expansion. To this end one defines dimensionless coupling constants $g = \Lambda^{d-2} T$ and $g_c = \Lambda^{d-2} T_c$ as was done in Eq. (3.1). Then the non-perturbative result of Eq. (3.59) becomes

$$m(g) \simeq c_d \cdot \Lambda \left( \frac{1}{g_c} - \frac{1}{g} \right)^{1/(d-2)} , \tag{3.60}$$

with the numerical coefficient given by $c_d = [\frac{1}{2}(d-2)\pi |\csc\left(\frac{d\pi}{2}\right)|]^{\frac{1}{d-2}}$. One welcome feature of this large-$N$ result is the fact that it provides an explicit value for the coefficient in Eq. (3.29), namely

$$c_d = \left( \frac{g_c}{a_0} \right)^{1/(d-2)} , \tag{3.61}$$

and thereby for the numerical factor $a_0$ appearing in Eqs. (3.29) and (3.22).

Again the physical, dimensionful mass $m$ in Eqs. (3.59) or (3.60) is required to be scale- and cutoff-independent as in Eq. (3.26)

$$\Lambda \frac{d}{d\Lambda} m[\Lambda, g(\Lambda)] = 0 , \tag{3.62}$$

or, more explicitly, using the expression for $m$ in Eq. (3.60),

$$\left[ \Lambda \frac{\partial}{\partial \Lambda} + \beta(g) \frac{\partial}{\partial g} \right] \Lambda \, (\frac{1}{g_c} - \frac{1}{g})^{1/(d-2)} = 0 , \tag{3.63}$$

which implies for the $O(N)$ $\beta$-function in the large $N$ limit the simple result

$$\beta(g) = (d-2)\, g\, (1 - g/g_c) . \tag{3.64}$$

The latter is valid again in the vicinity of the fixed point at $g_c$, due to the assumption, used in Eq. (3.59), of $m \ll \Lambda$. Note that it vanishes in $d = 2$, and for $g = 0$, in agreement with the $2+\varepsilon$ result of Eq. (3.14). Furthermore Eq. (3.64) gives the momentum dependence of the coupling at fixed cutoff. After integration, one finds for the momentum ($\mu$) dependence of the coupling at fixed cutoff $\Lambda$

$$\frac{g(\mu)}{g_c} = \frac{1}{1 - c\,(\mu_0/\mu)^{d-2}} \approx 1 + c\,(\mu_0/\mu)^{d-2} + \ldots \tag{3.65}$$

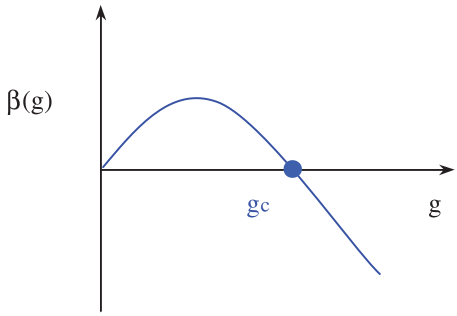



**Fig. 3.3** The $\beta$-function for the non-linear $\sigma$-model in the large-$N$ limit for $d > 2$.

with $c\mu_0^{d-2}$ the integration constant. The sign of $c$ then depends on whether one is on the right ($c > 0$) or on the left ($c < 0$) of the ultraviolet fixed point at $g_c$.

One notices therefore again that the general shape of $\beta(g)$ is of the type shown in Fig. 3.3, with $g_c$ a stable non-trivial UV fixed point, and $g = 0$ and $g = \infty$ two stable (trivial) IR fixed points. Once more, at the critical point $g_c$ the $\beta$-function vanishes and the theory becomes scale invariant. Furthermore one can check that again $\nu = -1/\beta'(g_c)$ where $\nu$ is the exponent in Eq. (3.59). As before one can re-write the physical mass $m$ for $2 < d < 4$ as

$$\xi^{-1}(g) = m(g) \propto \Lambda \exp\left(-\int^{g} \frac{dg'}{\beta(g')}\right) \ , \tag{3.66}$$

as was done previously in Eq. (3.34).

Another general lesson one learns is that Eq. (3.62),

$$\left[\Lambda \frac{\partial}{\partial \Lambda} + \beta(g)\frac{\partial}{\partial g}\right] m[\Lambda, g(\Lambda)] = 0 \ , \tag{3.67}$$

can be used to provide a non-perturbative definition for the $\beta$-function $\beta(g)$. If one sets $m = \Lambda F(g)$, with $F(g)$ a dimensionless function of $g$, then one has the simple result

$$\beta(g) = -\frac{F(g)}{F'(g)} \ . \tag{3.68}$$

Thus the knowledge of the dependence of the mass gap $m$ on the bare coupling $g$ fixes the shape of the $\beta$ function, at least in the vicinity of the fixed point. It should be clear then that the definition of the $\beta$-function per se, and therefore the scale dependence of $g(\mu)$ which follows from it [as determined from the solution of the differential equation $\mu \frac{\partial g}{\partial \mu} = \beta(g(\mu))$] is *not* necessarily tied to perturbation theory.

When $N$ is large but finite, one can develop a systematic $1/N$ expansion in order to evaluate the corrections to the picture presented above (Zinn-Justin, 2002). Corrections to the exponents are known up to order $1/N^2$, but the expressions are rather complicated for arbitrary $d$ and will not be reproduced here. In general it appears that the $1/N$ expansion is only asymptotic, and somewhat slowly convergent for useful values of $N$ in three dimensions.

## 3.4 Self-coupled Fermion Model

The non-linear sigma model is not an isolated example of a field theory that is not perturbatively renormalizable above two dimensions, although it is certainly by far the most thoroghly explored one. Here it seems worthwhile to mention a second example of a theory which naively is not perturbatively renormalizable in $d > 2$, and yet whose critical properties can again be worked out both in the $2+\varepsilon$ expansion, and in the large $N$ limit. It is described by an $U(N)$-invariant action containing a set of $N$ massless self-coupled Dirac fermions (Wilson, 1973; Gross and Neveu, 1974)

$$S(\psi,\bar\psi) = -\int d^d x [\bar\psi \cdot \partial\!\!\!/ \, \psi + \tfrac{1}{2}\Lambda^{d-2} u (\bar\psi \cdot \psi)^2] \ . \tag{3.69}$$

In even dimensions the discrete chiral symmetry $\psi \to \gamma_5 \psi$, $\bar\psi \to -\bar\psi\gamma_5$ prevents the appearance of a fermion mass term. Interest in the model resides in the fact that it exhibits a mechanism for dynamical mass generation and chiral symmetry breaking.

In two dimensions the fermion self-coupling constant is dimensionless, and after setting $d = 2+\varepsilon$ one is again ready to develop the full machinery of the perturbative expansion in $u$ and $\varepsilon$, as was done for the non-linear $\sigma$-model, since the model is again believed to be multiplicatively renormalizable in the framework of the $2+\varepsilon$ expansion. For the $\beta$-function one finds to three loops

$$\beta(u) = \varepsilon u - \frac{\bar N - 2}{2\pi} u^2 + \frac{\bar N - 2}{4\pi^2} u^3 + \frac{(\bar N - 2)(\bar N - 7)}{32\pi^3} u^4 + \dots \tag{3.70}$$

with the parameter $\bar N = N \operatorname{tr} \mathbf{1}$, where the last quantity is the identity matrix in the $\gamma$-matrix algebra. In two dimensions $\bar N = 2N$ and the model is asymptotically free; for $\bar N = 2$ the interaction is proportional to the Thirring one and the $\beta$-function vanishes identically.

As for the case of the non-linear $\sigma$-model, the solution of the renormalization group equations involves an invariant scale, which can be obtained (up to a constant which cannot be determined from perturbation theory alone) by integrating Eq. (3.70)

$$\xi^{-1}(u) = m(u) = \text{const.} \Lambda \exp\left[ -\int^u \frac{du'}{\beta(u')} \right] \ . \tag{3.71}$$

In two dimensions this scale is, to lowest order in $u$, proportional to

$$m(u) \underset{u \to 0}{\sim} \Lambda \exp\left[ -\frac{2\pi}{(\bar N - 2) u} \right] \ , \tag{3.72}$$

and thus non-analytic in the bare coupling $u$. Above two dimensions a non-trivial ultraviolet fixed point appears at

$$u_c = \frac{2\pi}{\bar N - 2} \varepsilon + \frac{2\pi}{(\bar N - 2)^2} \varepsilon^2 + \frac{(\bar N + 1)\pi}{2(\bar N - 2)^3} \varepsilon^3 + \dots \tag{3.73}$$

In the weak coupling phase $u < u_c$ the fermions stay massless and chiral symmetry is unbroken, whereas in the strong coupling phase $u > u_c$ (which is the only phase present in $d = 2$) chiral symmetry is broken, a fermion condensate arises and a non-perturbative fermion mass is generated. In the vicinity of the ultraviolet fixed point one has for the mass gap

$$m(u) \underset{u \to u_c}{\sim} \Lambda \, (u - u_c)^{\nu} \quad , \tag{3.74}$$

up to a constant of proportionality, with the exponent $\nu$ given by

$$\nu^{-1} \equiv -\beta'(u_c) = \varepsilon - \frac{\varepsilon^2}{\bar{N}-2} - \frac{(\bar{N}-3)\pi}{2(\bar{N}-2)^2}\,\varepsilon^3 + \dots \tag{3.75}$$

The rest of the analysis proceeds in a way that, at least formally, is virtually identical to the non-linear $\sigma$-model case. It need not be repeated here, as one can just take over the relevant formulas for the renormalization group behavior of $n$-point functions, for the running of the couplings, etc.

The existence of a non-trivial ultraviolet fixed point implies that the large momentum behavior above two dimensions is not given by naive perturbation theory; it is given instead by the critical behavior of the renormalized theory, in accordance with Eq. (3.70). In the weak coupling, small $u$ phase the scale $m$ can be regarded as a crossover scale between the free field behavior at large distance scales and the critical behavior which sets in at large momenta.

Finally, the same model can be solved exactly in the large $N$ limit. There too one can show that the model is characterized by two phases, a weak coupling phase where the fermions are massless and a strong coupling phase in which a chiral symmetry is spontaneously broken.

## 3.5 The Gravitational Case

In two dimensions the gravitational coupling becomes dimensionless, $G \sim \Lambda^{2-d}$, and the theory appears perturbatively renormalizable. In spite of the fact that the gravitational action reduces to a topological invariant in two dimensions, it would seem meaningful to try to construct, in analogy to what was suggested originally for scalar field theories (Wilson, 1973), the theory perturbatively as a double series in $\varepsilon = d - 2$ and $G$.

One first notices though that in pure Einstein gravity, with Lagrangian density

$$\mathscr{L} = -\frac{1}{16\pi G_0}\sqrt{g}R \quad , \tag{3.76}$$

the bare coupling $G_0$ can be completely reabsorbed by a field redefinition

$$g_{\mu\nu} = \omega\, g'_{\mu\nu} \quad , \tag{3.77}$$

with $\omega$ is a constant, and thus the renormalization properties of $G_0$ have no physical meaning for this theory. This simply follows from the fact that $\sqrt{g}R$ is homogeneous in $g_{\mu\nu}$, which is quite different from the Yang-Mills case. The situation changes though when one introduces a second dimensionful quantity to compare to. In the pure gravity case this contribution is naturally supplied by the cosmological constant term proportional to $\lambda_0$,

$$\mathscr{L} = -\frac{1}{16\pi G_0}\sqrt{g}R + \lambda_0\sqrt{g} \ . \tag{3.78}$$

Under a rescaling of the metric as in Eq. (3.77) one has

$$\mathscr{L} = -\frac{1}{16\pi G_0}\,\omega^{d/2-1}\sqrt{g'}R' + \lambda_0\,\omega^{d/2}\sqrt{g'} \ , \tag{3.79}$$

which is interpreted as a rescaling of the two bare couplings

$$G_0 \to \omega^{-d/2+1} G_0 \ , \quad \lambda_0 \to \lambda_0\,\omega^{d/2} \ , \tag{3.80}$$

leaving the dimensionless combination $G_0^d \lambda_0^{d-2}$ unchanged. Therefore only the latter combination has physical meaning in pure gravity. In particular, one can always choose the scale $\omega = \lambda_0^{-2/d}$ so as to adjust the volume term to have a unit coefficient. More importantly, it is physically meaningless to discuss separately the renormalization properties of $G_0$ and $\lambda_0$, as they are individually gauge-dependent in the sense just illustrated. These arguments should clarify why in the following it will be sufficient at the end to just focus on the renormalization properties of one coupling, such as Newton's constant $G_0$.

In general it is possible at least in principle to define quantum gravity in any $d > 2$ (Weinberg, 1979). There are $d(d+1)/2$ independent components of the metric in $d$ dimensions, and the same number of algebraically independent components of the Ricci tensor appearing in the field equations. The contracted Bianchi identities reduce the count by $d$, and so does general coordinate invariance, leaving $d(d-3)/2$ physical gravitational degrees of freedom in $d$ dimensions. At the same time, four space-time dimensions is known to be the lowest dimension for which Ricci flatness does not imply the vanishing of the gravitational field, $R_{\mu\nu\lambda\sigma} = 0$, and therefore the first dimension to allow for gravitational waves and their quantum counterparts, gravitons.

In a general dimension the position space tree-level graviton propagator of the linearized theory, given in $k$-space in Eq. (1.77), can be obtained by Fourier transform and is proportional to

$$\int d^d k\,\frac{1}{k^2}\,e^{ik\cdot x} \;=\; \frac{\Gamma\left(\frac{d-2}{2}\right)}{4\,\pi^{d/2}\,(x^2)^{d/2-1}} \ . \tag{3.81}$$

The static gravitational potential is then proportional to the spatial Fourier transform

$$V(r) \propto \int d^{d-1}\mathbf{k}\,\frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{\mathbf{k}^2} \sim \frac{1}{r^{d-3}} \ , \tag{3.82}$$

and can be shown to vanish in $d=3$. To show this one needs to compute the analog of Eq. (1.17) in $d$ dimensions, which is

$$-\frac{\kappa^2}{2}\int d^dx\left[T_{\mu\nu}\,\Box^{-1}\,T^{\mu\nu}\;-\;(d-2)^{-1}\,T_\mu{}^\mu\,\Box^{-1}\,T_\nu{}^\nu\right]$$
$$\rightarrow -\frac{d-3}{d-2}\,\frac{\kappa^2}{2}\int d^{d-1}x\,T^{00}\,\mathscr{G}\,T^{00}\;, \qquad (3.83)$$

where the Green's function $\mathscr{G}$ is the static limit of $1/\Box$, and $\kappa^2=16\pi G$. The above result implies that there are no Newtonian forces in $d$=2+1 dimensions (Deser, Jackiw and 't Hooft, 1984; Deser and Jackiw, 1984). The only fluctuations left in 3$d$ are possibly associated with the scalar curvature (Deser, Jackiw and Templeton, 1982).

The $2+\varepsilon$ expansion for pure gravity then proceeds as follows. First the gravitational part of the action

$$\mathscr{L}=-\frac{\mu^\varepsilon}{16\pi G}\sqrt{g}\,R\;, \qquad (3.84)$$

with $G$ dimensionless and $\mu$ an arbitrary momentum scale, is expanded by setting

$$g_{\mu\nu}\;\rightarrow\;\bar{g}_{\mu\nu}=g_{\mu\nu}+h_{\mu\nu}\;, \qquad (3.85)$$

where $g_{\mu\nu}$ is the classical background field and $h_{\mu\nu}$ the small quantum fluctuation. The quantity $\mathscr{L}$ in Eq. (3.84) is naturally identified with the bare Lagrangian, and the scale $\mu$ with a microscopic ultraviolet cutoff $\Lambda$, the inverse lattice spacing in a lattice formulation. Since the resulting perturbative expansion is generally reduced to the evaluation of Gaussian integrals, the original constraint (in the Euclidean theory)

$$\det g_{\mu\nu}(x)\,>0\;, \qquad (3.86)$$

is no longer enforced [the same is not true in the lattice regulated theory, where it plays an important role, see the discussion following Eq. (6.70)].

A gauge fixing term needs to be added, in the form of a background harmonic gauge condition,

$$\mathscr{L}_{gf}=\tfrac{1}{2}\alpha\sqrt{g}\,g_{\nu\rho}\left(\nabla_\mu h^{\mu\nu}-\tfrac{1}{2}\beta g^{\mu\nu}\nabla_\mu h\right)\left(\nabla_\lambda h^{\lambda\rho}-\tfrac{1}{2}\beta g^{\lambda\rho}\nabla_\lambda h\right)\;, \qquad (3.87)$$

with $h^{\mu\nu}=g^{\mu\alpha}g^{\nu\beta}h_{\alpha\beta}$, $h=g^{\mu\nu}h_{\mu\nu}$ and $\nabla_\mu$ the covariant derivative with respect to the background metric $g_{\mu\nu}$. The gauge fixing term also gives rise to a Faddeev-Popov ghost contribution $\mathscr{L}_{ghost}$ containing the ghost field $\psi_\mu$, so that the total Lagrangian becomes $\mathscr{L}+\mathscr{L}_{gf}+\mathscr{L}_{ghost}$.

In a flat background, $g^{\mu\nu}=\delta^{\mu\nu}$, one obtains from the quadratic part of the Lagrangian of Eqs. (3.84) and (3.87) the following expression for the graviton propagator

$$<h_{\mu\nu}(k)h_{\alpha\beta}(-k)>=$$

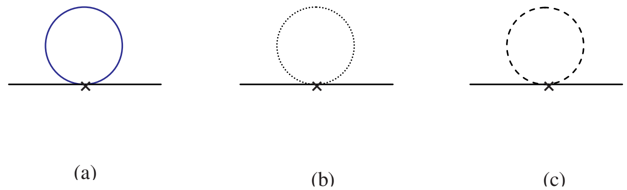



**Fig. 3.4** One-loop diagrams giving rise to coupling renormalizations in gravity. From *left* to *right*, graviton loop, ghost loop and scalar matter loop.

$$
\begin{aligned}
&\frac{1}{k^2}(\delta_{\mu\alpha}\delta_{\nu\beta}+\delta_{\mu\beta}\delta_{\nu\alpha})-\frac{2}{d-2}\frac{1}{k^2}\delta_{\mu\nu}\delta_{\alpha\beta}\\
&-\left(1-\frac{1}{\alpha}\right)\frac{1}{k^4}(\delta_{\mu\alpha}k_\nu k_\beta+\delta_{\nu\alpha}k_\mu k_\beta+\delta_{\mu\beta}k_\nu k_\alpha+\delta_{\nu\beta}k_\mu k_\alpha)\\
&+\frac{1}{d-2}\frac{4(\beta-1)}{\beta-2}\frac{1}{k^4}(\delta_{\mu\nu}k_\alpha k_\beta+\delta_{\alpha\beta}k_\mu k_\nu)\\
&+\frac{4(1-\beta)}{(\beta-2)^2}\left[2-\frac{3-\beta}{\alpha}-\frac{2(1-\beta)}{d-2}\right]\frac{1}{k^6}k_\nu k_\nu k_\alpha k_\beta \ .
\end{aligned}
\tag{3.88}
$$

Normally it would be convenient to choose a gauge $\alpha$=$\beta = 1$, in which case only the first two terms for the graviton propagator survive. But here it might be advantageous to leave the two gauge parameters unspecified, so that one can later show explicitly the gauge independence of the final result. In particular the gauge parameter $\beta$ is related to the gauge freedom associated with the possibility, described above, of rescaling the metric $g_{\mu\nu}$. Note also the presence of kinematical poles in $\varepsilon = d-2$ in the second, fourth and fifth term for the graviton propagator.

To illustrate explicitly the mechanism of coupling renormalization, the cosmological term will be discussed first, since the procedure is a bit simpler. The cosmological term $\sqrt{g}$ is first expanded by setting $g_{\mu\nu} = g_{\mu\nu} + h_{\mu\nu}$ with a flat background $g_{\mu\nu} = \delta_{\mu\nu}$. One has

$$\sqrt{g} = 1 + \tfrac{1}{2}h - \tfrac{1}{4}h_{\mu\nu}h^{\mu\nu} + \tfrac{1}{8}h^2 + O(h^3) \ , \tag{3.89}$$

with $h = h^{\mu}{}_{\mu}$. Terms linear in the fluctuation $h_{\mu\nu}$ are dropped, since in a properly chosen background such terms are expected to be absent. The one-loop correction to the 1 term in the above expression is then given by the tadpole diagrams for the two quadratic terms,

$$-\tfrac{1}{4}h_{\mu\nu}h^{\mu\nu} + \tfrac{1}{8}h^2 \rightarrow -\tfrac{1}{4} < h_{\mu\nu}h^{\mu\nu} > + \tfrac{1}{8} < h^2 > \ . \tag{3.90}$$

These are easily evaluated using the graviton propagator of Eq. (3.88). For the one loop divergences (see Fig. 3.4) associated with the $\sqrt{g}$ term one then obtains

$$\lambda_0 \to \lambda_0 \left[ 1 - \left( \frac{a_1}{\varepsilon} + \frac{a_2}{\varepsilon^2} \right) G \right] \tag{3.91}$$

with coefficients

$$\begin{aligned} a_1 &= -\frac{8}{\alpha} + 8\frac{(\beta-1)^2}{(\beta-2)^2} + 4\frac{(\beta-1)(\beta-3)}{\alpha(\beta-2)^2} \\ a_2 &= 8\frac{(\beta-1)^2}{(\beta-2)^2} \ . \end{aligned} \tag{3.92}$$

One notices that the kinematic singularities in the graviton propagator, proportional to $1/(d-2)$, can combine with the one loop ultraviolet divergent part of momentum integrals, as in

$$\frac{1}{\varepsilon} \int \frac{d^d k}{(2\pi)^d} \frac{1}{k^2} \sim \frac{1}{\varepsilon^2} \ , \tag{3.93}$$

to give terms of order $1/\varepsilon^2$ in Eq. (3.91). Generally it is better to separate the ultraviolet divergence from the infrared one, by using for example the following regulated integral

$$\int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2+\mu^2)^a} = \frac{1}{(4\pi)^{d/2}} \frac{\Gamma(a-d/2)}{\Gamma(a)} (\mu^2)^{d/2-a} \ , \tag{3.94}$$

for $a=1$ and $\mu \to 0$.

One can then follow the same procedure for the $\sqrt{g}R$ term. First one needs to expand the Einstein term to quadratic order in the quantum field $h_{\mu\nu}$

$$\begin{aligned} \sqrt{\bar{g}}\bar{R} &= \sqrt{g}R \\ &+ \sqrt{g}\left\{ \tfrac{1}{4}\nabla_\rho h^\mu{}_\nu \nabla^\rho h^\nu{}_\mu - \tfrac{1}{2}\nabla_\nu h^\nu{}_\mu \nabla_\rho h^{\rho\mu} + \tfrac{1}{2}R^\sigma{}_{\rho\mu\nu} h^\rho{}_\sigma h^{\mu\nu} \right\} + \dots \end{aligned} \tag{3.95}$$

where $\nabla_\mu$ denotes the covariant derivative with respect to the background metric $g_{\mu\nu}$. The complete expansion was given previously in Eq. (1.94). The same expansion then needs to be done for the gauge fixing term of Eq. (3.87) as well, and furthermore it is again convenient to choose as a background field the flat metric $g_{\mu\nu} = \delta_{\mu\nu}$. For the one loop divergences associated with the $\sqrt{g}R$ term one then finds

$$\frac{\mu^\varepsilon}{16\pi G} \to \frac{\mu^\varepsilon}{16\pi G} \left( 1 - \frac{b}{\varepsilon} G \right) \ , \tag{3.96}$$

with the coefficient $b$ given by (Gastmans et al, 1978; Christensen et al, 1980)

$$b = \frac{2}{3} \cdot 19 + \frac{4(\beta-1)^2}{(\beta-2)^2} \ . \tag{3.97}$$

Thus the one-loop radiative corrections modify the total Lagrangian to

$$\mathscr{L} \to -\frac{\mu^\varepsilon}{16\pi G} \left( 1 - \frac{b}{\varepsilon} G \right) \sqrt{g}R + \lambda_0 \left[ 1 - \left( \frac{a_1}{\varepsilon} + \frac{a_2}{\varepsilon^2} \right) G \right] \sqrt{g} \ . \tag{3.98}$$

Next one can make use of the freedom to rescale the metric, by setting

$$\left[1-\left(\frac{a_1}{\varepsilon}+\frac{a_2}{\varepsilon^2}\right)G\right]\sqrt{g}=\sqrt{g'} \ , \tag{3.99}$$

which restores the original unit coefficient for the cosmological constant term. The rescaling is achieved by the following field redefinition

$$g_{\mu\nu}=\left[1-\left(\frac{a_1}{\varepsilon}+\frac{a_2}{\varepsilon^2}\right)G\right]^{-2/d} g'_{\mu\nu} \ . \tag{3.100}$$

Hence the cosmological term is brought back into the standard form $\lambda_0\sqrt{g'}$, and one obtains for the complete Lagrangian to first order in $G$

$$\mathscr{L}\to-\frac{\mu^\varepsilon}{16\pi G}\left[1-\frac{1}{\varepsilon}(b-\tfrac{1}{2}a_2)G\right]\sqrt{g'}R'+\lambda_0\sqrt{g'} \ , \tag{3.101}$$

where only terms singular in $\varepsilon$ have been retained. From this last result one can finally read off the renormalization of Newton's constant

$$\frac{1}{G}\to\frac{1}{G}\left[1-\frac{1}{\varepsilon}(b-\tfrac{1}{2}a_2)\,G\right] \ . \tag{3.102}$$

From Eqs. (3.92) and (3.97) one notices that the $a_2$ contribution cancels out the gauge-dependent part of $b$, giving for the remaining contribution $b-\frac{1}{2}a_2=\frac{2}{3}\cdot 19$. Therefore the gauge dependence has, as one would have hoped on physical grounds, disappeared from the final answer. It is easy to see that the same result would have been obtained if the *scaled* cosmological constant $G\lambda_0$ had been held constant, instead of $\lambda_0$ as in Eq. (3.99). One important aspect of the result of Eq. (3.102) is that the quantum correction is negative, meaning that the strength of $G$ is effectively increased by the lowest order radiative correction.

In the presence of an explicit renormalization scale parameter $\mu$ the $\beta$-function for pure gravity is obtained by requiring the independence of the quantity $G_e$ (here identified as an effective coupling constant, with lowest order radiative corrections included) from the original renormalization scale $\mu$,

$$\mu\frac{d}{d\mu}G_e=0$$
$$\frac{1}{G_e}\equiv\frac{\mu^\varepsilon}{G(\mu)}\left[1-\frac{1}{\varepsilon}(b-\tfrac{1}{2}a_2)\,G(\mu)\right] \ . \tag{3.103}$$

To zero-th order in $G$, the renormalization group $\beta$-function entering the renormalization group equation

$$\mu\frac{\partial}{\partial\mu}G=\beta(G) \tag{3.104}$$

is just given by

$$\beta(G)=\varepsilon\,G+\ldots \tag{3.105}$$

The above result just follows from the trivial scale dependence of the classical, dimensionful gravitational coupling: to achieve a fixed given $G_e$, the dimensionless quantity $G(\mu)$ itself has to scale like $\mu^\varepsilon$. Next, to first order in $G$, one has from Eq. (3.103)

$$\mu \frac{\partial}{\partial \mu} G = \beta(G) = \varepsilon G - \beta_0 G^2 + O(G^3, G^2 \varepsilon) \ , \tag{3.106}$$

with $\beta_0 = \frac{2}{3} \cdot 19$. From the procedure outlined above it is clear that $G$ is the only coupling that is scale-dependent in pure gravity. As will be appreciated further below, the importance of the gravitational $\beta$-function $\beta(G)$ lies in the fact that it can be used either to determine the ultraviolet cutoff dependence of the bare coupling needed to keep the effective coupling fixed (as in Eq. (3.103), *or* to determine the momentum dependence of the physical coupling $G(k)$ for a *fixed* cutoff.

Matter fields can be included as well. When $N_S$ scalar fields and $N_F$ Majorana fermion fields are added, the results of Eqs. (3.97), (3.102) and (3.103) are modified to

$$b \to b - \tfrac{2}{3} c \ , \tag{3.107}$$

with $c = N_S + \frac{1}{2} N_F$ (the central charge of the Virasoro algebra in two dimensions), and therefore for the combined $\beta$-function of Eq. (3.106) to one-loop order one has $\beta_0 = \frac{2}{3}(19 - c)$. Of course one noteworthy aspect of the perturbative calculation is the appearance of a non-trivial ultraviolet fixed point at $G_c = (d-2)/\beta_0$ for which $\beta(G_c) = 0$, whose physical significance will be discussed further below.

To check their consistency, the one-loop calculations just described have been repeated by performing a number of natural variations. One modification consists in using the Thirring interaction

$$\mathscr{L} = -\frac{\mu^\varepsilon}{16\pi G} \sqrt{g} R + e \bar{\psi} i \gamma^\mu D_\mu \psi - k e \bar{\psi} \gamma^a \psi \bar{\psi} \gamma_a \psi \ , \tag{3.108}$$

instead of the cosmological constant term to set the scale for the metric. This results in a $\beta$-function still of the form of Eq. (3.106) but with $\beta_0 = \frac{2}{3}(25 - c)$. The slight discrepancy between the two results was initially attributed (Kawai and Ninomiya, 1990) to the well known problems related to the kinematic singularities of the graviton propagator in $2d$ discussed previously. To address this issue, a new perturbative expansion can be performed with the metric parametrized as

$$g_{\mu\nu} \to \bar{g}_{\mu\nu} = g_{\mu\rho} \left(e^h\right)^\rho{}_\nu \, e^{-\phi} \ , \tag{3.109}$$

where the conformal mode $\phi$ [responsible for the kinematic singularity in the second term on the r.h.s. of Eq. (3.88)] is explicitly separated out, and $h^\mu{}_\nu$ is now taken to be traceless, $h^\mu{}_\mu = 0$. Furthermore the conformal mode is made massive by adding a cosmological constant term $\lambda_0 \sqrt{g}$, which again acts as an infrared regulator. Repeating the calculation for the one loop divergences (Kawai, Kitazawa and Ninomiya, 1993a,b) one now finds $\beta_0 = \frac{2}{3}(25-c)$ which is consistent with the above quoted Thirring result.[1]

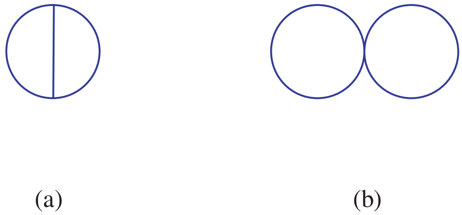



**Fig. 3.5** Two loop graviton diagrams in $2+\varepsilon$ gravity.

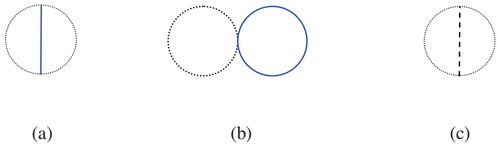



**Fig. 3.6** Two loop graviton-ghost diagrams in $2+\varepsilon$ gravity.

In the meantime the calculations have been laboriously extended to two loops (see Figs. 3.5 and 3.6) (Aida and Kitazawa, 1997), with the result

$$\mu \frac{\partial}{\partial \mu} G = \beta(G) = \varepsilon G - \beta_0 G^2 - \beta_1 G^3 + O(G^4, G^3 \varepsilon, G^2 \varepsilon^2) \ , \tag{3.110}$$

with $\beta_0 = \frac{2}{3}(25-c)$ and $\beta_1 = \frac{20}{3}(25-c)$.

Of some interest is the fact that $\mathcal{N}=1$ supergravity in $2+\varepsilon$ dimensions also seems to give rise to a non-trivial ultraviolet fixed point (Kojima, Sakai and Tanii, 1994). These authors consider a model with a vielbein $e_\mu{}^a$, a Majorana Rarita-Schwinger field $\psi_\mu$ and a real auxiliary scalar field $S$, with indices $a, b, \ldots$ and $\mu, \nu, \ldots$ running from 0 to $d-1$. The action taken to be of the form

$$I_{SG} \ = \ \frac{1}{16\pi G} \int d^d x \det e \left[ R + i \bar{\psi}_\mu \gamma^{\mu\nu\rho} D_\nu \psi_\rho - \frac{d-2}{d-1} S^2 \right] \ , \tag{3.111}$$

with $\gamma^{\mu\nu\rho} \equiv \frac{1}{3!}(\gamma^\mu \gamma^\nu \gamma^\rho \pm$ permutations). There are some subtleties associated with the dimensional reduction of supergravity that will not be discussed here. To lowest

[1] For a while there was considerable uncertainty about the magnitude of the graviton contribution to $\beta_0$, which was quoted originally as $38/3$ (Tsao, 1977), later as $2/3$ (Gastmans et al, 1978; Weinberg, 1977; Christensen and Duff, 1978), and more recently as $50/3$ (Kawai, Kitazawa and Ninomiya, 1993). As discussed in (Weinberg, 1979), the original expectation was that the graviton contribution should be $d(d-3)/2 = -1$ times the scalar contribution close to $d=2$, which would suggest for gravity the value $2/3$. Direct numerical estimates of the scaling exponent $\nu$ in the lattice theory for $d=3$ (Hamber and Williams, 1993) give, using Eq. (3.125), a value $\beta_0 \approx 44/3$ and are therefore in much better agreement with the larger, more recent values.

order in the $\varepsilon$ expansion these authors then find

$$\beta(G) \;=\; \varepsilon\, G - (9-c)\, G^2 + \ldots \tag{3.112}$$

and therefore a non-trivial ultraviolet fixed point close to two dimensions at $G_c = \varepsilon/(9-c)$, where $c$ is the central charge of the superconformal matter multiplet, such as $c$ free scalars and spinor fields. Taken at face value, these result would suggest that, besides ordinary Einstein gravity, also $\mathcal{N}=1$ supergravity could have a non-trivial strong coupling phase, which appears for $G > G_c$.

## 3.6 Phases of Gravity in 2+$\varepsilon$ Dimensions

The gravitational $\beta$-function of Eqs. (3.106) and (3.110) determines the scale dependence of Newton's constant $G$ for $d$ close to two. It has the general shape shown in Fig. 3.7. Because one is left, for the reasons described above, with a single coupling constant in the pure gravity case, the discussion becomes in fact quite similar to the non-linear $\sigma$-model case.

For a qualitative discussion of the physics it will be simpler in the following to just focus on the one loop result of Eq. (3.106); the inclusion of the two-loop correction does not alter the qualitative conclusions by much, as it has the same sign as the lower order, one-loop term. Depending on whether one is on the right ($G > G_c$) or on the left ($G < G_c$) of the non-trivial ultraviolet fixed point at

$$G_c = \frac{d-2}{\beta_0} + O[(d-2)^2] \;, \tag{3.113}$$

(with $G_c$ positive provided one has $c < 25$) the coupling will either flow to increasingly larger values of $G$, or flow towards the Gaussian fixed point at $G = 0$, respectively. In the following we will refer to the two phases as the strong coupling and weak coupling phase, respectively. Perturbatively one only has control on the small

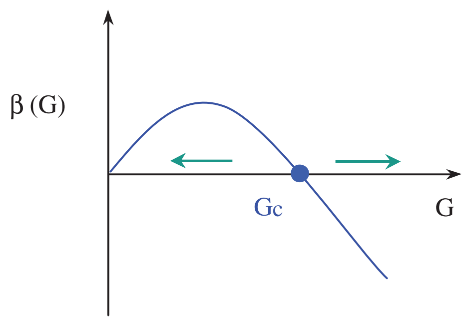



**Fig. 3.7** The renormalization group $\beta$-function for gravity in $2+\varepsilon$ dimensions. The arrows indicate the coupling constant flow as one approaches increasingly larger distance scales.

$G$ regime. When one then sets $d = 2$ only the strong coupling phase survives, so two-dimensional gravity is always strongly coupled within this picture.

The running of $G$ as a function of a sliding momentum scale $\mu = k$ in pure gravity can be obtained by integrating Eq. (3.106), and one finds

$$G(k^2) = \frac{G_c}{1 \pm a_0 \, (m^2/k^2)^{(d-2)/2}} \ , \tag{3.114}$$

with $a_0$ a positive constant and $m$ a mass scale. The choice of $+$ or $-$ sign is determined from whether one is to the left (+), or to right (-) of $G_c$, in which case the effective $G(k^2)$ decreases or, respectively, increases as one flows away from the ultraviolet fixed point towards lower momenta, or larger distances. Physically the two solutions represent a screening ($G < G_c$) and an anti-screening ($G > G_c$) situation. The renormalization group invariant mass scale $\sim m$ arises here as an arbitrary integration constant of the renormalization group equations (one could have absorbed the constant $a_0$ in $m$, but we will not do so here for reasons that will become clearer later).

While in the continuum perturbative calculation both phases, and therefore both signs, seem acceptable, the Euclidean lattice results on the other hand seem to rule out the weak coupling phase as pathological, in the sense that the lattice collapses into a two-dimensional branched polymer, as will be discussed later in Sect. 8.2.

At the opposite end, at energies sufficiently high to become comparable to the ultraviolet cutoff (the inverse lattice spacing in a lattice discretization), the gravitational coupling $G$ flows towards the ultraviolet fixed point,

$$G(k^2) \underset{k^2 \to \Lambda^2}{\sim} G(\Lambda) \ , \tag{3.115}$$

where $G(\Lambda)$ is the coupling at the cutoff scale $\Lambda$, to be identified with the bare (or lattice) coupling. Note that it would seem meaningless to consider, within this framework, momenta which are larger than the ultraviolet cutoff $\Lambda$. At such energies higher dimension operators (such as higher derivative curvature terms) are expected to become important and should therefore be included in the microscopic action.

Note that the result of Eq. (3.114) is quite different from the naive expectation based on straight perturbation theory in $d > 2$ dimensions (where the theory is not perturbatively renormalizable)

$$\frac{G(k^2)}{G} \sim 1 + \text{const.}\, G k^{d-2} + O(G^2) \ , \tag{3.116}$$

which gives a much worse ultraviolet behavior. The existence of a non-trivial ultraviolet fixed point alters the naive picture and drastically improves the ultraviolet behavior.

The $k^2$-dependent contribution in the denominator of Eq. (3.114) is the quantum correction, which at least within a perturbative framework is assumed to be small. In the weak coupling phase $G < G_c$ the coupling then flows towards the origin corresponding to a gravitational screening solution, which sounds a bit odd as one would not expect gravity to be screened. On the other hand the infrared growth of the coupling in the strong coupling phase $G > G_c$ can be written equivalently as

$$G(k^2) \simeq G_c \left[ 1 + a_0 \left( \frac{m^2}{k^2} \right)^{(d-2)/2} + \ldots \right] , \qquad (3.117)$$

where the dots indicate higher order radiative corrections, and which exhibits a number of interesting features. Firstly the fractional power suggests new non-trivial gravitational scaling dimensions, just as in the case of the non-linear $\sigma$-model. Furthermore, the quantum correction involves a new physical, renormalization group invariant scale $\xi = 1/m$ which cannot be fixed perturbatively, and whose size determines the scale for the quantum effects. In terms of the bare coupling $G(\Lambda)$, it is given by

$$m = A_m \cdot \Lambda \, \exp \left( - \int^{G(\Lambda)} \frac{dG'}{\beta(G')} \right) , \qquad (3.118)$$

which just follows from integrating $\mu \frac{\partial}{\partial \mu} G = \beta(G)$ and then setting as the arbitrary scale $\mu \to \Lambda$. Conversely, since $m$ is an invariant, one has $\Lambda \frac{d}{d\Lambda} m = 0$; the running of $G(\mu)$ in accordance with the renormalization group equation of Eq. (3.104) ensures that the l.h.s. is indeed a renormalization group invariant. The constant $A_m$ on the r.h.s. of Eq. (3.118) cannot be determined perturbatively, it needs to be computed by non-perturbative (lattice) methods, for example by evaluating invariant correlations at fixed geodesic distances. It is related to the constant $a_0$ in Eq. (3.117) by $a_0 = 1/(A_m^{1/\nu} G_c)$.

At the fixed point $G = G_c$ the theory is scale invariant by definition. In statistical field theory language the fixed point corresponds to a phase transition, where the correlation length $\xi = 1/m$ diverges and the theory becomes scale (conformally) invariant. In general in the vicinity of the fixed point, for which $\beta(G) = 0$, one can write

$$\beta(G) \underset{G \to G_c}{\sim} \beta'(G_c)\,(G - G_c) + O[(G - G_c)^2] \; . \qquad (3.119)$$

If one then defines the exponent $\nu$ by

$$\beta'(G_c) = -1/\nu \; , \qquad (3.120)$$

then from Eq. (3.118) one has by integration in the vicinity of the fixed point

$$m \underset{G \to G_c}{\sim} \Lambda \cdot A_m \, | G(\Lambda) - G_c |^{\nu} \; , \qquad (3.121)$$

which is why $\nu$ is often referred to as the mass gap exponent. Solving the above equation (with $\Lambda \to k$) for $G(k)$ one obtains back Eq. (3.117), with the constant $a_0$ there related to $A_m$ in Eq. (3.121) by $a_0 = 1/(A_m^{1/\nu} G_c)$ and $\nu = 1/(d-2)$.

That $m$ is a renormalization group invariant is seen from

$$\mu \frac{d}{d\mu} m \equiv \mu \frac{d}{d\mu} \left[ A_m \, \mu \, | G(\mu) - G_c |^{\nu} \right] = 0 \ , \tag{3.122}$$

provided $G$ runs in accordance with Eq. (3.117). To one-loop order one has from Eqs. (3.106) and (3.120) $\nu = 1/(d-2)$. When the bare (lattice) coupling $G(\Lambda) = G_c$ one has achieved criticality, $m = 0$. How far the bare theory is from the critical point is determined by the choice of $G(\Lambda)$, the distance from criticality being measured by the deviation $\Delta G = G(\Lambda) - G_c$.

Furthermore Eq. (3.121) shows how the (lattice) continuum limit is to be taken. In order to reach the continuum limit $a \equiv 1/\Lambda \to 0$ for fixed physical correlation $\xi = 1/m$, the bare coupling $G(\Lambda)$ needs to be tuned so as to approach the ultraviolet fixed point at $G_c$,

$$\Lambda \to \infty \ , \quad m \, \text{fixed} \ , \quad G \to G_c \ . \tag{3.123}$$

The fixed point at $G_c$ thus plays a central role in the cutoff theory: together with the universal scaling exponent $\nu$ it determines the correct unique quantum continuum limit in the presence of an ultraviolet cutoff $\Lambda$. Sometimes it can be convenient to measure all quantities in units of the cutoff and set $\Lambda = 1/a = 1$. In this case the quantity $m$ measured in units of the cutoff (i.e. $m/\Lambda$) has to be tuned to zero in order to construct the lattice continuum limit: for a fixed lattice cutoff, the continuum limit is approached by tuning the bare lattice $G(\Lambda)$ to $G_c$. In other words, the lattice continuum limit has to be taken in the vicinity of the non-trivial ultraviolet point.

The discussion given above is not altered significantly, at least in its qualitative aspects, by the inclusion of the two-loop correction of Eq. (3.110). From the expression for the two-loop $\beta$-function

$$\mu \frac{\partial}{\partial \mu} G = \beta(G) = \varepsilon \, G - \frac{2}{3} \, (25 - c) \, G^2 - \frac{20}{3} \, (25 - c) \, G^3 + \ldots \tag{3.124}$$

for $c$ massless real scalar fields minimally coupled to gravity, one computes the roots $\beta(G_c) = 0$ to obtain the location of the ultraviolet fixed point, and from it on can then determine the universal exponent $\nu = -1/\beta'(G_c)$. One finds

$$\begin{aligned} G_c &= \frac{3}{2\,(25-c)} \, \varepsilon - \frac{45}{2(25-c)^2} \, \varepsilon^2 + \ldots \\ \nu^{-1} &= \varepsilon + \frac{15}{25-c} \, \varepsilon^2 + \ldots \end{aligned} \tag{3.125}$$

which gives, for pure gravity without matter ($c = 0$) in four dimensions, to lowest order $\nu^{-1} = 2$, and $\nu^{-1} \approx 4.4$ at the next order.[2]

Also, in general higher order corrections to the results of the linearized renormalization group equations of Eq. (3.119) are present, which affect the scaling away from the fixed point. Let us assume that close to the ultraviolet fixed point at $G_c$ one can write for the $\beta$-function the following expansion

---

[2] If one does not expand the solution in $\varepsilon$, one finds from the two-loop result $\nu^{-1} = 2d - (1/6)(19 + \sqrt{60d - 95})$ which gives a smaller value $\approx 2.8$ in $d = 4$, as well as rough estimate of the uncertainty.

$$\beta(G) = -\tfrac{1}{\nu}\,(G-G_c) - \omega\,(G-G_c)^2 + \mathcal{O}[(G-G_c)^3] \ . \tag{3.126}$$

After integrating $\mu\frac{\partial}{\partial\mu}G = \beta(G)$ as before, one finds for the structure of the correction to $m$ [see for comparison Eq. (3.121)]

$$\left(\frac{m}{\Lambda}\right)^{1/\nu} = A_m\left[(G(\Lambda)-G_c) - \omega\,\nu\,(G(\Lambda)-G_c)^2 + \ldots\right] \ . \tag{3.127}$$

The hope of course is that these corrections to scaling are small, $(\omega \ll 1)$; in the vicinity of the fixed point the higher order term becomes unimportant when $|G-G_c| \ll 1/(\omega\nu)$. For the effective running coupling one then has

$$\frac{G(\mu)}{G_c} = 1 + a_0\left(\frac{m}{\mu}\right)^{1/\nu} + a_0\,\omega\,\nu\left(\frac{m}{\mu}\right)^{2/\nu} + \mathcal{O}\left(\left(\frac{m}{\mu}\right)^{3/\nu}\right) \ , \tag{3.128}$$

which gives an estimate for the size of the modifications to Eq. (3.117).

Finally, as a word of caution, one should mention that in general the convergence properties of the $2+\varepsilon$ expansion are not well understood. The poor convergence found in some better known cases is usually ascribed to the suspected existence of infrared renormalon-type singularities $\sim e^{-c/G}$ close to two dimensions, and which could possibly arise in gravity as well. At the quantitative level, the results of the $2+\varepsilon$ expansion for gravity therefore remain somewhat limited, and obtaining the three- or four-loop term still represents a daunting task. Nevertheless they provide, through Eqs. (3.117) and (3.121), an analytical insight into the scaling properties of quantum gravity close and above two dimensions, including the suggestion of a non-trivial phase structure and an estimate for the non-trivial universal scaling exponents [Eq. (3.125)]. The key question raised by the perturbative calculations is therefore: what remains of the above phase transition in four dimensions, how are the two phases of gravity characterized there non-perturbatively, and what is the value of the exponent $\nu$ determining the running of $G$ in the vicinity of the fixed point *in four dimensions*.

Finally we should mention that there are other continuum renormalization group methods which can be used to estimate the scaling exponents. An approach which is closely related to the $2+\varepsilon$ expansion for gravity is the derivation of approximate flow equations from the changes of the Legendre effective action with respect to a suitably introduced infrared cutoff $\mu$. The method can be regarded as a variation on Wilson's original momentum slicing technique for obtaining approximate renormalization group equations for lattice couplings. In the simplest case of a scalar field theory (Morris, 1994a,b) one starts from the partition function

$$\exp(W[J]) = \int [d\phi]\,\exp\left\{-\tfrac{1}{2}\phi\cdot C^{-1}\cdot\phi - I_\Lambda[\phi] + J\cdot\phi\right\} \ . \tag{3.129}$$

The $C \equiv C(k,\mu)$ term is taken to be an "additive infrared cutoff term". For it to be an infrared cutoff it needs to be small for $k < \mu$, ideally tending to zero as $k \to 0$, and such that $k^2C(k,\mu)$ is large when $k > \mu$. Since the method is only ultimately applied to the vicinity of the fixed point, for which all physical relevant scales are much smaller than the ultraviolet cutoff $\Lambda$, it is argued that the specific nature of this cutoff is not really relevant in the following. Taking a derivative of $W[J]$ with respect to the scale $\mu$ gives

$$\frac{\partial W[J]}{\partial \mu} = -\tfrac{1}{2}\left[\frac{\delta W}{\delta J}\cdot\frac{\partial C^{-1}}{\partial \mu}\cdot\frac{\delta W}{\delta J} + \mathrm{tr}\left(\frac{\partial C^{-1}}{\partial \mu}\frac{\delta^2 W}{\delta J\,\delta J}\right)\right] \ , \tag{3.130}$$

which can be re-written in terms of the Legendre transform $\Gamma[\phi] = -W[J] - \frac{1}{2}\phi\cdot C^{-1}\cdot\phi + J\cdot\phi$ as

$$\frac{\partial \Gamma[\phi]}{\partial \mu} = -\tfrac{1}{2}\,\mathrm{tr}\left[\frac{1}{C}\frac{\partial C}{\partial \mu}\cdot\left(1 + C\cdot\frac{\delta^2\Gamma}{\delta\phi\,\delta\phi}\right)^{-1}\right] \ , \tag{3.131}$$

where now $\phi = \delta W/\delta J$ is regarded as the classical field. The traces can then be simplified by writing them in momentum space. What remains to be done is first settle on a suitable cutoff function $C(k,\mu)$, and subsequently compute the effective action $\Gamma[\phi]$ in a derivative expansion, thus involving terms of the type $\partial^n\phi^m$, with $\mu$ dependent coefficients.

As far as the cutoff function is concerned, it is first written as $C(k,\mu) = \mu^{\eta-2}C(k^2/\mu^2)$ so as to include the expected anomalous dimensions of the $\phi$ propagator. To simplify things further, it is then assumed for the remaining function of a single variables that $C(q^2) = q^{2p}$ with $p$ a non-negative integer (Morris, 1994a,b). The subsequent derivative expansion gives for example for the $O(N)$ model in $d = 3$ to lowest order $O(\partial^0)$ an anomalous dimensions $\eta = 0$ for all $N$, and $\nu = 0.73$ for $N = 2$. At the next order $O(\partial^2)$ in the derivative expansion the method gives $\nu = 0.65$, compared to the best theoretical and experimental value $\nu = 0.67$ (Morris and Turner, 1998).

In the gravitational case one can proceed in a similar way. First the gravity analog of Eq. (3.131) is clearly

$$\frac{\partial \Gamma[g]}{\partial \mu} = -\tfrac{1}{2}\,\mathrm{tr}\left[\frac{1}{C}\frac{\partial C}{\partial \mu}\cdot\left(1 + C\cdot\frac{\delta^2\Gamma}{\delta g\,\delta g}\right)^{-1}\right] \ , \tag{3.132}$$

where now $g_{\mu\nu} = \delta W/\delta J_{\mu\nu}$ corresponds to the classical metric. The effective action itself contains the Einstein and cosmological terms

$$\Gamma_\mu[g] = -\frac{1}{16\pi G(\mu)}\int d^dx\sqrt{g}\,[R(g) - 2\lambda(\mu)] + \ldots \tag{3.133}$$

as well as gauge fixing and possibly higher derivative terms (Reuter, 1998; Reuter and Saueressig, 2002; Reuter, 2008). After the addition of a background harmonic gauge fixing term with gauge parameter $\alpha$, the choice of a suitable (scalar) cutoff function is required, $C^{-1}(k,\mu) = (\mu^2 - k^2)\theta(\mu^2 - k^2)$ (Litim, 2004), which is inserted into

$$\int [dh]\,\exp\left\{-\tfrac{1}{2}h\cdot C^{-1}\cdot h - I_\Lambda[g] + J\cdot h\right\} \ . \tag{3.134}$$

Note that this added momentum-dependent cutoff term violates both the weak field general coordinate invariance [see for instance Eq. (1.11)], as well as the general rescaling invariance of Eq. (3.79).

The solution of the resulting renormalization group equation for the two couplings $G(\mu)$ and $\lambda(\mu)$ is then truncated to the Einstein and cosmological term, a procedure which is equivalent to the derivative expansion discussed previously. A nontrivial fixed point in both couplings $(G^*, \lambda^*)$ is then found in four dimensions, with complex eigenvalues $\nu^{-1} = 1.667 \pm 4.308i$ for a gauge parameter choice $\alpha \to \infty$ [for general gauge parameter the exponents can vary by as much as seventy percent (Lauscher and Reuter, 2002)]. In the special limit of vanishing cosmological constant the equations simplify further and one finds a trivial Gaussian fixed point at $G = 0$, as well as a non-trivial ultraviolet fixed point with $\nu^{-1} = 2d(d-2)/(d+2)$, which in $d = 4$ gives now $\nu^{-1} = 2.667$. So in spite of the apparent crudeness of the lowest order approximation, an ultraviolet fixed point similar to the one found in the $2+\varepsilon$ expansion is recovered.

## 3.7 Running of $\alpha(\mu)$ in Gauge Theories

*QED* and *QCD* provide two invaluable illustrative cases where the running of the gauge coupling with energy is not only theoretically well understood, but also verified experimentally. This section is just intended to provide some analogies and distinctions between the two theories, in a way later suitable for a comparison with the gravitational case. Most of the results found in this section are well known (see, for example, Frampton, 2000), but the purpose here is to provide some contrast (and in some instances, a relationship) with the gravitational case.

In *QED* the non-relativistic static Coulomb potential is affected by the vacuum polarization contribution due to electrons (and positrons) of mass $m$. To lowest order in the fine structure constant, the contribution is from a single Feynman diagram involving a fermion loop. One finds for the vacuum polarization contribution $\omega_R(\mathbf{k^2})$ at small $\mathbf{k^2}$ the well known result (Itzykson and Zuber, 1980)

$$\frac{e^2}{\mathbf{k^2}} \to \frac{e^2}{\mathbf{k^2}[1+\omega_R(\mathbf{k^2})]} \sim \frac{e^2}{\mathbf{k^2}} \left[ 1 + \frac{\alpha}{15\pi} \frac{\mathbf{k^2}}{m^2} + O(\alpha^2) \right] , \tag{3.135}$$

which, for a Coulomb potential with a charge centered at the origin of strength $-Ze$ leads to well-known Uehling $\delta$ -function correction

$$V(r) = \left( 1 - \frac{\alpha}{15\pi} \frac{\Delta}{m^2} \right) \frac{-Ze^2}{4\pi r} = \frac{-Ze^2}{4\pi r} - \frac{\alpha}{15\pi} \frac{-Ze^2}{m^2} \delta^{(3)}(\mathbf{x}) . \tag{3.136}$$

It is not necessary though to resort to the small-$\mathbf{k^2}$ approximation, and in general a static charge of strength $e$ at the origin will give rise to a modified potential

$$\frac{e}{4\pi r} \rightarrow \frac{e}{4\pi r} Q(r) \tag{3.137}$$

with

$$Q(r) = 1 + \frac{\alpha}{3\pi} \ln \frac{1}{m^2 r^2} + \ldots \qquad mr \ll 1 \tag{3.138}$$

for small $r$, and

$$Q(r) = 1 + \frac{\alpha}{4\sqrt{\pi}\,(mr)^{3/2}} e^{-2mr} + \ldots \qquad mr \gg 1 \ , \tag{3.139}$$

for large $r$. Here the normalization is such that the potential at infinity has $Q(\infty) = 1$.[3] The reason we have belabored this example is to show that the screening vacuum polarization contribution would have dramatic effects in QED if for some reason the particle running through the fermion loop diagram had a much smaller (or even close to zero) mass. There are two interesting aspects of the (one-loop) result of Eqs. (3.138) and (3.139). The first one is that the exponentially small size of the correction at large $r$ is linked with the fact that the electron mass $m_e$ is not too small: the range of the correction term is $\xi = 2\hbar/mc = 0.78 \times 10^{-10} cm$, but would have been much larger if the electron mass had been a lot smaller.

In *QCD* (and related Yang-Mills theories) radiative corrections are also known to alter significantly the behavior of the static potential at short distances. The changes in the potential are best expressed in terms of the running strong coupling constant $\alpha_S(\mu)$, whose scale dependence is determined by the celebrated beta function of $SU(3)$ *QCD* with $n_f$ light fermion flavors

$$\mu \frac{\partial \alpha_S}{\partial \mu} = 2\beta(\alpha_S) = -\frac{\beta_0}{2\pi}\alpha_S^2 - \frac{\beta_1}{4\pi^2}\alpha_S^3 - \frac{\beta_2}{64\pi^3}\alpha_S^4 - \ldots \tag{3.140}$$

with $\beta_0 = 11 - \frac{2}{3}n_f$, $\beta_1 = 51 - \frac{19}{3}n_f$, and $\beta_2 = 2857 - \frac{5033}{9}n_f + \frac{325}{27}n_f^2$. The solution of the renormalization group equation of Eq. (3.140) then gives for the running of $\alpha_S(\mu)$

$$\alpha_S(\mu) = \frac{4\pi}{\beta_0 \ln \mu^2/\Lambda_{\overline{MS}}^2}\left[1 - \frac{2\beta_1}{\beta_0^2}\frac{\ln\left[\ln \mu^2/\Lambda_{\overline{MS}}^2\right]}{\ln \mu^2/\Lambda_{\overline{MS}}^2} + \ldots\right] \ , \tag{3.141}$$

(see Fig. 3.8). The non-perturbative scale $\Lambda_{\overline{MS}}$ appears as an integration constant of the renormalization group equations, and is therefore - by construction - scale independent. The physical value of $\Lambda_{\overline{MS}}$ cannot be fixed from perturbation theory alone, and needs to be determined by experiment, giving $\Lambda_{\overline{MS}} \simeq 220 MeV$.

In principle one can solve for $\Lambda_{\overline{MS}}$ in terms of the coupling at any scale, and in particular at the cutoff scale $\Lambda$, obtaining

[3] The running of the fine structure constant has recently been verified experimentally at *LEP*. The scale dependence of the vacuum polarization effects gives a fine structure constant changing from $\alpha(0) \sim 1/137.036$ at atomic distances to about $\alpha(m_{Z_0}) \sim 1/128.978$ at energies comparable to the $Z^0$ boson mass, in good agreement with the theoretical renormalization group prediction.

$$
\begin{aligned}
\Lambda_{\overline{MS}} &= \Lambda \exp\left( - \int^{\alpha_S(\Lambda)} \frac{d\alpha_S'}{2\beta(\alpha_S')} \right) \\
&= \Lambda \left( \frac{\beta_0\, \alpha_S(\Lambda)}{4\pi} \right)^{\beta_1/\beta_0^2} e^{-\frac{2\pi}{\beta_0\, \alpha_S(\Lambda)}} \left[ 1 + O[\alpha_S(\Lambda)] \right] \ . \qquad (3.142)
\end{aligned}
$$

In lattice QCD this is usually taken as the definition of the running strong coupling constant $\alpha_S(\mu)$. It then leads to an effective potential between quarks and anti-quarks of the form

$$
V(\mathbf{k^2}) \;=\; -\frac{4}{3} \frac{\alpha_S(\mathbf{k^2})}{\mathbf{k^2}} \ , \qquad (3.143)
$$

and the leading logarithmic correction makes the potential appear softer close to the origin, $V(r) \sim 1/(r \ln r)$.

When the *QCD* result is contrasted with the *QED* answer of Eqs. (3.135) and (3.136) it appears that the infrared small $\mathbf{k^2}$ singularity in Eq. (3.143) is quite serious. An analogous conclusion is reached when examining Eq. (3.141): the coupling strength $\alpha_S(\mathbf{k^2})$ diverges in the infrared due to the singularity at $k^2 = 0$. In phenomenological approaches to low energy *QCD* (Richardson, 1979) the uncontrolled growth in $\alpha_S(\mathbf{k^2})$ due to the spurious small-$k^2$ divergence is regulated by the dynamically generated *QCD* infrared cutoff $\Lambda_{\overline{MS}}$, which can then be shown to give a confining linear potential at large distances.

Not all physical properties can be computed reliably in weak coupling perturbation theory. In non-Abelian gauge theories a confining potential is found at strong coupling by examining the behavior of the Wilson loop (Wilson, 1973), defined for a large closed loop $C$ as

$$
W(C) = < \mathrm{tr}\, \mathscr{P} \exp\left\{ ig \oint_C A_\mu(x) dx^\mu \right\} > \ , \qquad (3.144)
$$

with $A_\mu \equiv t_a A_\mu^a$ and the $t_a$'s the group generators of $SU(N)$ in the fundamental representation. In the pure gauge theory at strong coupling, the leading contribution to the Wilson loop can be shown to follow an area law for sufficiently large loops

$$
W(C) \underset{A\to\infty}{\sim} \exp(-A(C)/\xi^2) \ , \qquad (3.145)
$$

where $A(C)$ is the minimal area spanned by the planar loop $C$. The quantity $\xi$ is the gauge field correlation length, defined for example from the exponential decay of the Euclidean correlation function of two infinitesimal loops separated by a distance $|x|$,

$$
G_\square(x) = < \mathrm{tr}\, \mathscr{P} \exp\left\{ ig \oint_{C_\varepsilon} A_\mu(x') dx'^\mu \right\}(x)\ \mathrm{tr}\, \mathscr{P} \exp\left\{ ig \oint_{C_\varepsilon} A_\mu(x'') dx''^\mu \right\}(0) >_c \ . \qquad (3.146)
$$

Here the $C_\varepsilon$'s are two infinitesimal loops centered around $x$ ands 0 respectively, suitably defined on the lattice as elementary square loops, and for which one has at sufficiently large separations

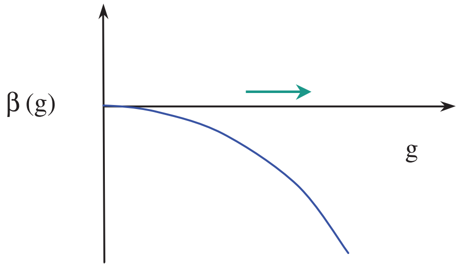



**Fig. 3.8** The *QCD* $\beta$-function in four dimensions, with an ultraviolet stable fixed point at $g = 0$.

$$G_{\Box}(x) \underset{|x|\to\infty}{\sim} \exp(-|x|/\xi) \ . \tag{3.147}$$

The inverse of the correlation length $\xi$ corresponds to the lowest mass excitation in the gauge theory, the scalar glueball, $m_0 = 1/\xi$. Notice that since the glueball mass $m_0$ is expected to be proportional to the parameter $\Lambda_{\overline{MS}}$ of Eq. (3.142) for small $g$, it is non-analytic in the gauge coupling.

# Chapter 4
# Hamiltonian and Wheeler-DeWitt Equation

## 4.1 Classical Initial Value Problem

In formulating the gravitational analogue of the classical initial value (or Cauchy) problem one needs to specify initial value data at some given time; one should then be able to determine the field configurations at some later time, by making appropriate use of the field equations. In such a program the first requirement is therefore a knowledge of $g_{\mu\nu}$ and $\partial_0 g_{\mu\nu}$ everywhere on some spatial hypersurface at a given initial time $x^0 = t$. If one can extract from the field equations the quantities $\partial_0^2 g_{\mu\nu}$, then these can be used to fix $g_{\mu\nu}$ and $\partial_0 g_{\mu\nu}$ at a later time $x^0 = t + \delta t$. The process could then be iterated, and one would eventually obtain a full solution for the metric $g_{\mu\nu}$ valid at all subsequent times.

It would seem at first that the above procedure should work, as there are ten second time derivatives of the metric, and ten field equations. But this is not so, since the field equations are not all independent due to the Bianchi identity,

$$\left(R^{\mu\nu} - \tfrac{1}{2} g^{\mu\nu} R\right)_{;\nu} = 0 \ , \tag{4.1}$$

which implies that the Einstein tensor $G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R$ satisfies

$$\partial_0 G^{\mu 0} = -\partial_i G^{\mu i} - \Gamma^{\mu}_{\nu\lambda} G^{\lambda\nu} - \Gamma^{\nu}_{\nu\lambda} G^{\mu\lambda} \ . \tag{4.2}$$

One can verify that the right-hand side only contains first and second time derivatives of the metric, which implies that $G^{\mu 0}$ does not contain second time derivatives of the metric. As a result, one cannot use the four field equations

$$G^{\mu 0} = 8\pi G \, T^{\mu 0} \ , \tag{4.3}$$

as time evolution equations for the metric. Instead, these should be regarded as *constraints* to be imposed on the initial conditions, that is on the quantities $g_{\mu\nu}$ and $\partial_0 g_{\mu\nu}$ at an initial time $x^0 = t$.

The remaining six field equations

$$G^{ij} = 8\pi G T^{ij} \tag{4.4}$$

should then be viewed as the true dynamical equations. In order to write down these equations, it will be necessary to solve for the ten second derivatives $\partial_0^2 g_{\mu\nu}$. But further thought reveals that this is not possible, as one can only solve for the six quantities $\partial_0^2 g_{ij}$, since there are only six equations, thereby leaving the remaining four quantities $\partial_0^2 g_{\mu 0}$ undetermined. Nevertheless the resulting ambiguity in the solution is not unexpected, due to the freedom of performing arbitrary coordinate transformations. One would therefore expect that the ambiguity present in the solution could be resolved by imposing a suitable coordinate condition, such as the harmonic one,

$$\partial_\nu \left(\sqrt{g} g^{\mu\nu}\right) = 0 \quad . \tag{4.5}$$

In this case, after differentiating with respect to time, one obtains

$$\partial_0^2 \left(\sqrt{g}\, g^{\mu 0}\right) = -\partial_0 \partial_i \left(\sqrt{g}\, g^{\mu i}\right) \quad , \tag{4.6}$$

which would then allow one to determine the remaining second time derivatives of the metric $\partial^2 g_{\mu 0}$.

Furthermore one might worry that the constraint on the initial data needs to be imposed at every step of the calculation. This is not so, as the Bianchi identity $(G^{\mu\nu} - 8\pi G T^{\mu\nu})_{;\nu} = 0$, together with the initial data constraint $G^{\mu 0} = 8\pi T^{\mu 0}$ and the second derivatives from Eq. (4.4), guarantees that at the initial time $x^0 = t$ one has

$$\partial_0 \left(G^{\mu 0} - 8\pi G\, T^{\mu 0}\right) = 0 \quad . \tag{4.7}$$

So that the constraint on the initial data will still be satisfied at a later time $x^0 = t + \delta t$. It would seem therefore that the above method will lead to a consistent scheme for determining the time evolution of the metric $g_{\mu\nu}$, provided one makes a suitable choice of coordinate conditions, so as to avoid the problem of having some components of the metric remaining undetermined. The key issue of identifying the relevant dynamical equations, as well as the constraints, in a Hamiltonian formulation of gravity was first addressed in a series of papers by Dirac (Dirac, 1958; 1959).

## 4.2 First Order Formulation

One natural way of posing the initial value problem is, as in classical mechanics, in terms of a Hamiltonian. In developing a Hamiltonian approach to gravity it is useful to first write the action in first order form, where the metric $g$ and connection $\Gamma$ are considered as independent variables. In the Lagrangian formulation of general relativity one considers variations of the Einstein-Hilbert action with respect to the metric $g_{\mu\nu}$, which then yield the field equations. But if one wants to obtain a canonical form of these equations, that is a set of equations of the type $\dot{q} = \partial H / \partial p$,

$\dot{p} = -\partial H/\partial q$, then one needs to re-write the equations of motion in such a way that only first derivatives of the metric appear.

In the first order (Palatini) formulation of general relativity one writes for the Einstein-Hilbert pure gravity action

$$I = \frac{1}{16\pi G} \int d^4x \sqrt{g}\, g^{\mu\nu} R_{\mu\nu}(\Gamma) \ , \tag{4.8}$$

with $R_{\mu\nu}(\Gamma)$ now considered only as a function of the affine connection $\Gamma$,

$$R_{\mu\nu}(\Gamma) = \partial_\lambda \Gamma^\lambda_{\mu\nu} - \partial_\nu \Gamma^\lambda_{\mu\lambda} + \Gamma^\lambda_{\mu\nu}\Gamma^\sigma_{\lambda\sigma} - \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\nu\lambda} \ . \tag{4.9}$$

Variation of the gravitational action then requires that

$$\frac{1}{16\pi G} \int d^4x \ \delta[\sqrt{g}\, g^{\mu\nu} R_{\mu\nu}] = 0 \ . \tag{4.10}$$

First by varying with respect to the metric $g^{\mu\nu}$ one obtains the Einstein field equations,

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = 0 \ . \tag{4.11}$$

At this point the metric $g_{\mu\nu}$ and the connection $\Gamma^\lambda_{\mu\nu}$ are still thought of as independent variables, and a relationship between the two needs to be established before one can claim to have reproduced correctly the field equations.

The variation of $R_{\mu\nu}$ can be simplified by virtue of the Palatini identity

$$\delta R^\lambda_{\ \mu\nu\sigma} = \delta\Gamma^\lambda_{\mu\sigma;\nu} - \delta\Gamma^\lambda_{\mu\nu;\sigma} \ . \tag{4.12}$$

After integrating by parts one can then show that the term involving the variation of the connection $\Gamma$ implies

$$\partial_\lambda g_{\mu\nu} - g_{\nu\sigma}\Gamma^\sigma_{\mu\lambda} - g_{\mu\sigma}\Gamma^\sigma_{\nu\lambda} = 0 \ , \tag{4.13}$$

(normally one writes this as $g_{\mu\nu;\lambda} = 0$). The last equation can then be solved to give the usual relationship between the connection $\Gamma$ and the metric $g$ in Riemannian geometry, namely

$$\Gamma^\lambda_{\mu\nu}(g) = \tfrac{1}{2}\, g^{\lambda\sigma} \Big( \partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu} \Big) \ . \tag{4.14}$$

Instead of using the metric form, one can introduce local Lorentz frames and write the gravitational action in terms of vierbeins (tetrads) $e^a_\mu(x)$ and spin connections $\omega^{ab}_\mu(x)$. In this formalism (see for example Weinberg, 1972) the metric is written as

$$g_{\mu\nu}(x) = \eta_{ab}\, e^a_\mu(x)\, e^b_\nu(x) \ , \tag{4.15}$$

where one can think of the four covariant vector fields $e^a_\mu$ as relating locally non-inertial coordinate system (described by $g_{\mu\nu}$) to locally inertial ones (described by the flat metric $\eta_{ab}$), with $e^a_\mu e_a^\nu = \delta^\mu{}_\nu$. In terms of flat ($\xi^a$) and curved ($x^\mu$) coordinates,

$$d\,\xi^a \;=\; \frac{\partial\,\xi^a}{\partial\,x^\mu}\,dx^\mu \quad , \tag{4.16}$$

with transformation matrix

$$e^a_\mu \;\equiv\; \frac{\partial\,\xi^a}{\partial\,x^\mu} \quad . \tag{4.17}$$

Now one can construct a new coordinate-scalar Lorentz vector covariant derivative $\mathscr{D}_a$ defined as

$$\mathscr{D}_a \;=\; e_a{}^\mu\left(\partial_\mu + \Gamma_\mu\right) \quad , \tag{4.18}$$

with the matrix $\Gamma_\mu$ given by

$$\Gamma_\mu \;=\; \tfrac{1}{2}\,\sigma^{ab}\,e_a{}^\nu\,\nabla_\mu\,e_{b\nu} \quad . \tag{4.19}$$

The matrices $\sigma_{ab}$ satisfying the usual commutation relations for the generators of the Lorentz group,

$$\left[\sigma_{ab},\sigma_{cd}\right] \;=\; \eta_{bc}\,\sigma_{ad} - \eta_{ac}\,\sigma_{bd} + \eta_{bd}\,\sigma_{ac} - \eta_{ad}\,\sigma_{bc} \quad , \tag{4.20}$$

with $\sigma_{ab} = -\sigma_{ba}$. The spin connection $\omega^{ab}_\mu$ is given in terms of the $e^a_\mu$'s by

$$\omega^{ab}_\mu \;=\; \tfrac{1}{2}e^{a\nu}(\partial_\mu e^b_\nu - \partial_\nu e^b_\mu) + \tfrac{1}{4}\,e^{a\rho}\,e^{b\sigma}(\partial_\sigma e^c_\rho - \partial_\rho e^c_\sigma)e_{\mu\,c} - (a \leftrightarrow b) \quad . \tag{4.21}$$

It has a number of important properties, the first of which is a relationship to the usual affine connection $\Gamma^\lambda_{\mu\nu}$,

$$\Gamma^\lambda_{\mu\nu} \;=\; e^\lambda_a\,e_{b\mu}\,\omega^{ab}_\nu + e^\lambda_a\,\partial_\nu\,e^a_\mu \quad . \tag{4.22}$$

The second property is related to the fact that the quantities $\omega^{ab}_\mu$ have been constructed in such a way that the covariant derivative of $e^a_\mu$ is identically zero,

$$\begin{aligned} D_\mu e^a_\nu &= \partial_\mu\,e^a_\nu - \Gamma^\lambda_{\mu\nu}\,e^a_\lambda + \omega^{ab}_\mu\,e_{b\nu} \\ &= \partial_\mu\,e^a_\nu - e_{b\nu}\,\omega^{ab}_\mu - \partial_\mu\,e^a_\nu + \omega^{ab}_\mu\,e_{b\nu} \;=\; 0 \quad . \end{aligned} \tag{4.23}$$

The new formulation is essential for incorporating fermions into general relativity. Thus on a fermion field the correct definition of a covariant derivative involves the spin connection $\omega^{ab}_\mu$, $D_\mu\,\psi = (\partial_\mu + \frac{1}{2}\sigma_{ab}\,\omega^{ab}_\mu)\,\psi$.

Finally one can construct the curvature tensor in the $e^a_\mu$ basis,

$$R^{ab}_{\mu\nu} \;=\; \partial_\mu\,\omega^{ab}_\nu - \partial_\nu\,\omega^{ab}_\mu + \omega^a_{\mu c}\,\omega^{cb}_\nu - \omega^a_{\nu c}\,\omega^{cb}_\mu \quad , \tag{4.24}$$

and the vierbein field can be used later to transform it into the ordinary Riemann tensor

$$R^\lambda{}_{\sigma\mu\nu} \;=\; e^\lambda_a\,e_{b\sigma}\,R^{ab}_{\mu\nu} \quad . \tag{4.25}$$

In this formalism the pure gravitational action can be written as

$$I = \frac{1}{16\pi G} \int d^4x \, |\det e| \left( e_a^\mu \, e_b^\nu \, R_{\mu\nu}^{ab} - 2\lambda \right) \ , \tag{4.26}$$

where we have added, for later reference, a cosmological constant term. The above action can be shown to reproduce, after varying with respect to $e_a^\mu$ and $\omega_\mu^{ab}$, the Einstein field equations. By construction it is invariant under both local Lorentz and general coordinate transformations.

A closely related approach is the starting point for the loop quantization of gravity. There one introduces a self-dual connection defined as

$$A_\mu^{ab} = \tfrac{1}{2} \left( \omega_\mu^{ab} - \tfrac{1}{2} i \, \varepsilon_{cd}^{ab} \, \omega_\mu^{cd} \right) \ , \tag{4.27}$$

where the dual ($^*$) of an antisymmetric two index object is defined here as $f^{*ab} \equiv -\frac{1}{2} i \varepsilon_{cd}^{ab} f^{cd}$ with $f^{**} = f$. From the field $A_\mu^{ab}$ one can define a curvature $F_{\mu\nu}^{ab}$, and from it an action

$$I = \frac{1}{8\pi G} \int d^4x \, |\det e| \left( e_a^\mu \, e_b^\nu \, F_{\mu\nu}^{ab} - \lambda \right) \ , \tag{4.28}$$

which can also be shown to reproduce correctly the classical field equations (Ashtekar, 1986a,b; 2004). These variables then provide the basis for the so-called loop quantum gravity method, for which recent reviews can be found in (Smolin, 2003; Thiemann, 2007a,b).

## 4.3 Arnowitt-Deser-Misner (ADM) Formalism

The next step in developing a Hamiltonian formulation for gravity is to introduce a time-slicing of space-time, by introducing a sequence of spacelike hypersurfaces, labeled by a continuous time coordinate $t$ (Arnowitt, Deser and Misner, 1960; 1962). The invariant distance $ds^2 = -d\tau^2$ is then written as

$$\begin{aligned} -d\tau^2 &= g_{\mu\nu} \, dx^\mu dx^\nu \\ &= g_{ij} \, (dx^i + N^i dt)(dx^j + N^j dt) - N^2 dt^2 \\ &= g_{ij} \, dx^i dx^j + 2 g_{ij} N^i dx^j dt - (N^2 - g_{ij} N^i N^j) dt^2 \ , \end{aligned} \tag{4.29}$$

where $x^i$ $(i = 1,2,3)$ are coordinates on a three-dimensional manifold. The relationship between the quantities $d\tau$, $dt$, $dx_i$, $N$ and $N_i$ here basically expresses the Lorentzian version of Pythagoras' theorem (see Fig. 4.1).

Indices are raised and lowered with $g_{ij}(x)$ $(i, j = 1,2,3)$, which denotes the three-metric on the given spacelike hypersurface, and $N(x)$ and $N^i(x)$ are the lapse and shift functions, respectively. These last two quantities describe the lapse of proper time ($N$) between two infinitesimally close hypersurfaces, and the corresponding

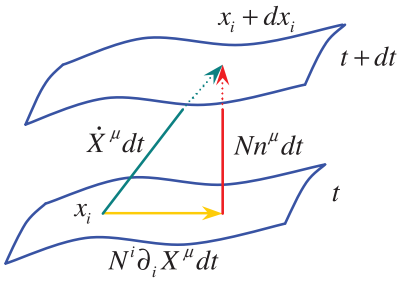



**Fig. 4.1** Illustration of the lapse and shift functions. $N_i$ is the shift function and $N$ the lapse function. In terms of a normal to the surface $n^\mu$ one has $\partial_t X^\mu = N n^\mu + N^i \partial_i X^\mu$.

shift in spatial coordinate ($N^i$). It will be useful in the following to mark four-dimensional quantities by the prefix $^4$, so that all un-marked quantities will refer to three dimensions (and are occasionally marked explicitly by a $^3$). In terms of the original four-dimensional metric $^4g_{\mu\nu}$ one has

$$\begin{pmatrix} {}^4g_{00} & {}^4g_{0j} \\ {}^4g_{i0} & {}^4g_{ij} \end{pmatrix} = \begin{pmatrix} N_k N^k - N^2 & N_j \\ N_i & g_{ij} \end{pmatrix} , \tag{4.30}$$

and for its inverse

$$\begin{pmatrix} {}^4g^{00} & {}^4g^{0j} \\ {}^4g^{i0} & {}^4g^{ij} \end{pmatrix} = \begin{pmatrix} -1/N^2 & N^j/N^2 \\ N_i/N^2 & g^{ij} - N^i N^j/N^2 \end{pmatrix} , \tag{4.31}$$

which then gives for the spatial metric and the lapse and shift functions

$$g_{ij} = {}^4g_{ij} \qquad N = \left(-{}^4g^{00}\right)^{-1/2} \qquad N_i = {}^4g_{0i} \ . \tag{4.32}$$

For the volume element one has

$$\sqrt{{}^4g} = N\sqrt{g} \ , \tag{4.33}$$

where the latter involves the determinant of the three-metric, $g = \det g_{ij}$. As usual $g^{ij}$ denotes the inverse of the matrix $g_{ij}$.

In terms of these quantities, the Einstein-Hilbert Lagrangian of General Relativity can then be written, up to an overall multiplicative constant, in the following (first-order) form

$$\begin{aligned} \mathscr{L} = \sqrt{{}^4g}\,{}^4R = &-g_{ij}\,\partial_t \pi^{ij} - N R^0 - N_i R^i \\ &-2\,\partial_i\left(\pi^{ij} N_j - \tfrac{1}{2}\pi N^i + \nabla^i N \sqrt{g}\right) \ , \end{aligned} \tag{4.34}$$

where one has defined the following quantities:

$$\pi^{ij} = \sqrt{{}^4g}\left({}^4\Gamma^0_{kl} - g_{kl}\,{}^4\Gamma^0_{mn}\, g^{mn}\right) g^{ik} g^{lj}$$

$$
\begin{aligned}
R^0 &= -\sqrt{g}\left[{}^3R + g^{-1}\left(\tfrac{1}{2}\pi^2 - \pi^{ij}\pi_{ij}\right)\right] \\
R^i &= -2\nabla_j \pi^{ij} \ .
\end{aligned}
\tag{4.35}
$$

Here the symbol $\nabla^i$ denotes covariant differentiation with respect to the index $i$ using the spatial three-metric $g_{ij}$, and ${}^3R$ is the scalar curvature computed from this metric. Also note that the affine connection coefficients $\Gamma^k_{ij}$ have been eliminated in favor of the spatial derivatives of the metric $\partial_k g_{ij}$, and one has defined $\pi = \pi^i{}_i$.

A particularly simple form for the Lagrangian of Eq. (4.34) is obtained when the term involving a spatial divergence and a total time derivative is omitted. In this case

$$
\begin{aligned}
\mathscr{L} &= \pi^{ij}\,\partial_t g_{ij} \ - \ N R^0 \ - \ N_i R^i \\
R^\mu \ &= \ R^\mu(g_{ij}, \pi_{ij}) \ .
\end{aligned}
\tag{4.36}
$$

Since the quantities $N$ and $N^i$ do not appear in the $\pi^{ij}\,\partial_t g_{ij}$ part, they are interpreted as Lagrange multipliers, and the "Hamiltonian" density

$$
\mathscr{H} \ = \ N R^0 + N_i R^i \ = \ 0 \ , \tag{4.37}
$$

vanishes as a result of the constraints.

Varying the first order Lagrangian of Eq. (4.34) with respect to $g_{ij}$, $N_i$, $N$ and $\pi_{ij}$ one then obtains a set of equations which are equivalent to Einstein's field equations. First varying with respect to $\pi_{ij}$ one obtains an equation which can be viewed as defining $\pi_{ij}$ (in analogy to $\dot{q} = \partial H / \partial p$)

$$
\partial_t g_{ij} \ = \ 2N g^{-1/2}\left(\pi_{ij} - \tfrac{1}{2} g_{ij}\,\pi\right) + \nabla_j N_i + \nabla_i N_j \ . \tag{4.38}
$$

Varying with respect to the spatial metric $g_{ij}$ gives the time evolution for $\pi_{ij}$ (in analogy to $\dot{p} = -\partial H / \partial q$),

$$
\begin{aligned}
\partial_t \pi^{ij} = &-N\sqrt{g}\left({}^3R^{ij} - \tfrac{1}{2} g^{ij}\,{}^3R\right) + \tfrac{1}{2} N g^{-1/2} g^{ij}\left(\pi^{kl}\pi_{kl} - \tfrac{1}{2}\pi^2\right) \\
&-2N g^{-1/2}\left(\pi^{ik}\pi_k^j - \tfrac{1}{2}\pi\pi^{ij}\right) + \sqrt{g}\left(\nabla^i\nabla^j N - g^{ij}\,\nabla^k\nabla_k N\right) \\
&+\nabla_k\left(\pi^{ij}N^k\right) - \nabla_k N^i\,\pi^{kj} - \nabla_k N^j\,\pi^{ki} \ .
\end{aligned}
\tag{4.39}
$$

Finally varying with respect to the lapse ($N$) and shift ($N^i$) functions gives

$$
R^0(g_{ij}, \pi_{ij}) = 0 \qquad R^i(g_{ij}, \pi_{ij}) = 0 \ , \tag{4.40}
$$

which can be viewed as the four constraint equations ${}^4G^0_\mu = {}^4R^0_\mu - \frac{1}{2}\delta^0_\mu\,{}^4R = 0$, expressed for this choice of metric decomposition (Dirac, 1958). The above constraints can be considered as analogous to Gauss's law $\partial_i F^{i0} = \nabla \cdot \mathbf{E} = 0$ in electromagnetism.

One can count the number of physical degrees of freedom of the gravitational field, and verify that one obtains the expected answer. There are twelve dynamical variables $g_{ij}$ and $\pi_{ij}$, but the Bianchi identities reduce the number by four. An additional four constraints can be imposed through coordinate conditions, giving a residual of four fields describing two degrees of freedom, corresponding to the two helicity states of the linearized gravitational field, which describes a massless spin two particle.

## 4.4 Orthogonal Decomposition in the Linearized Theory

The linearized gravitational field case is the easiest to work out. As usual one assumes boundary conditions such that some or all the field vanish at infinity, where space is assumed to be flat. One needs an orthogonal decomposition of the metric into trace part, longitudinal part and transverse-traceless part, which is achieved by writing for any symmetric tensor

$$f_{ij} \;=\; f_{ij}^{TT} \;+\; f_{ij}^{T} \;+\; \partial_i f_j \;+\; \partial_j f_i \;\;, \tag{4.41}$$

which similar to what is done in electromagnetism, where the vector potential $\mathbf{A}$ is written as a transverse and a longitudinal part. Here one writes

$$\begin{aligned} f_i &= (1/\partial^2)\left[\partial_j f_{ij} - \tfrac{1}{2}(1/\partial^2)\,\partial_i \partial_j \partial_k f_{kj}\right] \\ f^T &= f_{ii} - (1/\partial^2)\,\partial_i \partial_j f_{ij} \\ f_{ij}^{TT} &= f_{ij} - f_{ij}^{T}[f] - \partial_i f_j[f] - \partial_j f_i[f] \;\;, \end{aligned} \tag{4.42}$$

for the longitudinal, trace and transverse-traceless part respectively, with the quantity $f_{ij}^{T}$ defined by

$$f_{ij}^{T} \;\equiv\; \tfrac{1}{2}\left[\,\delta_{ij} f^T \,-\, (1/\partial^2)\,\partial_i\partial_j f^T\right] \;\;. \tag{4.43}$$

In the above expressions $\partial^2 \equiv \partial_i\partial_i$. To this order one can show that both $g^T$ and $\pi^i$ vanish. Furthermore $\pi^T$ and $g_i$ also can be eliminated by a choice of coordinate condition, such as

$$t \;=\; (-1/2\,\partial^2)\,\pi^T \;\;, \tag{4.44}$$

giving $N_i = g_0 i = 0$ everywhere, as well as $N = 1$.

This finally leaves $g_{ij}^{TT}$ and $\pi^{ijTT}$ as the only two remaining canonically conjugate variables in the linearized theory, with fundamental equal time Poisson bracket

$$\left\{ g_{ij}^{TT}(\mathbf{x}),\, \pi^{kl\,TT}(\mathbf{x}') \right\} \;=\; \delta^{kl}{}_{ij}(\mathbf{x}-\mathbf{x}') \;\;, \tag{4.45}$$

and all other equal time Poisson brackets equal to zero. The modified Dirac $\delta$ function on the r.h.s. ensures that the transversality constraint on the fields appearing on the l.h.s. is not violated. The use of a fundamental Poisson bracket then allows a straightforward transition to a quantum mechanical treatment via the usual replacement $\{q,p\} \;\rightarrow\; (1/i\hbar)\;[\hat{q},\hat{p}]$.

## 4.5 Intrinsic and Extrinsic Curvature, Hamiltonian

Some of the quantities introduced in the previous section (such as ${}^3R$) describe intrinsic properties of the spacelike hypersurface. Some others can be related to the extrinsic properties of such a hypersurface when embedded in four-dimensional space. Consider spacetime as sliced up (foliated) by a one-parameter family of spacelike hypersurfaces $x^\mu = x^\mu(x^i, t)$. One then has for the intrinsic metric within the spacelike hypersurface

$$g_{ij} \; = \; g_{\mu\nu} X_i^\mu X_j^\nu \quad \text{with} \quad X_i^\mu \; \equiv \; \partial_i x^\mu \quad , \tag{4.46}$$

while the extrinsic curvature is given in terms of the unit normals to the spacelike surface, $U^\mu$,

$$K_{ij}(x^k, t) \; = \; -\nabla_\mu U_\nu X_i^\nu X_j^\mu \quad . \tag{4.47}$$

In this language, the lapse and shift functions appear in the expression

$$\partial_t x^\mu \; = \; N U^\mu + N^i X_i^\mu \quad . \tag{4.48}$$

Then the Einstein tensor $G_{\mu\nu}$ can be projected into directions normal $(\perp)$ and tangential $(i)$ to the hypersurface, with the result

$$\begin{aligned} G_{\perp\perp} \equiv G_{\mu\nu} U^\mu U^\nu \; &= \; -\tfrac{1}{2}\,(K_{ij}K^{ij} - K^2 - {}^3\!R) \\ G_{i\perp} \equiv -G_{\mu\nu} X_i^\mu U^\nu \; &= \; \nabla_j (K_i^j - K\delta_i^j) \\ G^i_j \equiv G_{\mu\nu} X^{\mu i} X_j^\nu \; &= \; -\partial_t (K^i_j - K\,\delta^i_j) + K\,K^i_j \\ &\quad -\tfrac{1}{2}(K^m_n K^n_m + K^2)\,\delta^i_j + {}^3G^i_j \quad , \end{aligned} \tag{4.49}$$

with $K = g^{ij} K_{ij} = K^i{}_i$ the trace of the matrix $\mathbf{K}$.

Then in the canonical formalism the momentum can be expressed in terms of the extrinsic curvature as

$$\pi^{ij} \; = \; -\sqrt{g}\,(K^{ij} - K\,g^{ij}) \quad . \tag{4.50}$$

By inverting this last relationship, one can then replace the extrinsic curvatures $K$ by the six momenta $\pi$. Then the quantities $G_{\perp\perp}$ and $G_{i\perp}$ in Eq. (4.49) can be expressed entirely in terms of $g_{ij}$ and $\pi^{ij}$ within a given spacelike hypersurface,

$$\begin{aligned} -2\sqrt{g}\,G_{\perp\perp} \equiv H \; &= \; 2Kg^{-1/2}\left(\pi_{ij}\pi^{ij} - \tfrac{1}{2}\pi^2\right) - \frac{1}{2K}\sqrt{g}\,{}^3\!R \\ 2\sqrt{g}\,G_{i\perp} \equiv H_i \; &= \; -2Kg^{-1/2}\nabla_j \pi_i^j \quad , \end{aligned} \tag{4.51}$$

with $\pi = \pi^i{}_i$. The last two statements are essentially equivalent to the definitions in Eq. (4.35). On the other hand the expression involving $G^i_j$ in Eq. (4.49) can be rearranged to give

$$\partial_t \pi^{ij}(x,t) \; = \; \left\{\pi^{ij}(x,t), H_N + H_{\mathbf{N}}\right\} \; + \; N\,G^{ij} \quad , \tag{4.52}$$

while Eq. (4.50) gives the evolution equation for the intrinsic metric,

$$\partial_t g_{ij}(x,t) = \left\{ g_{ij}(x,t), H_N + H_{\mathbf{N}} \right\} . \tag{4.53}$$

Here $\{A,B\} \equiv \sum_r (\partial A/\partial q_r \partial B/\partial p_r - \partial A/\partial p_r \partial B/\partial q_r)$ is the classical Poisson bracket. Its use is motivated here by the fact that the transition from classical to quantum mechanics can be affected easily by promoting the Poisson bracket to a quantum commutator,

$$\{H,O\} \rightarrow \tfrac{1}{i\hbar}\,[\hat{H},\hat{O}] , \tag{4.54}$$

thus leading in a natural and simple way from Eqs. (4.52) and (4.53) to the Heisenberg equations of motions for the canonically conjugate quantum operators $\hat{g}_{ij}$ and $\hat{\pi}^{ij}$. The quantities $H_N$ and $H_{\mathbf{N}}$ are given by

$$H_N \equiv \int d^3x N(x) H(x) , \quad H_{\mathbf{N}} \equiv \int d^3x N^i(x) H_i(x) . \tag{4.55}$$

In this notation, the Einstein field equations in the absence of sources $G_{\mu\nu} = 0$ are equivalent to the initial value constraint

$$H(x) = H_i(x) = 0 , \tag{4.56}$$

supplemented by the canonical evolution equations of Eqs. (4.52) and (4.53), with $G_{ij} = 0$. The quantity

$$\mathbf{H} = \int d^3x \left[ N(x) H(x) + N^i(x) H_i(x) \right] , \tag{4.57}$$

should then be regarded as the Hamiltonian for classical general relativity. Then Eq. (4.56) is equivalent to

$$\mathbf{H} = 0 , \tag{4.58}$$

for all $N$ and $N_i$.

## 4.6 Matter Source Terms

When matter is added to the Einstein-Hilbert Lagrangian,

$$I[g,\phi] = \frac{1}{16\pi G} \int d^4x \sqrt{g}\, {}^4\!R[g_{\mu\nu}(x)] + I_\phi[g_{\mu\nu},\phi] , \tag{4.59}$$

where $\phi(x)$ are some matter fields, it is easy to see that the action within the ADM parametrization of the metric coordinates needs to be modified to

$$I[g,\pi,\phi,\pi_\phi,N] = \int dt\, d^3x \left( \tfrac{1}{16\pi G}\, \pi^{ij}\, \partial_t g_{ij} + \pi_\phi\, \partial_t \phi - N T - N^i T_i \right) . \tag{4.60}$$

One still has the same definitions as before for the (Lagrange multiplier) lapse and shift function, namely $N = (-{}^4g^{00})^{-1/2}$ and $N^i = g^{ij}\,{}^4g_{0j}$. The gravitational constraints are modified as well, since now one defines

$$\begin{aligned} T &\equiv \tfrac{1}{16\pi G} H(g_{ij}, \pi^{ij}) \;+\; H^{\phi}(g_{ij}, \pi^{ij}, \phi, \pi_{\phi}) \\ T_i &\equiv \tfrac{1}{16\pi G} H_i(g_{ij}, \pi^{ij}) \;+\; H_i^{\phi}(g_{ij}, \pi^{ij}, \phi, \pi_{\phi}) \;, \end{aligned} \tag{4.61}$$

with the first part describing the gravitational part already given in Eq. (4.51),

$$\begin{aligned} H(g_{ij}, \pi^{ij}) &= G_{ij,kl}\pi^{ij}\pi^{kl} \;-\; \sqrt{g}\,{}^3R + 2\lambda\sqrt{g} \\ H_i(g_{ij}, \pi^{ij}) &= -2\,g_{ij}\nabla_k\,\pi^{jk} \;, \end{aligned} \tag{4.62}$$

here conveniently re-written using the (inverse of) the DeWitt supermetric of Eq. (2.14),

$$G_{ij,kl} \;=\; \tfrac{1}{2}\,g^{-1/2}\left(g_{ik}g_{jl} + g_{il}g_{jk} + \alpha\,g_{ij}g_{kl}\right) \;, \tag{4.63}$$

with parameter $\alpha = -1$. In the previous expression a cosmological term (proportional to $\lambda$) has been added as well, for future reference. For the matter part one has

$$\begin{aligned} H^{\phi}(g_{ij}, \pi^{ij}, \phi, \pi_{\phi}) &= \sqrt{g}\,T_{00}\,(g_{ij}, \pi^{ij}, \phi, \pi_{\phi}) \\ H_i^{\phi}(g_{ij}, \pi^{ij}, \phi, \pi_{\phi}) &= -\sqrt{g}\,T_{0i}\,(g_{ij}, \pi^{ij}, \phi, \pi_{\phi}) \;. \end{aligned} \tag{4.64}$$

We note here that the (inverse of the) DeWitt supermetric in Eq. (4.63) can be used to define a distance in the space of three-metrics $g_{ij}(x)$. Consider an infinitesimal displacement of such a metric $g_{ij} \to g_{ij} + \delta g_{ij}$, and define the natural metric $G$ on such deformations by considering a distance in function space

$$\|\delta g\|^2 \;=\; \int d^3x N(x)\; G^{ij,kl}(x)\;\delta g_{ij}(x)\,\delta g_{kl}(x) \;. \tag{4.65}$$

Here the lapse $N(x)$ is an essentially arbitrary but positive function, to be taken equal to one in the following. The quantity $G^{ij,kl}(x)$ is the three-dimensional version version of the DeWitt supermetric,

$$G^{ij,kl} \;=\; \tfrac{1}{2}\sqrt{g}\left(g^{ik}g^{jl} + g^{il}g^{jk} + \bar{\alpha}\,g^{ij}g^{kl}\right) \;, \tag{4.66}$$

with the parameter $\alpha$ of Eq. (4.63) related to $\bar{\alpha}$ in Eq. (4.66) by $\bar{\alpha} = -2\alpha/(2+3\alpha)$, so that $\alpha = -1$ gives $\bar{\alpha} = -2$. One can then verify that for any choice of $\alpha \neq -2/3$

$$G^{ij,ab}\,G_{kl,ab} \;=\; \tfrac{1}{2}\left(\delta_k^i\delta_l^j \;+\; \delta_l^i\delta_k^j\right) \;. \tag{4.67}$$

As shown originally by DeWitt, in three dimensions the supermetric has signature $(+,+,+,+,+,-)$, implying infinitely many negative signs, one for each spatial point $\mathbf{x}$. The negative directions in function space can be shown to correspond to constant conformal displacements in the three-metric $\delta g_{ij}(x) = \delta\Omega^2 g_{ij}(x)$ (Giulini, 1995).

## 4.7 Wheeler-DeWitt Equation

Within the framework of the previous construction, a transition from the classical to the quantum description of gravity is obtained by promoting $g_{ij}$, $\pi^{ij}$, $H$ and $H_a$ to quantum operators, with $\hat{g}_{ij}$ and $\hat{\pi}^{ij}$ satisfying canonical commutation relations. In particular the classical constraints now select a physical vacuum state $|\Psi\rangle$, such that in the source free case

$$\hat{H}\,|\Psi\rangle \;=\; 0 \qquad \hat{H}_i\,|\Psi\rangle \;=\; 0 \;, \tag{4.68}$$

and in the presence of sources more generally

$$\hat{T}\,|\Psi\rangle \;=\; 0 \qquad \hat{T}_i\,|\Psi\rangle \;=\; 0 \;. \tag{4.69}$$

As in ordinary non-relativistic quantum mechanics, one can choose different representations for the canonically conjugate operators $\hat{g}_{ij}$ and $\hat{\pi}^{ij}$. In the functional *position representation* one sets

$$\hat{g}_{ij}(\mathbf{x}) \;\rightarrow\; g_{ij}(\mathbf{x}) \qquad \hat{\pi}^{ij}(\mathbf{x}) \;\rightarrow\; -i\hbar\cdot 16\pi G\cdot \frac{\delta}{\delta g_{ij}(\mathbf{x})} \;. \tag{4.70}$$

In further developing the analogy with standard non-relativistic quantum mechanics, one notices that in this picture the quantum states become wave functionals of the three-metric $g_{ij}(\mathbf{x})$,

$$|\Psi\rangle \;\rightarrow\; \Psi[g_{ij}(\mathbf{x})] \;. \tag{4.71}$$

The two quantum constraint equations in Eq. (4.69) become the Wheeler-DeWitt equation

$$\left\{ -\,16\pi G\cdot G_{ij,kl}\,\frac{\delta^2}{\delta g_{ij}\,\delta g_{kl}} \;-\; \frac{1}{16\pi G}\,\sqrt{g}\left({}^3\!R - 2\lambda\right) + \hat{H}^{\phi} \right\} \Psi[g_{ij}(\mathbf{x})] \;=\; 0 \;, \tag{4.72}$$

and the diffeomorphism, or momentum, constraint

$$\left\{ 2\,i\,g_{ij}\,\nabla_k\,\frac{\delta}{\delta g_{jk}} + \hat{H}_i^{\phi} \right\} \Psi[g_{ij}(\mathbf{x})] \;=\; 0 \;. \tag{4.73}$$

The last constraint implies that the gradient of $\Psi$ on the superspace of $g_{ij}$'s and $\phi$'s is zero along those directions that correspond to gauge transformations, i.e. diffeomorphisms on the three dimensional manifold whose points are labeled by the coordinates $\mathbf{x}$.

A number of basic issues need to be addressed before one can gain a full and consistent understanding of the dynamical content of the theory. These include the problem of operator ordering in the above equations (in particular regarding the $\pi^2$ term, which classically can be written in a number of different and equivalent ways), a discussion of what is meant by the time variable, and how it can be suitably defined in concrete models, for example in cosmological applications. In addition one needs to be specific about a suitable Hilbert space, which entails at some point a specific choice for the inner product of wave functionals over the space $\Sigma$ (and thus a notion of self-adjointness for operators), for example in the Schrödinger form

$$\langle \Psi | \Phi \rangle \; = \; \int_\Sigma d\mu[g] \, \Psi^*[g_{ij}] \, \Phi[g_{ij}] \;\; , \tag{4.74}$$

where $d\mu[g]$ is some appropriate measure over the three-metric $g$. The latter does not seem to be the only choice, since a Klein Gordon inner product could be used instead, which is not positive definite.

Another peculiar property of the Wheeler-DeWitt equation, and which distinguishes it from the usual Schrödinger equation $H\Psi = i\hbar\partial_t\Psi$, is the absence of an explicit time coordinate. As a result the r.h.s. term of the Schrödinger equation is here entirely absent. The reason is of course diffeomorphism invariance of the underlying theory, which expresses now the fundamental quantum equations in terms of fields $g_{ij}$, and not coordinates. As a result the Wheeler-DeWitt equation contains no explicit time evolution parameter, a problem that is usually referred to as the *problem of time* (see for example Kuchař, 1992). Nevertheless in some cases it seems possible to assign the interpretation of "time coordinate" to some specific variable entering the Wheeler-DeWitt equation, such as the overall spatial volume or the magnitude of some scalar field. But in general a consistent and unambiguous prescription does not seem to be known yet.

## 4.8 Semiclassical Expansion of the Wheeler-DeWitt Equation

The simplest approach to finding solutions to the Wheeler-DeWitt equations, besides working with the linearized theory, is to expand around the classical theory. One writes a (WKB-type) ansatz for the wave functional $\Psi$,

$$\Psi[g_{ij}(\mathbf{x})] \; = \; \exp\left\{ \tfrac{i}{16\pi G} \, S[g_{ij}] \right\} \; \Phi[g_{ij}(\mathbf{x})] \;\; , \tag{4.75}$$

where the action function $S[g_{ij}]$ is a solution of the Hamilton-Jacobi equations for classical gravity (Peres, 1962),

$$\begin{aligned} G_{ij,kl} \, \frac{\delta S}{\delta g_{ij}} \, \frac{\delta S}{\delta g_{kl}} \; - \; \sqrt{g} \left( {}^3R - 2\lambda \right) = 0 \\ 2 \, i \, g_{ij} \, \nabla_k \, \frac{\delta S}{\delta g_{jk}} = 0 \;\; , \end{aligned} \tag{4.76}$$

and $\Phi[g_{ij}]$ some new wave functional to be determined. From the original Wheeler-DeWitt equation one then obtains a new set of equations for the wavefunctional $\Phi[g_{ij}]$,

$$\left\{ -2i\, G_{ij,kl}\, \frac{\delta S}{\delta g_{ij}}\, \frac{\delta}{\delta g_{kl}} \;-\; i\, G_{ij,kl}\, \frac{\delta^2 S}{\delta g_{ij}\, \delta g_{kl}} \right.$$
$$\left. -16\pi G \cdot G_{ij,kl}\, \frac{\delta^2}{\delta g_{ij}\, \delta g_{kl}} \;+\; \hat{H}^{\phi} \right\} \Phi[g_{ij}(\mathbf{x})] \;=\; 0 \;. \tag{4.77}$$

$$\left\{ 2\, i\, g_{ij}\, \nabla_k\, \frac{\delta}{\delta g_{jk}} + \hat{H}_i^{\phi} \right\} \Phi[g_{ij}(\mathbf{x})] \;=\; 0 \;. \tag{4.78}$$

The next step consists in approximating (again, in analogy with the semiclassical expansion in non-relativistic quantum mechanics), the solution by neglecting second derivative terms $\delta^2 S/\delta g^2$ and $\delta^2/\delta g^2$ terms in the equations for $\Phi[g_{ij}]$ (Wheeler, 1964). The latter step is usually justified by regarding (or assuming) the back-reaction of quantum matter on the gravitational field as small.

The resulting truncated Wheeler-deWitt equations then become, to first order in $\delta/\delta g$,

$$\left\{ -2\, i\, G_{ij,kl}\, \frac{\delta S}{\delta g_{ij}}\, \frac{\delta}{\delta g_{kl}} \;+\; \hat{H}^{\phi} \right\} \Phi[g_{ij}(\mathbf{x})] = 0$$
$$\left\{ 2\, i\, g_{ij}\, \nabla_k\, \frac{\delta}{\delta g_{jk}} + \hat{H}_i^{\phi} \right\} \Phi[g_{ij}(\mathbf{x})] = 0 \;. \tag{4.79}$$

Furthermore the wavefunction $\Phi[g_{ij}(\mathbf{x})]$ is now evaluated along a solution of the classical field equations $g_{ij}(\mathbf{x},t)$. This means that $S[g_{ij}]$ is first determined from a solution of the classical Hamilton-Jacobi equations

$$\partial_t\, g_{ij} \;=\; N\, G_{ij,kl}\, \frac{\delta S}{\delta g_{kl}} - 2(D_i N_j + D_j N_i) \;, \tag{4.80}$$

with [see Eq. (4.73)]

$$D_i \;\equiv\; -\frac{2}{i}\, \nabla_j\, \frac{\delta}{\delta g_{ij}} \;, \tag{4.81}$$

after which the relevant derivatives $\delta S/\delta g$ are inserted in Eq. (4.77).

To make further progress, one need to be more specific about the form of the lapse ($N$) and shift ($N^i$) functions. One can show that $\Phi[g_{ij}]$ satisfies an evolution equation of the type

$$\frac{\partial}{\partial t}\, \Phi(t) \;=\; \int d^3x\, \partial_t\, g_{ij}(\mathbf{x})\, \frac{\delta}{\delta g_{ij}(\mathbf{x})}\, \Phi[g_{mn}] \;, \tag{4.82}$$

as well as Eqs. (4.77) and (4.78). It was DeWitt who first showed that these are equivalent to the Schrödinger equation for quantized matter field in a classical gravitational background $g_{ij}(\mathbf{x})$,

$$i \frac{\partial}{\partial t} \Phi(t) = \hat{H}^{\phi}[g_{ij}] \Phi(t) , \tag{4.83}$$

with a background $g_{ij}$-dependent matter field Hamiltonian

$$\hat{H}^{\phi}[g_{ij}] = \int d^3x \left\{ N(\mathbf{x}) \hat{H}^{\phi}(\mathbf{x}) + N^i(\mathbf{x}) \hat{H}_i^{\phi}(\mathbf{x}) \right\} . \tag{4.84}$$

A discussion on how the Wheeler-deWitt equation and its semiclassical expansion relate to the conventional covariant Feynman diagram picture can be found in (Barvinsky and Kiefer, 1998; Barvinsky, 1998).

## 4.9 Connection with the Feynman Path Integral

In principle any solution of the Wheeler-deWitt equation corresponds to a possible quantum state of the universe. It is also clear the effects of the boundary conditions on the wavefunction will act to severely restrict the class of possible solutions. In ordinary quantum mechanics these are determined by the physical context of the problem and some set of external conditions. In the case of the universe as a whole the situation is less clear, and in many approaches some suitable set of boundary conditions are postulated instead, based on general arguments involving concepts such as simplicity or economy.

One proposal (Hartle and Hawking, 1983) is to restrict the quantum state of the universe by requiring that the wave function $\Psi$ be determined by a path integral over compact Euclidean metrics. The wave function would then be given by

$$\Psi[g_{ij}, \phi] = \int [dg_{\mu\nu}] [d\phi] \exp\left(-\hat{I}[g_{\mu\nu}, \phi]\right) , \tag{4.85}$$

where $\hat{I}$ is the Euclidean action for gravity plus matter

$$\hat{I} = -\frac{1}{16\pi G} \int d^4x \sqrt{g} (R - 2\lambda) - \frac{1}{8\pi G} \int d^3x \sqrt{g_{ij}} K - \int d^4x \sqrt{g} \mathscr{L}_m . \tag{4.86}$$

The functional integral would be restricted to those four-metrics which have the induced metric $g_{ij}$ and the matter field $\phi$ as given on the boundary surface $S$. One would expect, as is already the case in non-relativistic quantum mechanics where the path integral with a boundary surface satisfies the Schrödinger equation (Feynman and Hibbs, 1963), that the wavefunction constructed in this way would also automatically satisfy the Wheeler-DeWitt equation. In (Hartle and Hawking, 1983) this is shown to be indeed the case.

Other approaches to the issue of boundary conditions and the construction of the wave functional can be found, among others, in (Linde, 1998), where an anti-Wick rotation $t = i\tau$ is suggested, which might improve convergence issues involving the gravitational conformal mode, but causes irreparable damage in the non-gravitational path integrals, and in (Vilenkin, 1998) where tunneling minisuperspace models are examined and considered as phenomenologically viable, based on claims that the Hartle-Hawking wafe functions might tend to disfavor inflationary evolution. A very recent reference addressing these issues is (Hartle, Hawking and Hertog, 2008).

## 4.10 Minisuperspace

The quantum state of a gravitational system is described, in the Wheeler-deWitt framework just introduced, by a wave function $\Psi$ which is a functional of the three-metric $g_{ij}$ and the matter fields $\phi$. In general the latter could contain fields of arbitrary spins, but here we will consider for simplicity just one single component scalar field $\phi(x)$.

The wavefunction $\Psi$ will then obey the zero energy Schrödinger-like equation of Eqs. (4.72) and (4.73),

$$\left\{ -16\pi G \cdot G_{ij,kl} \frac{\delta^2}{\delta g_{ij}\,\delta g_{kl}} - \frac{1}{16\pi G}\sqrt{g}\left({}^3\!R - 2\lambda\right) + \sqrt{g}\,T_{00}(\partial/\partial\phi,\phi) \right\} \Psi[g_{ij},\phi] \;=\; 0 \;, \tag{4.87}$$

with inverse supermetric

$$G_{ij,kl} \;=\; \tfrac{1}{2}\, g^{-1/2} \left(g_{ik}g_{jl} + g_{il}g_{jk} - g_{ij}g_{kl}\right) \;, \tag{4.88}$$

and momentum constraint

$$\left\{ i\nabla_i \frac{\delta}{\delta g_{ij}} - \tfrac{1}{2}\sqrt{g}\,T^{0j}(\partial/\partial\phi,\phi) \right\} \Psi[g_{ij},\phi] \;=\; 0 \;. \tag{4.89}$$

Then quantum state described by $\Psi$ is a functional on the infinite dimensional manifold $W$ consisting of all positive definite metrics $g_{ij}(x)$ and matter fields $\phi(x)$ on a spacelike three-surface $S$. On this space there is a natural metric $\Gamma(N)$

$$ds^2[\delta g,\delta\phi] = \int \frac{d^3x\,d^3x'}{N(x)} \left[ G^{ij,kl}(x,x')\,\delta g_{ij}\,\delta g_{kl}(x') + \sqrt{g}\,\delta^3(x-x')\,\delta\phi(x)\delta\phi(x') \right] , \tag{4.90}$$

where

$$\begin{aligned} G^{ij,kl}(x,x') &= G^{ij,kl}(x)\,\delta^3(x-x') \\ G^{ij,kl}(x) \;&=\; \tfrac{1}{2}\sqrt{g}\left[ g^{ik}(x)g^{jl}(x) + g^{il}(x)g^{jk}(x) - 2\,g^{ij}(x)g^{kl}(x) \right] \;, \end{aligned} \tag{4.91}$$

is the inverse of the deWitt supermetric $G_{ij,kl}$.

Due to ambiguities in the choice of operator ordering in the Wheeler-DeWitt equations, not all terms and their coefficients can be fixed in a unique way. General symmetry requirements (here covariance in superspace) restrict the Hamiltonian $H = \int d^3x N(x) H(x)$ in Eqs. (4.62) and (4.72) to be of the following form (Hawking and Page, 1986 a,b)

$$H = -\tfrac{1}{2}\nabla^2 + \beta \cdot 16\pi G \int d^3x N g^{-1/2} g_{ij} \frac{\delta}{\delta g_{ij}} + \varepsilon + V \tag{4.92}$$

$$V = \int d^3x N \sqrt{g} \left[ \tfrac{1}{16\pi G} (-{}^3R + 2\lambda) + U(\phi) \right] \tag{4.93}$$

where $\nabla^2$ the covariant Laplacian in the function space $W$ with metric $\Gamma(N)$ [see Eq. (4.90)],

$$\varepsilon = \xi \mathscr{R}(g) + \eta \int d^3x \sqrt{g} \;, \tag{4.94}$$

is a scalar term involving the scalar curvature on the function space $W$, $\xi$ and $\eta$ two constants, and

$$U = T_{00} - \tfrac{1}{2} \pi_\phi^2 \;. \tag{4.95}$$

Since in general the $\beta$-term violates the self-adjointness requirement on $H$, one sets $\beta = 0$. The most natural (and simplest choice) for $\xi$, the coefficient of the scalar curvature term $\mathscr{R}$, is zero. The $\eta$ term can be re-absorbed into a shift of the cosmological term $\lambda$. We shall not dwell here on the rather technical point that in general the function $N$ enters non-linearly in the superspace connection on $W$, and therefore in $H$, which then spoils the interpretation of $N$ as a Lagrange multiplier.

In general the wavefunction for all the dynamical variables of the gravitational field in the universe is difficult to calculate, since an infinite number of degrees of freedom are involved: the infinitely many values of the metric at all spacetime points, and the infinitely many values of the matter field $\phi$ at the same points. One option is to restrict the choice of variable to a finite number of suitable degrees of freedom (Blyth and Isham, 1975; Hartle and Hawking, 1983). As a result the overall quantum fluctuations are severely restricted, since these are now only allowed to be nonzero along the surviving dynamical directions. If the truncation is severe enough, the transverse-traceless nature of the graviton fluctuation is lost as well. Also, since one is not expanding the quantum solution in a small parameter, it can be difficult to estimate corrections. In a cosmological context, it seems natural to consider initially a homogeneous and isotropic model, and restrict the function space to two variables, the scale factor $a(t)$ and a minimally coupled homogeneous scalar field $\phi(t)$ (Hawking and Page, 1986a,b). The space-time metric is given by

$$d\tau^2 = N^2(t)\, dt^2 - g_{ij}\, dx^i dx^j \;. \tag{4.96}$$

The three-metric $g_{ij}$ is then determined entirely by the scale factor $a(t)$,

$$g_{ij} = a^2(t)\, \tilde{g}_{ij} \;, \tag{4.97}$$

with $\tilde{g}_{ij}$ a time-independent reference three-metric with constant curvature,

$$ {}^3\tilde{R}_{ijkl} = k\left(\tilde{g}_{ij}\,\tilde{g}_{kl} - \tilde{g}_{il}\,\tilde{g}_{jk}\right) \ , \tag{4.98} $$

and $k = 0, \pm 1$ corresponding to the flat, closed and open universe case respectively. In this case the minisuperspace $W$ is two dimensional, with coordinates $a$ and $\phi$, and supermetric $\Gamma(N)$

$$ ds^2[a,\phi] = N^{-1}(-a\,da^2 + a^3\,d\phi^2) \ . \tag{4.99} $$

It is important to note the indefinite nature of the supermetric, a general feature of general relativity. From the above expression for $ds^2[a,\phi]$ one obtains the Laplacian in the metric $\Gamma(N)$ needed in Eq. (4.92),[1]

$$ -\tfrac{1}{2}\nabla^2(a,\phi) = \frac{N}{2a^2}\left\{\frac{\partial}{\partial a}\,a\,\frac{\partial}{\partial a} - \frac{1}{a}\,\frac{\partial^2}{\partial\phi^2}\right\} \ . \tag{4.100} $$

Since the space is homogeneous, the diffeomorphism constraint is trivially satisfied. Also $N$ is independent of $g_{ij}$ so in the homogeneous case it can be taken to be a constant, conveniently chosen as $N = 1$. This is because the lapse function $N(t)$ represents the freedom to reparametrize the time coordinate $t$; it has no kinetic term and hence no canonical momentum, its only effect being to ensure the constraint $H = 0$. In light of the previous discussions, once the constraint has been imposed, one is free to choose a time coordinate such that $N = 1$. Furthermore, the supercurvature $\mathscr{R}(g)$ is zero for this choice of metric, removing one source of ambiguity in the equation.

The simplest and most straightforward way to derive the Wheeler-DeWitt equation for these models is to start from the gravitational action written in terms of the appropriate metric coefficients,

$$ I = V_3 \int dt\,N\,\left\{\tfrac{3}{8\pi G}[-N^{-2}\,a\dot{a}^2 + k\,a - \tfrac{1}{3}\,\lambda\,a^3] + \tfrac{1}{2}\,a^3\,[N^{-2}\dot{\phi}^2 - 2U(\phi)]\right\} \ , \tag{4.101} $$

with $V_3$ the volume of three-space (for a three-sphere $V_3 = 2\pi^2$), and $U(\phi)$ a potential for the scalar field $\phi$. In the following we will assume for simplicity that $\phi$ is a free scalar field with $U(\phi) = \frac{1}{2}m^2\phi^2$ and $m$ the scalar field's mass. From the canonical momenta derived from the action in Eq. (4.101),

$$ p_a = -\frac{3}{4\pi G}N^{-1}\,a\,\dot{a} \qquad p_\phi = N^{-1}\,a^3\,\dot{\phi} \qquad p_N = 0 \ , \tag{4.102} $$

one can then construct the classical Hamiltonian,

[1] The ambiguity regarding the operator ordering of $p^2/a = a^{-(q+1)}pa^q p$ in the Wheeler-DeWitt equation can in principle be retained by writing for the above operator $\nabla^2$ the expression $-(N/a^{q+1})\left\{(\partial/\partial a)\,a^q\,(\partial/\partial a) - (\partial^2/\partial\phi^2)\right\}$, but this does not seem to affect the qualitative nature of the solutions. The case discussed in the text corresponds to $q = 1$, but $q = 0$ seems even simpler.

$$\begin{aligned} H &= p_a \dot{a} + p_\phi \dot{\phi} + p_N \dot{N} - L(p_a, a, p_\phi, \phi, p_N, N) \\ &= -\bar{G} a^{-1} p_a^2 + \tfrac{1}{2} a^{-3} p_\phi^2 - \bar{k} a + \bar{\lambda} a^3 + \tfrac{1}{2} m^2 a^3 \phi^2 \ , \end{aligned} \tag{4.103}$$

where we have defined $\bar{G} \equiv 2\pi G/3$, $\bar{\lambda} = \lambda/8\pi G$ and $\bar{k} = 3k/8\pi G$ (occasionally a system of units is used in which $\frac{4\pi}{3} G = 1$), and we have set for the lapse function $N = 1$.[2] Classically the resulting equations of motion coincide, as expected, with the Friedman equations for $a$ and $\phi$. Quantum-mechanically one obtains from Eq. (4.103) the reduced Wheeler-DeWitt equation $H\Psi = 0$, which reads[3]

$$\left\{ \bar{G} \frac{1}{a^2} \frac{\partial}{\partial a} a \frac{\partial}{\partial a} - \frac{1}{2a^3} \frac{\partial^2}{\partial \phi^2} - \bar{k} a + \bar{\lambda} a^3 + \tfrac{1}{2} m^2 a^3 \phi^2 \right\} \Psi(a, \phi) = 0 \ . \tag{4.104}$$

This quantum Schrödinger-type equation can be contrasted with the classical Friedman-Robertson-Walker equations of motion for $a(t)$ and $\phi(t)$ (here, as far as $a(t)$ is concerned, they are essentially Newtonian, as shown originally by Milne and McCrea) , and which read

$$\begin{aligned} \dot{a}^2 &= \frac{4\pi G}{3} a^2 \left( \dot{\phi}^2 + m^2 \phi^2 \right) - k + \tfrac{1}{3} \lambda a^2 \\ \ddot{\phi} &= -3 a^{-1} \dot{a} \dot{\phi} - m^2 \phi \ , \end{aligned} \tag{4.105}$$

subject to some initial conditions at $t = t_0$.

One might wonder if in this case one really needs the full machinery of the Wheeler-deWitt equation to obtain the correct quantum-mechanical minisuperspace equation. The classical FRW equations for $a(t)$ and $\dot{\phi} = 0$ in Eq. (4.105) can after all be derived from an action which is similar, but not identical, to the one in Eq. (4.101). From the canonical momentum $p_a$ one then obtains a classical Hamiltonian, which leads to a Schrödinger-like equation $\hat{H}\Psi = 0$ after the substitution $p_a \to (\hbar/i)\partial/\partial a$. Using this procedure one finds for $p_\phi = 0$ an equation for $\Psi(a)$ of the form

$$\left\{ -\frac{1}{2} \frac{\partial^2}{\partial a^2} - \frac{8\pi G}{3} \rho a^3 - \frac{\lambda}{3} a^3 - \frac{2\pi G}{3} m^2 a^2 \phi^2 + \frac{k}{2} \right\} \Psi(a) = 0 \ , \tag{4.106}$$

---

[2] One can advocate the point of view that for a genuine canonical quantization the system has to be reduced to true canonical form *before* quantization. In this scheme the constraint $H = 0$ is solved first at the classical level, thereby eliminating here the variable $p_a$. The remaining gravitational variable $a$ is then picked as a time coordinate, $t = a$ (Blyth and Isham, 1975). In general it seems this procedure leads to unconstrained Hamiltonians involving square roots.

[3] Some authors have expressed concerns about possible singularities of the Wheeler-deWitt equation at $a = 0$, and prefer therefore to use as field coordinate $\alpha = \log(a/a_0)$. We shall not pursue this choice here. A full quantum treatment generally requires the presence of an ultraviolet cutoff $\Lambda$, which would suggest the imposition of a lower bound of the type $a > 1/\Lambda$.

which is quite different in structure from Eq. (4.104), although some significant similarities are still recognizable.

## 4.11 Solution of Simple Minisuperspace Models

Returning to the minisuperspace Wheeler-DeWitt equation of Eq. (4.104), one can consider, as an illustrative example, the simplest case of no matter ($\phi = 0$), in the flat case $k = 0$. Then the solution to the quantum equation on the half-line $a > 0$, Eq. (4.104), is given for $\lambda \neq 0$ by a linear combination of Bessel functions

$$\Psi(a) \;=\; c_1\, J_0(\tfrac{1}{3}\sqrt{\tilde{\lambda}}\, a^3) + c_2\, Y_0(\tfrac{1}{3}\sqrt{\tilde{\lambda}}\, a^3) \ , \tag{4.107}$$

with $\tilde{\lambda} = \bar{\lambda}/\bar{G}$, $c_1$ and $c_2$ arbitrary constants, and $\Psi(a)$ further constrained by a normalization condition such as $\int_0^\infty d\mu(a)\,|\Psi(a)|^2 = 1$.[4] Because of the restriction $a > 0$ one would expect that $\Psi(0) = 0$; this would then possibly exclude the second solution which is singular at the origin. Note that the condition $a > 0$ can be traced back to the requirement on the three-metric $\det g_{ij} > 0$.

It should be clear that in general the quantum behavior of the solutions is expected to be quite different from the classical one. In the latter case one imposes some initial conditions on the scale factor at some time $t_0$, which then determines $a(t)$ at all later times. In the minisuperspace view of quantum cosmology one has to instead impose a condition on the wavepacket $\Psi$ at $a = 0$. Due to their simplicity, in general it is possible to analyze the solutions to the minisuperspace Wheeler-deWitt equation in a rather complete way, given some sensible assumptions on how $\Psi(a,\phi)$ should behave, for example, when the scale factor $a$ approaches zero. For $k = 1$ (closed universe) and $\bar{\lambda} = 0$ it is convenient to express the Wheeler-DeWitt equation for $\Psi[a,\phi]$ in the field coordinates

$$x \;=\; a\sinh\phi \qquad y \;=\; a\cosh\phi \ , \tag{4.108}$$

which leads to

$$\frac{1}{2}\left[\frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial x^2} + V(x,y)\right]\Psi(x,y) \;=\; 0 \ , \tag{4.109}$$

with a "potential" $V$

$$V(x,y) \;=\; x^2 - y^2 + m^2(x^2-y^2)^2\,\mathrm{arctanh}^2(x/y) \ . \tag{4.110}$$

Then the physical domain $a > 0$ corresponds, in the new parametrization, to $y > |x|$. To fully determine the solution, a condition needs to be imposed on or near the

[4] For a general operator ordering, parametrized by $q$, the solution is modified to a different combination of Bessel functions, $\Psi(a) = c_1\, a^{(1-q)/2} J_{(1-q)/6}(\frac{1}{3}\sqrt{\tilde{\lambda}}\, a^3) + c_2\, a^{(1-q)/2} J_{-(1-q)/6}(\frac{1}{3}\sqrt{\tilde{\lambda}}\, a^3)$.

"light cone" $y \approx |x|$, and one possible choice is $\Psi = 1$. Presumably a better procedure would be an approximate determination in this region of $\Psi$ via an Euclidean functional integral. The qualitative behavior of the solutions in general depends crucially on whether one is in the region $V < 0$ (where the solution $\Psi$ increases exponentially) or in the region $V > 0$ (where the solution for $\Psi$ is oscillatory). From the nature of the solution and its semiclassical correspondence it has been argued that the behavior close to $y = |x|$ ($a = 0$ in the original variable) is consistent with a minimum radius or "bounce" at $V = 0$ (Hawking and Wu, 1984; Ochiai and Sato, 2000).

How do the properties of the solutions to the Wheeler-DeWitt equation applied to minisuperspace models depend on additional terms that might enter into the original gravitational action? To answer this question, one could consider the addition of higher derivative terms, such as

$$\hat{I} = -\frac{1}{16\pi G}\int d^4x\sqrt{g}\,\left[R - \alpha C^{\mu\nu\rho\sigma}C_{\mu\nu\rho\sigma} + \beta R^2\right] + \text{s.t.} \;, \tag{4.111}$$

where $C_{\mu\nu\rho\sigma}$ is the Weyl tensor, and "s.t." refers to surface terms (Hawking and Luttrell, 1984a,b). Following a procedure analogous to the one discussed in the previous case, one obtains for a homogeneous isotropic universe the following Wheeler-DeWitt equation for $\Psi(a,R)$, or, more conveniently, for $\Psi(x,y)$

$$\frac{1}{2}\left[\frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial x^2} + V(x,y)\right]\Psi(x,y) \;=\; 0 \;, \tag{4.112}$$

now with "potential" $V$

$$V(x,y) \;=\; x^2 - y^2 + g x^2 (x-y)^2 \;, \tag{4.113}$$

with $g \equiv 1/18\pi\beta$, and variables $x$ and $y$ given by $x = 2\beta aR$ and $y = 2a(1+\beta R)$, where $R$ is the scalar curvature.

As in the previous example, one is interested in a solution for $a > 0$ which corresponds to $y > x$, with boundary condition $\Psi(x = y) = 1$ (see Fig. 4.2). But this is not sufficient to determine the Cauchy data, and one has to make some additional guess on what a semiclassical wavefunction might look like in the vicinity of $x = y$ or $x = -y$, which suggests

$$\Psi(y = -x) \;=\; \exp(-a^4/36\pi\beta) \;. \tag{4.114}$$

These conditions are then in principle sufficient to determine the wavefunction $\Psi(x,y)$.

Physically the oscillatory behavior of the wavefunction then suggests the existence of small amplitude oscillations superimposed on an overall expansion or contraction. These oscillations in radius would cause particle creation in any matter fields present, which in turn would damp the oscillation.

As one last illustrative example consider a minisuperspace model for a universe filled with matter (or radiation) of uniform density, such that $\rho(a) = M/a^\sigma$ where $\sigma = 3$ (matter) or 4 (radiation); $\sigma = 0$ would correspond to a pure vacuum energy.

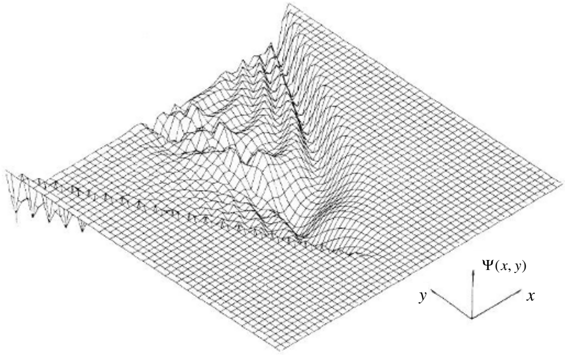



**Fig. 4.2** Minisuperspace wavefunction $\Psi(x,y)$ for the problem in Eq. (4.112), gravity with higher derivative terms, in the region $y > |x|$.

In the following we will consider for simplicity only the case $\sigma = 3$, corresponding to non-relativistic matter. The classical Friedman equations for $\lambda = 0$ give

$$\dot{a}^2 + k - \frac{8\pi G}{3} a^3 \rho(a) \;=\; 0 \;, \tag{4.115}$$

with $k = 0, \pm 1$, and subject to some initial conditions at $t = t_0$. The above equation can be regarded as a classical one-dimensional mechanics problem, with an inverted parabolic potential $V(a) = \frac{k}{2} - \frac{8\pi G}{3} a^3 \rho(a)$. Introducing, as before, the canonical momentum derived from the appropriate classical Lagrangian one finds for the classical Hamiltonian

$$H \;=\; -\bar{G} a^{-1} p^2 - \bar{k} a + a^3 \rho(a) = 0 \;, \tag{4.116}$$

with $\bar{G} = \frac{2\pi}{3} G$ and $\bar{k} = \frac{3}{8\pi G}$. After setting $p^2/a = a^{-q+1}\, p\, a^q\, p$, with $q$ a parameter introduced to describe an operator ordering ambiguity, and replacing $p \to -i\partial/\partial a$ one obtains for the Wheeler-DeWitt equation

$$\left\{ \alpha \frac{1}{a^2} \frac{\partial^2}{\partial a^2} + q\, \alpha \frac{1}{a^3} \frac{\partial}{\partial a} - k + \frac{8\pi G}{3} a^2 \rho(a) \right\} \Psi(a) \;=\; 0 \;, \tag{4.117}$$

with $\alpha \equiv 4\bar{G}^2 = 16\pi^2 G^2/9$. Then for the choice $q = 1$ one can re-write the equation as a one-dimensional zero-energy stationary state Schrödinger-like problem,

$$\left\{ \frac{1}{a^2} \frac{\partial}{\partial a} a \frac{\partial}{\partial a} - \frac{k}{\alpha} a + \frac{8\pi G}{3\alpha} a^3 \rho(a) \right\} \Psi(a) \;=\; 0 \;, \tag{4.118}$$

which in the non-relativistic matter case ($\rho(a) = M/a^3$) corresponds to one-dimensional quantum motion with potential $V_{eff}(a) = (k/\alpha)\, a^2 - \beta\, a$, with $\beta \equiv 8\pi G M/3\alpha > 0$. The shape of the potential for $k = 1$ is an inverted $U$ going through the origin, and a maximum at $a_0 = \beta\alpha/2k$, so that quantum mechanical tunneling through the barrier is possible. The vanishing of the wavefunction at the origin $\psi(0) = 0$ would imply here the absence of a singularity in the quantum case. The solution in this case can be found explicitly,

$$\Psi(a) = c_1 J_0(\gamma a^{3/2}) + c_2 Y_0(\gamma a^{3/2}) \ , \tag{4.119}$$

with $\gamma = (32\pi GM/9\alpha)^{1/2}$. The regular and irregular solutions are shown in Fig. (4.3) for $\gamma = 1$. A slow oscillating decay of the wavefunction $\Psi(a)$ for large scale factor $a$ can be seen as the quantum counterpart to the power-law increase of the scale factor $a(t)$ for large times. A more complete treatment of this problem would include matter as a dynamical variable, with its own equations of motion, as was done in the scalar field case. An update on the methods of minisuperspace models can be found in the recent review (Page, 2002).

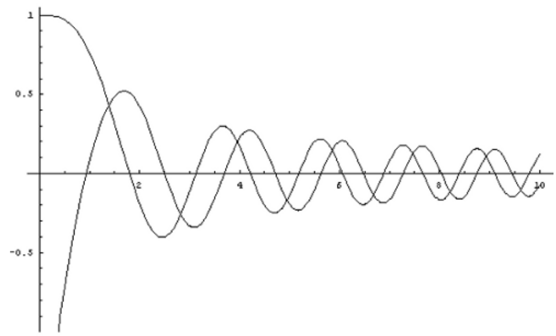



**Fig. 4.3** Minisuperspace wavefunction $\Psi(a)$ for the problem in Eq. (4.118), describing the fluctuations in the scale factor $a$ for non-relativistic matter, shown here in the physical region $a > 0$.

Finally one should mention that the problem of singularity avoidance in quantum cosmology has been, and remains, of great interest. To what extent is the singular behavior of the scale factor (and curvature invariants) at the big-bang singularity avoided by quantum fluctuations? In the models presented so far the singularity at $a = 0$ appears at the boundary of phase space, and has measure zero. The quantum probability would therefore seem to be vanishingly small. Based on our understanding of path integrals such a result should not come as unexpected: singular solutions of classical field equations could end up having zero measure in the full quantum path integral; after all one can prove in a number of simpler cases that *all* classical (differentiable) paths have zero measure in the Feynman path integral (Glimm and Jaffe, 1981). Even if the probability at the origin were finite it is not clear that it would necessarily cause a problem. The analogy with the Hydrogen atom comes to mind, where the Coulomb potential is singular at the origin. Quantum-mechanically, even though the electron in an $s$-wave state has a finite probability of being at the origin, this does not cause a problem as the electron just moves through. In general it appears that, at least within the context of a variety of minisuperspace models, normalizable solutions to the Wheeler-DeWitt equations can be made to vanish at the classical singularity (Kamenshchik, Kiefer and Sandhofer, 2007; Kiefer, 2008).

In concluding the discussion on minisuperspace models as tools for studying the physical content of the Wheeler-DeWitt equation it seems legitimate to ask the following question: to what extent can results for such models, and specifically the ones discussed in the previous section, be representative of what might or might not happen in the full quantum theory?

To such purpose let us consider a field theory model that is non-trivial, yet much simpler that gravity: a self-coupled quantum scalar field $\phi(x)$ in four dimensions with Lagrangian

$$\mathscr{L}_\phi = \tfrac{1}{2}(\partial_\mu \phi)^2 - \tfrac{1}{2} m^2 \phi^2 - \tfrac{1}{4} \lambda \phi^4 \ , \tag{4.120}$$

with a mass term proportional to $m^2$ and a quartic self-interaction proportional to the dimensionless coupling $\lambda$. The perturbative treatment in four dimensions (Wilson, 1973; for a review see Zinn-Justin, 2002) shows that the quantum theory is strongly interacting at short distances, whereas the long distance behavior is determined by the Gaussian fixed point at $\lambda = 0$. The theory essentially becomes non-interacting at large distances, with an effective running coupling $\lambda(p^2) \sim_{p^2 \to 0} 1/|\log(p^2/\Lambda^2)| \to 0$.

A minisuperspace approximation to such a theory would consist in neglecting altogether all spatial derivatives of the quantum field, by setting $\nabla\phi = 0$. The model is then described by a single degree of freedom $\phi(t)$, with Lagrangian

$$\mathscr{L}_\phi = \tfrac{1}{2} \dot{\phi}^2 - \tfrac{1}{2} m^2 \phi^2 - \tfrac{1}{4} \lambda \phi^4 \ . \tag{4.121}$$

From the expression for the momentum variable $\pi_\phi = \dot{\phi}$, and the substitution $\hat{\pi}_\phi \to (\hbar/i)\partial/\partial\phi$, one then obtains the quantum Schrödinger equation for stationary states in this reduced phase space, which is

$$\left\{ -\frac{1}{2} \frac{\partial^2}{\partial \phi^2} + U(\phi) - k \right\} \Psi(\phi) = 0 \ , \tag{4.122}$$

with $U(\phi) = \frac{1}{2} m^2 \phi^2 + \frac{1}{4} \lambda \phi^4$ and $k$ the energy eigenvalue. Normalizable solutions are of course the usual Hermite eigenfunctions of the quantum harmonic oscillator in the position representation. The ground state wavefunction (with $k = \frac{1}{2} m$) would then roughly correspond to the minisuperspace solution of the Wheeler-DeWitt equation $H\Psi = 0$ for gravity.

The shortcomings of such a minisuperspace truncation of the original quantum theory of Eq. (4.120) are now becoming evident: the model no longer contains any propagating degrees of freedom along the spatial directions $\mathbf{x}$. Since $\phi$ is assumed to be constant in $\mathbf{x}$, any correlations in the spatial directions are absent, which is troubling since for the free part one knows that such vacuum correlations are not negligible: they are given by

$$\langle \phi(\mathbf{x},0) \, \phi(\mathbf{x}',0) \rangle \sim \frac{1}{|\mathbf{x} - \mathbf{x}'|^2} \ , \tag{4.123}$$

and cannot be considered small in any sense (in particular they diverge as $\mathbf{x} \to \mathbf{x}'$, or, equivalently, become sensitive to an ultraviolet cutoff at short distances).

Furthermore, since one is dealing with a finite number of degrees of freedom, there are no radiative corrections, no renormalization effects, and thus no scale dependent couplings arising from the quantum corrections. In other words, the model is still in many ways essentially *classical*, in spite of the appearance of some quantum variables such as $\hat{\pi}_\phi$. In particular the short distance, and therefore high particle density, behavior of the theory, which in the full treatment becomes strongly coupled at short distance due to the UV growth of the coupling $\lambda$, is *not* described correctly by the minisuperspace model.

It would seem that similar concerns could be raised regarding the minisuperspace approximation to quantum gravity. The infinitely many degrees of freedom of the metric in this case are just reduced to a few, such as $a(t)$. As a result, the transverse-traceless nature of quantum fluctuations in the linearized limit is no longer apparent. The spatial quantum fluctuations of the metric, which are expected to acquire divergent contributions due to the ultraviolet renormalization effects of four-dimensional quantum field theories, and quantum gravity in particular, are set to zero, and quantum correlations in the spatial directions are entirely neglected. Finally the highly degenerate (and therefore genuinely quantum mechanical) nature of the strongly coupled graviton gas is not taken into account, one more aspect which could, possibly, be quite relevant for early time cosmology.

## 4.12 Quantum Hamiltonian for Gauge Theories

It is of interest to see how the Hamiltonian approach has fared for ordinary $SU(N)$ gauge theories, whose non-trivial infrared properties (confinement, chiral symmetry breaking) cannot be seen to any order in perturbation theory, and require therefore some sort of non-perturbative approach, based for example on a strong coupling expansion. In the continuum one starts from the Yang-Mills action

$$I = -\frac{1}{4g^2} \int d^4x F_{\mu\nu}^a F^{\mu\nu a} \quad , \tag{4.124}$$

with field strength

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f_{abc} A_\mu^b A_\nu^c \quad , \tag{4.125}$$

and gauge fields are $A_\mu^a$ with $(a = 1 \dots N^2 - 1)$, where the quantities $f_{abc}$ are the structure constants of the Lie group, such that the generators satisfy $[T_a, T_b] = i f_{abc} T_c$. The gauge invariant energy-momentum tensor for this theory is given by

$$T^{\mu\nu} = F_a^{\mu\rho} F_{\rho a}^{\nu} - \tfrac{1}{4} \eta^{\mu\nu} F_a^{\rho\sigma} F_a^{\sigma\rho} \quad , \tag{4.126}$$

and in terms of the fields strengths $E_a^i = F_a^{i0}$ and $B_a^i = -\frac{1}{2} \varepsilon_{ijk} F_a^{jk}$ (with $i, j, k = 1, 2, 3$) one has

$$T^{00} = \tfrac{1}{2}\left(E_a^i E_a^i + B_a^i B_a^i\right) \tag{4.127}$$

and

$$T^{0i} = \varepsilon_{ijk} E_a^j B_a^k \ . \tag{4.128}$$

$T^{00}$ could be interpreted as a Hamiltonian density, and therefore be used to construct the quantum Hamiltonian, were it not for the fact that some of the degrees of freedom, as shown below, are unphysical.

It is convenient to rewrite the gauge field Lagrangian in first order form (see for example Itzykson and Zuber, 1980), For concreteness we will discuss here the $SU(2)$ case with $f_{abc} = \varepsilon_{abc}$ (in the following bold-face vectors will therefore refer to iso-vectors with color index $a$). The Lagrangian is then

$$\mathscr{L} = \tfrac{1}{4}\mathbf{F}_{\mu\nu}\cdot\mathbf{F}^{\mu\nu} - \tfrac{1}{2}\mathbf{F}_{\mu\nu}\cdot\left(\partial^\mu\mathbf{A}^\nu - \partial^\nu\mathbf{A}^\mu + g\,\mathbf{A}^\mu\times\mathbf{A}^\nu\right) \ . \tag{4.129}$$

The Euler-Lagrange equations give

$$\mathbf{F}_{\mu\nu} = \partial_\mu\mathbf{A}_\nu - \partial_\nu\mathbf{A}_\mu + g\,\mathbf{A}_\mu\times\mathbf{A}_\nu \tag{4.130}$$

and

$$\partial^\mu\mathbf{F}_{\mu\nu} + g\,\mathbf{A}^\mu\times\mathbf{F}_{\mu\nu} = 0 \ , \tag{4.131}$$

with time evolution equations

$$\begin{aligned} \partial_0\mathbf{A}_i &= \mathbf{F}_{0i} + (\nabla_i + g\,\mathbf{A}_i\times)\,\mathbf{A}_0 \\ \partial_0\mathbf{F}_{0i} &= (\nabla_j + g\,\mathbf{A}_j\times)\,\mathbf{F}_{ji} - g\,\mathbf{A}_0\times\mathbf{F}_{0i} \ . \end{aligned} \tag{4.132}$$

The field canonically conjugate to $\mathbf{A}_i$ is $\mathbf{F}_{0i}$ (with the chromo-electric field having been defined as $E_a^i = F_a^{i0}$).

On the other hand the field canonically conjugate to $A_0$ vanishes, since there is no $\partial_0 A^0$ term in the Lagrangian,

$$\boldsymbol{\pi}^0 = 0 \ , \tag{4.133}$$

so this field must be treated as a dependent variable. In Dirac's language this is called a primary constraint. From the second equation of motion, Eq. (4.131) one has

$$(\nabla_k + g\mathbf{A}_k\times)\,\mathbf{F}_{k0} = 0 \ , \tag{4.134}$$

which is the analog of Gauss's equation $\nabla\cdot\mathbf{E} = 0$ in electrodynamics. Equivalently the last constraint can be written in terms of canonical momenta $\boldsymbol{\pi}_i$ as

$$(\nabla_k + g\mathbf{A}_k\times)\,\boldsymbol{\pi}_k = 0 \ , \tag{4.135}$$

which is sometimes referred to as a secondary constraint, since it involves the use of the equations of motion. Eq. (4.134) tells us that not all conjugate momenta $\mathbf{F}_{k0}$ are independent, and one needs therefore to impose a gauge condition, such as

$$\nabla_k\mathbf{A}_k = 0 \ , \tag{4.136}$$

which then implies that $\mathbf{A}$ is transverse.

The next step is to separate out the transverse and longitudinal parts in $\mathbf{F}_{0i}$ by setting $\mathbf{F}_{0i}^{L} = -\nabla_i \mathbf{f}$ and $\mathbf{F}_{0i}^{T} = \mathbf{E}_i$, solve for $\mathbf{f}$ using the constraint equation of Eq. (4.134), and eliminate $\mathbf{A}_0$ to obtain the physical Hamiltonian

$$H = \tfrac{1}{2} \int d^3x \left[ \mathbf{E}_i^2 + \mathbf{B}_i^2 + (\nabla_i \mathbf{f})^2 \right] \quad . \tag{4.137}$$

The last terms represents the instantaneous Coulomb interaction that is characteristic of this gauge.

## 4.13 Lattice Regularized Hamiltonian for Gauge Theories

The previous section discussed the Hamiltonian formulation for Yang-Mills theories in the continuum. In view of applying the Hamiltonian method to quantum gravity, it is of interest to see how such an approach is implemented in $SU(N)$ gauge theories at strong coupling first. In gauge theories the only known way to deal in a systematic way with the strong coupling problem is to formulate gauge theories on a lattice, obtain an appropriate lattice Hamiltonian by taking the continuous time limit, and from it develop a strong coupling expansion for physical states such as glueballs and hadrons. One would hope that a similar procedure could be applied to gravity as well. It is in order to understand the general ideas and issues better that the gauge theory methods will be described in some detail first.

A lattice regularized form of the gauge action in Eq. (4.124) was given in (Wilson, 1974). The theory is defined on a $d$-dimensional hyper-cubic lattice with lattice spacing $a$, vertices labeled by an index $n$ and directions by $\mu$ (see Fig 4.4). The group elements $U_{n\mu} = \exp iagA_\mu^a T_a$ are defined in the fundamental representation, and reside on the links of the lattice. The pure gauge Euclidean action involves a sum of traces of path-ordered products [with $U_{-\mu}(n+\nu) = U_\mu^\dagger(n)$] of unitary $U_\mu(n)$ matrices around an elementary square loop ("plaquettes", here denoted by $\square$),

$$I[U] = -\frac{a^{4-d}}{4g^2} \sum_{\square} \mathrm{tr} \left[ UUU^\dagger U^\dagger + h.c. \right] \quad . \tag{4.138}$$

From now on we will discuss exclusively the case $d = 4$. The action is locally gauge invariant with respect to the change

$$U_\mu(n) \rightarrow V^\dagger(n) U_\mu(n) V(n+\nu) \quad , \tag{4.139}$$

where $V$ is an arbitrary $SU(N)$ matrix defined on the lattice sites. The product of four $U$ matrices around a plaquettes can be shown, using the Baker-Hausdorff formula $\exp(A)\exp(B) = \exp(A + B + \frac{1}{2}[A,B] + \ldots)$, to give the exponential of the lattice field strength tensor in the limit of small lattice spacing,

$$U_\mu(n)U_\nu(n+\mu)U_{-\mu}(n+\mu+\nu)U_{-\nu}(n+\nu) \approx \exp\left[ia^2 g\mathscr{F}_{\mu\nu}(A)\right] \qquad (4.140)$$

with

$$\mathscr{F}_{\mu\nu}(A) = \partial_\mu A_\nu - \partial_\nu A_\mu + ig\left[A_\mu, A_\nu\right] , \qquad (4.141)$$

and $A_\mu \equiv A^a_\mu T^a$. Taking the trace of the expression in Eq. (4.140) and replacing the sums over lattice points by integrals,

$$a^4 \sum_{\text{sites}} \to \int d^4x , \qquad (4.142)$$

then gives, up to an irrelevant additive constant, the continuum Yang-Mills action of Eq. (4.124).

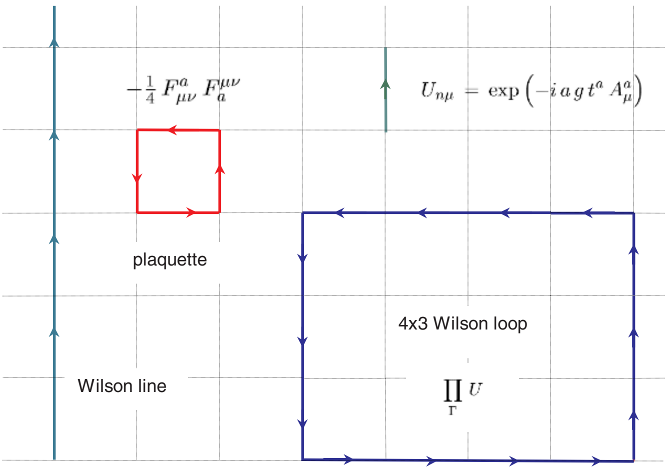



**Fig. 4.4** An elementary plaquette in Wilson's lattice gauge theory. Shown is also a small Wilson loop and a Wilson line, closed by the lattice's periodicity.

The next step is to define the path integral as

$$Z(g^2) = \int [dU_H] \exp\left(-I[U]\right) , \qquad (4.143)$$

where $[dU_H]$ is the Haar measure over the group $SU(N)$, one copy for each lattice link variable $U$. In this formulation only gauge-invariant terms have non-vanishing expectation values, there are no auxiliary conditions, no Faddeev-Popov ghosts, and the gauge volume only contributes a finite factor since there are a finite number of lattice points N$\to \infty$.

A lattice regularized Hamiltonian can be defined on a purely spatial lattice, by taking the zero lattice spacing limit in the time direction (Kogut, 1983). The process can be regarded as the inverse to the one usually follows in deriving the Feynman path integral from the matrix elements of the quantum time evolution operator $e^{-iHt}$. One starts by writing the lattice partition function $Z$ in terms of a product of transfer matrices $T(U_{t+1}, U_t)$ connecting the fields on successive time slices, labeled by $t$, and integrated over all intermediate variables $U$,

$$Z = \operatorname{tr} T^L = \int dU_{t+1}\, dU_t\, dU_{t-1} \dots T(U_{t+1}, U_t)\, T(U_t, U_{t-1})\, T(U_{t-1}, U_{t-2}) \dots \tag{4.144}$$

with $L$ is the total time extent of the lattice. In order to take the zero lattice spacing limit in the time direction, it is convenient to distinguish the lattice spacing in this direction by the symbol $a_0$. Local gauge invariance further allows one to set all the link variables in the time direction to unity, $U_{n0} = 1$, or $A^a_{n0} = 0$ in this lattice version of the temporal gauge. Consequently gauge invariance will need to be imposed as a constraint, which will eventually take the form of a discrete version of Gauss's law.

With this choice of temporal gauge, the action can be decomposed in a part that involves only the spatial plaquettes, and a remainder involving pairs of oppositely oriented link variables separated by a single time step,

$$\begin{aligned} I[U] = & -\frac{1}{4g^2} \sum_{\text{spatial}\,\square} \operatorname{tr}\left[ UUU^\dagger U^\dagger + h.c. \right] \\ & - \frac{1}{4g^2} \sum_{\text{spatial} <nm>} \operatorname{tr}\left[ U_{nm} U^\dagger_{n+a_0 t, m+a_0 t} + h.c. \right] \; . \end{aligned} \tag{4.145}$$

After labeling the gauge variables on two neighboring time slices by $U$ and $U'$, one can write the transfer matrix element $T(U, U')$ as

$$\begin{aligned} \langle U | T | U' \rangle = \exp \Bigg\{ & \frac{1}{4g^2} \sum_{\text{spatial}\,\square} \operatorname{tr}\left[ UUU^\dagger U^\dagger + h.c. \right] + \operatorname{tr}\left[ U'U'U^{\dagger'} U^{\dagger'} + h.c. \right] \\ & + \frac{1}{4g^2} \sum_{\text{links} <nm>} \operatorname{tr}\left[ U_{nm} U^{\dagger'}_{nm} + h.c. \right] \Bigg\} \; . \end{aligned} \tag{4.146}$$

In general the matrix elements of the transfer matrix $T$ have a rather complicated form, but in the limit $a_0 \to 0$ one can write $T \simeq 1 - a_0 H + O(a_0^2)$ and extract from $T$ an expression for the Hamiltonian $H$. But one notices further that the last two terms in Eq. (4.146) involve only temporal loops, with two (gauge fixed) links pointing in the time direction. When written out explicitly, they contain terms of the type

$$\operatorname{tr}[U^\dagger(t+1)\, U(t) + h.c.] \; , \tag{4.147}$$

which, up to irrelevant constant contributions, can be written as combinations of lattice differences in the time direction,

$$\mathrm{tr}\,\frac{1}{a_0}\left[U^\dagger(t+1) - U^\dagger(t)\right]\,\frac{1}{a_0}\left[U(t+1) - U(t)\right] \; . \tag{4.148}$$

In the limit as $a_0 \to 0$ these turn into a combination of time derivatives of the form $\dot{U}^\dagger \dot{U}$, and in this limit the exponent inside the path integral involves therefore the quantity

$$L \;=\; \frac{a}{4g^2}\sum_{\text{links}} \mathrm{tr}\,\dot{U}^\dagger\dot{U} + \sum_{\square}\frac{1}{4ag^2}\,\mathrm{tr}\left[UUU^\dagger U^\dagger + h.c.\right] \; . \tag{4.149}$$

The next step is an elimination of the $\dot{U}$ angular variables in favor of the local generators of rotations (Creutz, 1977). First from the above action one can construct a Hamiltonian in the usual way, by defining

$$H \;=\; \sum_{\text{links nm}}\left(\dot{U}^\dagger_{nm}\,\frac{\partial L}{\partial \dot{U}^\dagger_{nm}} + \dot{U}_{nm}\,\frac{\partial L}{\partial \dot{U}_{nm}}\right) \;-\; L(\dot{U}, U, \dot{U}^\dagger, U^\dagger) \; , \tag{4.150}$$

which in this case gives

$$H \;=\; \frac{a}{4g^2}\sum_{\text{links}} \mathrm{tr}\,\dot{U}^\dagger\dot{U} - \sum_{\square}\frac{1}{4ag^2}\,\mathrm{tr}\left[UUU^\dagger U^\dagger + h.c.\right] \; . \tag{4.151}$$

The $\dot{U}$ variables can now be eliminated by introducing generators of local rotations $E_i^a(\mathbf{n})$, defined on the links (with spatial directions labeled by $i, j = 1,2,3$) and satisfying the commutation relations

$$[E_i^a(\mathbf{n}), U_j(\mathbf{m})] \;=\; T^a\, U_i(\mathbf{n})\,\delta_{ij}\,\delta_{\mathbf{nm}} \; , \tag{4.152}$$

along with the $SU(N)$ generator algebra commutation relation

$$\left[E_i^a(\mathbf{n}), E_j^b(\mathbf{m})\right] \;=\; i\,f^{abc}\,E_i^c(\mathbf{n})\,\delta_{ij}\,\delta_{\mathbf{nm}} \; . \tag{4.153}$$

Since $E^a$ generates local rotations, it can be written explicitly in term of the $U$'s. An infinitesimal local gauge rotation of the $U$ link matrices is achieved by

$$U_i(\mathbf{n}) \;\to\; (1 + i\varepsilon_a T^a)\, U_i(\mathbf{n}) \; . \tag{4.154}$$

The generator for such a symmetry of the original Lagrangian $L$ is by Noether's theorem

$$\begin{aligned} E^a \;&=\; \frac{\partial L}{\partial \dot{U}_{ij}}\,(iT^a U)_{ij} + \frac{\partial L}{\partial \dot{U}^\dagger_{ij}}\,(iT^a U)^\dagger_{ij} \\ &= i\,\frac{a}{4g^2}\left(\mathrm{tr}\,\dot{U}^\dagger T^a U - h.c.\right) \; . \end{aligned} \tag{4.155}$$

In terms of the operators $E^a$ one then has for the first term in the Hamiltonian $H$

$$E^a E^a = \frac{a}{2g^4} \operatorname{tr} \dot{U}^\dagger \dot{U} \ , \tag{4.156}$$

after using the normalization condition on the $SU(N)$ generators $T^a$, $\operatorname{tr} T^a T^b = \frac{1}{2} \delta^{ab}$. This finally gives for the Hamiltonian of Wilson's lattice gauge theory (Kogut and Susskind, 1975)

$$H = \frac{g^2}{2a} \sum_{\text{links}} E^a E^a - \sum_{\square} \frac{1}{4ag^2} \operatorname{tr} \left[ UUU^\dagger U^\dagger + h.c. \right] \ . \tag{4.157}$$

The first term in Eq. (4.157) is the lattice analog of the electric field term $\mathbf{E}^2$, while the second term is a lattice discretized version, involving lattice finite differences, of the magnetic field, $(\nabla \times \mathbf{A})^2$ term. In this picture the analog of Gauss's law is a constraint, which needs to be enforced on physical states at each spatial site $\mathbf{n}$

$$\sum_{i=1}^{6} E_i^a(\mathbf{n}) \, |\Psi\rangle = 0 \ . \tag{4.158}$$

In the special case of the group $SU(2)$, the generators of group rotations in Eq. (4.153) are just the usual angular momentum operators $J_a(\mathbf{n})$, $a = 1,2,3$. The system can be regarded therefore as a collection of quantum rotators, with a kinetic term defined on the links and proportional to $\mathbf{J}^2$ [with eigenvalue $j(j+1)$], and a potential, or link coupling, term. An appropriate basis in the extreme strong coupling limit is then represented by a suitable product of angular momentum eigenstates

$$|\Psi\rangle = \prod_{\mathbf{n},i} |j,m\rangle_{\mathbf{n},i} \ . \tag{4.159}$$

In this limit the $\mathbf{B}^2$ term can be regarded as a perturbation, whose action on the above state can then be determined from the commutation relation in Eq. (4.152).

Even simpler is the Abelian case $U(1)$. Here only one angle variable $\theta_\mu(n)$ survives on each link. In the position representation one writes for the electric field operator $E \equiv J_\theta = -i\partial/\partial\theta$, with integer eigenvalues $m$ and eigenfunctions $e^{im\theta}/\sqrt{2\pi}$. The remainder of the Hamiltonian then involves for each spatial plaquette the term $\cos\theta_\square$ with

$$\theta_\square = \theta_\mu(n) + \theta_\nu(n+\mu) - \theta_\mu(n+\nu) - \theta_\nu(n) \ . \tag{4.160}$$

It is characteristic of this lattice gauge theory model, compact electrodynamics, that the gauge variables are angles. For weak enough coupling $g = e$ one can consider them as variables ranging over the whole real line, and ordinary QED is recovered.

In general, and irrespective of the symmetry group chosen, the ground state in the strong coupling $g^2 \to \infty$ limit has all the $SU(N)$ rotators in their ground state, which for example for $SU(2)$ means all $j = 0$. In this limit the Hamiltonian has the simple form

$$H = \frac{g^2}{2a} \sum_{\text{links}} E_i^a E_i^a \ . \tag{4.161}$$

Then the vacuum is a state for which each link is in a color singlet state

$$E_i^a \, |0\rangle > = 0 \ . \tag{4.162}$$

The lowest order excitation of the vacuum is a "boxciton" state, with one unit of chromo-electric field on each link of an elementary lattice square, with energy

$$E_\Box = 4 \cdot \frac{g^2}{2a} \frac{N^2-1}{2N} \ . \tag{4.163}$$

The mass of the lowest excitation in the theory is usually referred to as the mass gap, the energy gap between the vacuum and the first excited state. Note that if one creates states out of the vacuum by having the Hamiltonian act on it, there is no need to separately enforce the Gauss law constraint, as the states obtained in this way automatically satisfy the constraint.

By the same kind of arguments, the static potential between a quark and an anti-quark pair, separated by a distance $R$, increases linearly with distance causing linear confinement at strong coupling. This is due to the fact that, given two static quarks separated form each other by this distance $R$ [and described by fermion operators $\psi(\mathbf{n})$ and $\psi^\dagger(\mathbf{n})$], a number of link variables $U$ proportional to the distance between the two has to be laid down,

$$\psi^\dagger(\mathbf{n}) \left( \prod_{\mathbf{n} \to \mathbf{n}+R} U \right) \psi(\mathbf{n}+R) \ , \tag{4.164}$$

in order to construct the manifestly gauge invariant color singlet state. Then each link variable which is not in the ground state costs a unit of energy proportional to $1/g^2$.

All of this applies to the strong coupling limit of the theory. Raleigh-Schrödinger perturbation theory can then be used to compute corrections to such results, in principle to arbitrarily high order in $1/g^2$. But ultimately one is interested in the limit $g^2 \to 0$, corresponding to the ultraviolet, asymptotic freedom fixed point of the non-abelian gauge theory, and to the lattice continuum limit $a \to 0$. Therefore in order to recover the original theory's continuum limit, one needs to examine a limit where the mass gap in units of the lattice spacing goes to zero, $am(g) \to 0$.

It is easy to see that this limit corresponds to an infinite correlation length in lattice units, since the correlation length is given by $\xi(g) = 1/m$. The last result can most easily be seen from the Lehman representation of the connected vacuum field correlations in the Euclidean time ($\tau$) direction

$$\langle 0|\,\phi(\tau)\,\phi(0)\,|0\rangle \;=\; \langle 0|\,e^{H\tau}\phi(0)e^{-H\tau}\,\phi(0)\,|0\rangle \;=\; \sum_n \langle 0|\phi(0)|n\rangle|^2\, e^{-(E_n-E_0)\tau} \quad , \tag{4.165}$$

(with $H|n\rangle = E_n|n\rangle$). For sufficiently large Euclidean time one has therefore an asymptotic decay of the correlation function $\langle 0|\phi(\tau)\phi(0)|0\rangle \sim e^{-\tau/\xi}$, with $\xi$ related to the mass gap $m$ by $1/\xi = E_1 - E_0 = m$. The zero lattice spacing limit so described is a crucial step in fully recovering the properties (rotational or Lorentz invariance, asymptotic freedom, massless perturbative gluon excitations) of the original continuum theory.

It is possible to develop a systematic strong coupling $\frac{1}{g^2}$ expansion on the lattice based on the Hamiltonian in Eq. (4.157). Alternatively one can develop a strong coupling expansion starting from the original Wilson formulation in Eqs. (4.138) and (4.143). To this purpose one expands the exponential of the lattice action in terms of group characters $\chi_j$

$$\exp\left[\beta\,\mathrm{tr}(U_\Box)\right] \;=\; F_0 \left[1 + \sum_{j\neq 0} d_j\, c_j(g^2) \chi_j(U_\Box)\right] \quad , \tag{4.166}$$

with $U_\Box = UUU^\dagger U^\dagger$ around a single plaquette labeled by $\Box$, and $\beta = 2N/g^2$. The sum is over all non-trivial irreducible representations of $SU(N)$, here labeled by $j$. The partition function of Eq. (4.143) can then be re-expressed as

$$Z(g^2) \;=\; \int [dU_H] \prod_\Box \left[1 + \sum_{j\neq 0} d_j\, c_j(g^2) \chi_j(U_\Box)\right] \quad . \tag{4.167}$$

Relatively long series for $Z(g^2)$ (and derived quantities), for the string tension and for the glueball correlation function can then be obtained by this method using diagrammatic techniques, such as the linked cluster expansion, borrowed from statistical mechanics (Drouffe and Itzykson, 1978; Münster, 1981; 1982).

## 4.14 Lattice Hamiltonian for Quantum Gravity

In constructing a discrete Hamiltonian for gravity one has to decide first what degrees of freedom one should retain on the lattice. There are a number of possibilities, depending on which continuum theory one chooses to discretize, and at what stage. So, for example, one could start with a discretized version of Cartan's formulation, and define vierbeins and spin connections on a flat hypercubic lattice. Later one could define the transfer matrix for such a theory and construct a suitable Hamiltonian.

Another possibility, which is the one we choose to pursue here, is to use the more economical Regge discretization for gravity, with edge lengths defined on a random lattice as the primary dynamical variables. Even in this specific case several avenues for discretization are possible. One could discretize the theory from the very beginning, while it it is still formulated in terms of an action, and introduce for it lapse and shift functions, extrinsic and intrinsic discrete curvatures, etc. Alternatively one could try to discretize the continuum Wheeler-deWitt equation directly, a procedure that makes sense in the lattice formulation, as these equations are still given in terms of geometric objects, for which the Regge theory is well suited. It is the latter approach which we will outline here.

The starting point for the following discussion is therefore the Wheeler-DeWitt equation for pure gravity in the absence of matter, Eq. (4.72),

$$\left\{ -(16\pi G)^2 \, G_{ij,kl}(\mathbf{x}) \, \frac{\delta^2}{\delta g_{ij}(\mathbf{x}) \, \delta g_{kl}(\mathbf{x})} \; - \; \sqrt{g(\mathbf{x})} \; \left( {}^3\!R(\mathbf{x}) - 2\lambda \right) \right\} \Psi[g_{ij}(\mathbf{x})] \; = \; 0 \; , \tag{4.168}$$

and the diffeomorphism constraint of Eq. (4.73),

$$\left\{ 2\, i \, g_{ij}(\mathbf{x}) \, \nabla_k(\mathbf{x}) \, \frac{\delta}{\delta g_{jk}(\mathbf{x})} \right\} \Psi[g_{ij}(\mathbf{x})] \; = \; 0 \; . \tag{4.169}$$

Note that there is a constraint on the state $|\Psi\rangle$ at every $\mathbf{x}$, each of the form $\hat{H}(\mathbf{x})\,|\Psi\rangle = 0$ and $\hat{H}_i\,(\mathbf{x})|\Psi\rangle = 0$.

On a simplicial lattice (see Sect. 6.1 for a discussion of the lattice formulation for gravity) one knows that deformations of the squared edge lengths are linearly related to deformations of the induced metric, within a given simplex $s$ and based on a vertex 0, as discussed elsewhere here, for example in Eq. (6.3),

$$g_{ij}(s) \; = \; \tfrac{1}{2} \left( l_{0i}^2 + l_{0j}^2 - l_{ij}^2 \right) \; . \tag{4.170}$$

Note that in the following discussion only edges and volumes along the spatial direction are involved. Furthermore, one can introduce in a natural way a lattice analog of the DeWitt supermetric of Eq. (4.66), by writing

$$\| \, \delta l^2 \, \|^2 \; = \; \sum_{ij} G^{ij}(l^2) \; \delta l_i^2 \; \delta l_j^2 \; , \tag{4.171}$$

with the quantity $G^{ij}(l^2)$ suitably defined on the space of squared edge lengths (Lund and Regge, 1974; Hartle, Miller and Williams, 1997). Through a straightforward exercise of varying the squared volume of a given simplex $s$ in $d$ dimensions

$$V^2(s) \; = \; \left( \tfrac{1}{d!} \right)^2 \det g_{ij}[l^2(s)] \; , \tag{4.172}$$

to quadratic order in the metric (on the r.h.s.), or in the squared edge lengths (on the l.h.s.), one finds the identity

$$\frac{1}{V(l^2)} \sum_{ij} \frac{\partial^2 V^2(l^2)}{\partial l_i^2 \partial l_j^2} \; \delta l_i^2 \; \delta l_j^2 \; = \; \tfrac{1}{d!} \sqrt{\det(g_{ij})} \left[ \, g^{ij} g^{kl} \delta g_{ij} \delta g_{kl} - g^{ij} g^{kl} \delta g_{jk} \delta g_{li} \, \right] \; . \tag{4.173}$$

The r.h.s. of this equation contains precisely the expression appearing in the continuum supermetric of Eq. (2.14) or Eq. (4.66) (for a specific choice of the parameter $\alpha = -2$), while the l.h.s. contains the sought-for lattice supermetric. One is therefore led to the identification

$$G^{ij}(l^2) \; = \; - \, d! \sum_s \frac{1}{V(s)} \, \frac{\partial^2 \, V^2(s)}{\partial l_i^2 \, \partial l_j^2} \; . \tag{4.174}$$

In spite of the appearance of a sum over simplices $s$, $G^{ij}(l^2)$ is quite local (in correspondence with the continuum, where it is ultra-local), since the derivatives on the r.h.s. vanish when the squared edge lengths in question are not part of the same simplex. The sum over $s$ therefore only extends over those few tetrahedra which contain either the $i$ or the $j$ edge.

At this point one is finally ready to write a lattice analog of the Wheeler-DeWitt equation for pure gravity, which reads

$$\left\{ - \, (16 \pi G)^2 \, G_{ij}(l^2) \, \frac{\partial^2}{\partial l_i^2 \, \partial l_j^2} \; - \; \sqrt{g(l^2)} \, \left[ \, {}^3\!R(l^2) \, - \, 2 \lambda \, \right] \right\} \, \Psi[l^2] \; = \; 0 \;\; , \tag{4.175}$$

with $G_{ij}(l^2)$ the inverse of the matrix $G^{ij}(l^2)$. Note that the lattice supermetric is dimensionful, $G_{ij} \sim l^{4-d}$ and $G^{ij} \sim l^{d-4}$ in $d$ spacetime dimensions, and it might be useful from now on to introduce a lattice spacing $a$ (or momentum cutoff $\Lambda = 1/a$) and express all dimensionful quantities $(G, \lambda, l_i)$ in terms of this lattice spacing.

One notes that Eqs. (4.168) or (4.175) express a constraint equation for each "point" in space. On the lattice these points are replaced by a set of edge labels $i$, with a few edges clustered around each vertex, in a way that depends on the dimensionality and the local lattice coordination number. To be more specific, the first term in Eq. (4.175) contains derivatives with respect to edges $i$ and $j$ connected by a matrix element $G_{ij}$ which is nonzero only if $i$ and $j$ are close to each other, essentially nearest neighbor. One would therefore expect that the first term could be represented by a sum of edge contributions all from within one $d-1$-simplex (a tetrahedron in three dimensions) $s$. The second term containing ${}^3\!R(l^2)$ in Eq. (4.175) is also local in the edge lengths: it only involves a handful of edge lengths which enter into the definition of areas, volumes and angles around the point $\mathbf{x}$, and follows from the fact that the local curvature at the original point $\mathbf{x}$ is completely determined by the values of the edge lengths clustered around $i$ and $j$. Apart from some geometric factors, it describes, through a deficit angle $\delta_h$, the parallel transport of a vector around an elementary dual lattice loop. One would expect that it should be adequate to represent this second term by a sum over contributions over all $d-3$-dimensional hinges (edges in three dimensions) $h$ attached to the simplex $s$, giving therefore in three dimensions

$$\left\{ -(16\pi G)^2 \sum_{i,j \subset s} G_{ij}(s) \frac{\partial^2}{\partial l_i^2 \partial l_j^2} - \sum_{h \subset s} l_h \delta_h + 2\lambda V_s \right\} \Psi[l^2] = 0 \ . \quad (4.176)$$

Here $\delta_h$ is the deficit angle at the hinge $h$, $l_h$ the corresponding edge length, $V_s = \sqrt{g(s)}$ the volume of the tetrahedron centered on $s$, and $G_{ij}(s)$ obtained from Eq. (4.63)

$$G_{ij,kl}(s) = \tfrac{1}{2} g^{-1/2}(s) \left[ g_{ik}(s) g_{jl}(s) + g_{il}(s) g_{jk}(s) - g_{ij}(s) g_{kl}(s) \right] \ , \quad (4.177)$$

with the induced metric $g_{ij}(s)$ within a simplex $s$ given in Eq. (4.170).

It is in fact quite encouraging that the discrete equation in Eqs. (4.175) and (4.176) is very similar to what one would derive in Regge lattice gravity by doing the $3+1$ split of the lattice metric carefully from the very beginning (Piran and Williams, 1986; Williams and Tuckey, 1990). These authors also derived a lattice Hamiltonian in three dimensions, written in terms of lattice momenta conjugate to the edge length variables. In this formulation the Hamiltonian constraint equations have the form

$$\begin{aligned} H_n &= \tfrac{1}{4} \sum_{\alpha \in n} G_{ij}^{(\alpha)} \pi^i \pi^j - \sum_{\beta \in n} \sqrt{g_\beta}\, \delta_\beta \\ &= \tfrac{1}{4} \sum_{\alpha \in n} \frac{1}{V_\alpha} \left[ (\mathrm{tr}\, \pi^2)_\alpha - \tfrac{1}{2} (\mathrm{tr}\, \pi)^2_\alpha \right] - \sum_{\beta \in n} \sqrt{g_\beta}\, \delta_\beta \ , \\ &= 0 \end{aligned} \quad (4.178)$$

with $H_n$ defined on the lattice site $n$. The sum $\sum_{\alpha \in n}$ extends over neighboring tetrahedra labeled by $\alpha$, whereas the sum $\sum_{\beta \in n}$ extends over neighboring edges, here labeled by $\beta$. $G_{ij}^{(\alpha)}$ the inverse of the DeWitt supermetric at the site $\alpha$, and $\delta_\beta$ the deficit angle around the edge $\beta$. $\sqrt{g_\beta}$ is the dual (Voronoi) volume associated with the edge $\beta$. In this discrete formulation there is also an additional semi-local constraint at each vertex $n$, corresponding to the continuum constraint

$$\frac{1}{\sqrt{g}} \left[ \mathrm{tr}\, \pi^2 - \tfrac{1}{2} (\mathrm{tr}\, \pi)^2 \right] - \sqrt{g}\, {}^3 R = 0 \ . \quad (4.179)$$

The lattice Wheeler-DeWitt equation of Eq. (4.175) has an interesting structure, which is in part reminiscent of the Hamiltonian for lattice gauge theories, Eq. (4.157). The first, local kinetic term is the gravitational analogue of the electric field term $E_a^2$. It contains momenta which can be considered as conjugate to the squared edge length variables. The second local term involving ${}^3R(l^2)$ is the analog of the magnetic $(\nabla \times A_a)^2$ term in the lattice Hamiltonian of Eq. (4.175). In the absence of a cosmological constant term, the first and second term have opposite sign, and need to cancel out exactly on physical states in order to give $H(\mathbf{x})\Psi = 0$. On the other hand, the last term proportional to $\lambda$ has no gauge theory analogy, and is, as expected, genuinely gravitational.

It is important to note that the squared edge lengths take on only positive values $l_i^2 > 0$, a fact that would seem to imply that the wavefunction has to vanish when the edge lengths do, $\Psi(l^2 = 0) \simeq 0$. In addition one has some rather complicated constraints on the squared edge lengths, due to the triangle inequalities. These ensure that the areas of triangles and the volumes of tetrahedra are always positive. As a result one would expect an average soft local upper bound on the squared edge lengths of the type $l_i^2 \lesssim l_0^2$ where $l_0$ is an average edge length, $\langle l_i^2 \rangle = l_0^2$. The term "soft" refers to the fact that while large values for the edge lengths are possible, these should nevertheless enter with a relatively small probability, due to the small phase space available in this region.

These considerations have some consequences already in the strong coupling limit of the theory. For sufficiently strong coupling (large Newton constant $G$) the first term in Eq. (4.175) is dominant, which shows again some similarity with what one finds for non-abelian gauge theories for large $g$, Eq. (4.157). It is then easy to see, both from the constraint $l_i > 0$ and the triangle inequalities, that the spectrum of this operator is discrete. In particular the mass gap, the spacing between the lowest eigenvalue and the first excited state, is of the same order as the ultraviolet cutoff. One can argue that this is in fact a general feature of the strong coupling theory, where one is far removed from a lattice continuum limit. The latter has to be taken in the vicinity of a non-trivial ultraviolet fixed point, if such a fixed point can be found. One would then anticipate that the excitation spectrum would become denser as one approaches the lattice continuum limit, in accordance with the existence of a massless spin two particle in this limit.

Note that in the lattice theory the operator ordering ambiguity has not gone away either: in principle one would have to check that different orderings give the same physical results, whichever way those are defined (for example in terms of vacuum expectation values of invariant operators, or quantum correlations of invariant operators at fixed geodesic distance along the spatial directions).

Irrespective of its specific form, it is in general possible to simplify the lattice Hamiltonian constraint in Eqs. (4.175) and (4.176) by using scaling arguments, as one does often in ordinary non-relativistic quantum mechanics. After setting for the scaled cosmological constant $\lambda = 8\pi G \lambda_0$ and dividing the equation out by common factors, it can be recast in the slightly simpler form

$$\left\{ -\alpha\, a^6 \cdot \frac{1}{\sqrt{g(l^2)}}\, G_{ij}(l^2)\, \frac{\partial^2}{\partial l_i^2\, \partial l_j^2} - \beta\, a^2 \cdot {}^3\!R(l^2) + 1 \right\} \Psi[l^2] \;=\; 0 \;, \qquad (4.180)$$

where one finds it useful to define a dimensionless Newton's constant, as measured in units of the cutoff $\bar{G} \equiv 16\pi G/a^2$, and a dimensionless cosmological constant $\bar{\lambda}_0 \equiv \lambda_0 a^4$, so that in the above equation one has $\alpha = \bar{G}/\bar{\lambda}_0$ and $\beta = 1/\bar{G}\bar{\lambda}_0$. Furthermore the edge lengths have been rescaled so as to be able to set $\lambda_0 = 1$ in lattice units (it is clear from the original gravitational action that the cosmological constant $\lambda_0$ simply multiplies the total spacetime volume, thereby just shifting around the overall scale for the problem). Schematically Eq. (4.176) is therefore of the form

$$\left\{ -\bar{G}\Delta_q - \frac{1}{\bar{G}}\, {}^3\!R(q) + 1 \right\} \Psi[q] \;=\; 0 \;, \tag{4.181}$$

with $\Delta_q$ a discretized form of the covariant Laplacian, acting locally on the function space of the $q = l^2$ variables; on near-transverse traceless modes it is expected to have positive eigenvalues. Furthermore, at this point the similarity with the lattice Hamiltonian for non-abelian gauge theories in Eq. (4.157) has become evident.

Due to the triangle inequalities, finding a solution of Eq. (4.180) for all lattice points might not be easy; in principle it could be done numerically. For $N$ lattice $q$ variables, the solution for $\Psi(q)$ is expected to be in general a linear combination of $N$ wave functions. But if one is only interested in the lowest $p^2 \approx 0$ excitations of the theory, one could perhaps approximate the Laplacian term by its lowest eigenvalue $\sim (\pi/L)^2$ where $L$ is the linear size of the spatial system (for a spatial volume ${}^3V$ one would use $L \simeq ({}^3V)^{1/3}$). Furthermore the local three-curvature operator ${}^3\!R(q)$ involves an elementary loop on the lattice, with size of the order of the average lattice spacing $l_0$. From dimensional arguments one would expect this term on the average to contribute ${}^3\!R \simeq c_0/l_0^2 + c_1/L^2$, the first piece representing a subtraction, and the second one a correction dependent on the boundary conditions in $\mathbf{x}$. Inserting this expression into Eq. (4.181) one finds

$$c_0 \;=\; l_0^2\,\bar{G} \qquad c_1 \;=\; \pi^2\,\bar{G}^2. \tag{4.182}$$

The first condition amounts to requiring a critical value for $\bar{G}$, reminiscent of the ultraviolet fixed point condition for $G$ in the $2+\varepsilon$ expansion. If the theory develops self-consistently a non-perturbative scale $\xi$ by a mechanism analogous to dimensional transmutation, then one would replace $L \to \xi$ in the above expressions.

# Chapter 5
# Semiclassical Gravity

## 5.1 Cosmological Wavefunctions

The basic idea of semiclassical gravity is that for energies well below the Planck scale it is adequate to treat matter as quantum mechanical, but keep gravity classical. The underlying assumption is that quantum gravity effects do not extend to large distances, and that the Planck length $l_P = \sqrt{G}$ is the only length that can set the scale for such quantum corrections, and not, for example, a non-vanishing cosmological constant $\lambda$. One is therefore lead to consider quantum fields embedded in a general curved background, whose local curvature radii are generally expected to be much larger than the Planck length. The field equations will then be written as

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R + \lambda\, g_{\mu\nu} \;=\; 8\pi G \, \langle T_{\mu\nu} \rangle \;\; , \tag{5.1}$$

where $\langle T_{\mu\nu} \rangle$ is some suitable local quantum average of the energy-momentum tensor for matter fields.

One possible framework for discussing semiclassical gravity is the Wheeler-DeWitt equation. Generally this equation will exhibit, even for pure quantum gravity without matter, a vast number of solutions. These will in part be restricted by the imposition of suitable, physically or otherwise, motivated, boundary conditions. In the end the problem remains of how to concretely construct such wave functionals, beyond the simplest minisuperspace approximation. Another possibility is to use the Feynman path integral approach. There the wavefunction is represented by an Euclidean functional integral over four-metrics and matter fields, weighted by the Euclidean action $\hat{I}$ for gravity plus with matter (Hartle and Hawking, 1983),

$$\Psi[g_{ij}, \phi] \;=\; \int_M [d g_{\mu\nu}]\, [d\phi] \, \exp\left\{ - \frac{1}{\hbar} \, \hat{I}(g_{\mu\nu}, \phi) \right\} \;\; . \tag{5.2}$$

The Euclidean theory is obtained as usual from the Lorentzian path integral by a standard Wick rotation $t = -i\tau$, thereby continuing the spacetime metric to a Riemannian signature. The sign of the Wick rotation is not arbitrary, since the action for ordinary matter fields needs to be reproduced correctly. In the above expression the Feynman path integral is defined over those metrics $g_{\mu\nu}(x)$ and matter fields $\phi(x)$ which have the hypersurface $\partial M$ as a boundary, where the fields take on values specified in the argument of the wavefunction $\Psi$. In addition the fields are possibly subject to some other sort of boundary condition at some earlier time. One of the earliest discussions of the connection between the canonical approach to gravity and the Feynman sum over histories method can be found in (Leutwyler, 1964) and (DeWitt, 1967a,b,c).

An essential ingredient of the proposal is that only closed universes should be included in the path integral. This if often referred to as the no boundary (or perhaps more appropriately, the one boundary) proposal. The motivation is to pursue an analogy with ordinary field theory, where vacuum to vacuum amplitudes are obtained in the limit of large (Euclidean) time. Thus the integral is intended to extend over an appropriate class of Euclidean compact four-geometries with compact boundary on which the induced metric is $h_{ij}$, and an appropriate class of Euclidean matter field configurations which match the values given on the boundary. The wave functional so obtained is usually denoted by $\Psi_0$ and referred to as the state of minimum excitation. It would seem inappropriate to refer to it as a state of zero energy, as there is no natural definition of energy, just as there is no natural definition of time, and furthermore one expects the total energy of a closed universe to add up to zero when both matter (positive) and gravitational (negative) contributions are added up.

It is clear that there are several advantages to this formalism (Halliwell and Louko, 1989a,b,c; Halliwell and Hartle, 1990). Since it is based on path integrals, one can develop a systematic semi-classical expansion to compute the wavefunction, or even non-pertubative methods based on saddle point expansions or importance samplings. Secondly, it gives a relatively unambiguous prescription for constructing a wave functional which is a solution to the Wheeler-DeWitt equation, including an explicit specification of the boundary conditions. Finally it should shed light on the issue of cosmological singularities, since it allows the calculation of quantum-mechanical amplitudes (which might be finite or zero) to transition across those singularities. The problem is of course that, as discussed previously, the path integral is affected by severe ultraviolet divergences due to the fact that gravity is not perturbatively renormalizable, and the semiclassical expansion does little to remove these divergences. Secondly, the Euclidean path integral is unbounded from below, and some sort of prescription for integrating over the metric has to be given, such as integrating over complex conformal factors, and selecting the appropriate complex extrema about which to expand the metric.

A close relationship between the Schrödinger equation (to which the Wheeler-DeWitt equation itself is closely related) and the Feynman path integral arises already in non-relativistic quantum mechanics (Feynman and Hibbs, 1963). The non-relativistic wavefunction for a spinless particle $\Psi(\mathbf{x},t)$ satisfies the integral equation

$$\Psi(\mathbf{x},t) = \int d^3\mathbf{x} \; G(\mathbf{x},t;\mathbf{x}',t') \, \Psi(\mathbf{x}',t') \; . \tag{5.3}$$

Physically the above expression means that the total quantum-mechanical amplitude for a particle to arrive at $(\mathbf{x},t)$ is a sum over all possible values $\mathbf{x}'$ of the total amplitude to arrive at the $(\mathbf{x}',t')$ (which is given by $\Psi(\mathbf{x}',t')$, multiplied by the amplitude to go from $\mathbf{x}$ to $\mathbf{x}'$, which is given by the propagator $G(\mathbf{x},t;\mathbf{x}',t')$. The propagator itself corresponds to a special situation: the amplitude where the particle started out at precisely $(\mathbf{x}',t')$.

Let us recall that in Feynman's formulation of quantum mechanics the propagator $G$ is expressed as a sum over all paths connecting initial and final points, weighted by an action $I$,

$$G(\mathbf{x}_f,t_f;\mathbf{x}_i,t_i) \;\equiv\; \langle\, \mathbf{x}_f,t_f\,|\,\mathbf{x}_i,t_i\,\rangle \;=\; \int_{\mathbf{x}_i(t_i)}^{\mathbf{x}_f(t_f)} [d\mathbf{x}(t)]\; \exp\left\{\frac{i}{\hbar}\, I[\mathbf{x}(t)]\right\} \;. \qquad (5.4)$$

The paths $\mathbf{x}(t)$ contributing to the integral are known to be continuous, but not necessarily differentiable (one can give arguments in support of the statement that differentiable paths have measure zero), which requires in general that the above integral be carefully defined on a lattice of $N$ points with spacing $a$, with the limit $a \rightarrow 0$, $N \rightarrow \infty$ taken at the end.

Returning to the gravitational case, the question arises then of how to compute the path integral in Eq. (5.2), even in the absence of matter, and what boundary conditions need to be imposed. In gravity the analogue of Eq. (5.4) is the quantum mechanical amplitude

$$\langle\, g_{ij}^{(f)},\phi^{(f)}\,|\,g_{ij}^{(i)},\phi^{(i)}\,\rangle \;=\; \int_{g_{ij}^{(i)},\phi^{(i)}}^{g_{ij}^{(j)},\phi^{(f)}} [dg_{\mu\nu}]\,[d\phi]\,\exp\left\{-\frac{1}{\hbar}\,\hat{I}(g_{\mu\nu},\phi)\right\} \;, \qquad (5.5)$$

where the functional integral is over all four-geometries that match the initial $(i)$ and final $(f)$ field configurations on the two spacelike surfaces. One noteworthy aspect of such gravitational amplitudes is the fact that, since all intervening four-geometries are summed over, there is no notion of unique intervening proper time interval: the proper time distance between the two hypersurfaces will depend on the specific choice of interpolating four-geometry in the ensemble.

As mentioned previously, in computing the *ground state* wave functional $\Psi$ of Eq. (5.2) the proposal has been put forward to functionally integrate over all metrics associated with *compact* Euclidean four-geometries specified by $g_{\mu\nu}$, with a given three-metric $g_{ij}$ on the boundary. For obvious reasons this is usually referred to as the "no-boundary" proposal. It elegantly bypasses the issue of having to specify a boundary or continuity condition on cosmological singularities, by suitably restricting the choice of geometries at "initial" times. In this approach the wave functional for pure gravity is given by (from now on we set again $\hbar = 1$)

$$\Psi[g_{ij}] \;=\; \int_M [dg_{\mu\nu}]\,\exp\left\{-\hat{I}(g_{\mu\nu})\right\} \;, \qquad (5.6)$$

with an Euclidean action containing both volume $(M)$ and boundary $(\partial M)$ terms,

$$\hat{I}[g_{\mu\nu}] \;=\; -\frac{1}{16\pi G}\int_M d^4x\sqrt{g}\,(R-2\lambda) - \frac{1}{8\pi G}\int_{\partial M} d^3x\sqrt{g_{ij}}\,K \;, \tag{5.7}$$

where $K$ is the trace of the extrinsic curvature on the boundary. Note that the inner product between two wave functionals is obtained by gluing together two wave functionals and integrating over the fields on their common boundary, which is located on a spatial hypersurface,

$$\begin{aligned} \langle\, \Psi[g_{ij},\phi]\,|\,\Psi'[g_{ij},\phi]\,\rangle &= \int [dg_{ij}]\,[d\phi]\,\bar{\Psi}[g_{ij},\phi]\,\Psi'[g_{ij},\phi] \\ &= \int_{M,M'} [dg_{\mu\nu}]\,[d\phi]\,\exp\left\{-\frac{1}{\hbar}\,\hat{I}(g_{\mu\nu},\phi)\right\} \;, \end{aligned} \tag{5.8}$$

is interpreted as an integral over all of space, half of it to the left of the spacelike hypersurface ($M$), and half of it to the right ($M'$).

To evaluate the path integral defining $\Psi$ one option is to use the method of steepest descent, which is equivalent to the semi-classical, or $WKB$, approximation and produces an answer in the form of an expansion in powers of $\hbar$. In quantum field theory such an expansion is equivalent to expanding in the number of loop diagrams associated with perturbative Feynman diagrams. The leading term to the wave functional $\Psi$ is then the classical contributions, and the next correction is determined by the quadratic quantum fluctuations around the chosen background. After the fluctuations are integrated over via a Gaussian integral formula, one obtains a functional determinant, whose effects then determine the leading quantum correction. The wave functional will then take the form

$$\Psi[g_{ij}] \;=\; P[g_{ij}]\,\exp\left\{-\hat{I}_{cl}(g_{ij})\right\} \;, \tag{5.9}$$

where $\hat{I}_{cl}(g_{ij})$ is the classical Euclidean action associated with the saddle point (if there is more than one, an additional sum will be required), and $P[g_{ij}]$ is a prefactor whose form is determined by the expansion of the Euclidean action to quadratic order and the subsequent functional integration. One sets for the four-metric

$$g_{\mu\nu} \;\rightarrow\; \bar{g}_{\mu\nu} \;=\; g_{\mu\nu} + h_{\mu\nu} \;, \tag{5.10}$$

where $h_{\mu\nu}$ is a perturbation of the saddle-point four-metric $g_{\mu\nu}$ (which therefore satisfies $\delta\hat{I}/\delta g_{\mu\nu} = 0$), vanishing on the boundary. The prefactor $P$ in Eq. (5.9) is then given formally by the integral,

$$P[g_{ij}] \;=\; \int_M [dh_{\mu\nu}]\,\exp\left\{-\hat{I}_2(h_{\mu\nu})\right\} \;, \tag{5.11}$$

with $\hat{I}_2$ the contribution to the action $\hat{I}$ quadratic in the metric perturbation $h_{\mu\nu}$, and itself also a function of the background metric $g_{\mu\nu}$. The integral over the $h_{\mu\nu}$ fluctuations will involve zero modes from the gauge degrees of freedom, which will need to be factored out. Or, equivalently, one needs to restrict the functional integration to physical degrees of freedom.

To carry the program through in specific cases, one first needs a choice of suitable background metric, then an explicit expression for the second variation of the action around this metric, and finally a procedure for evaluating the contribution of the quantum fluctuation around this background. With the metric written as in Eq. (5.10) one finds for the second variation of the action for $\lambda = 0$

$$\hat{I}_2[h] \;=\; -\frac{1}{32\pi G}\int d^4x\,\sqrt{g}\,h^{\mu\nu}\,G_{\mu\nu} \tag{5.12}$$

with

$$G_{\mu\nu} \;=\; \nabla_\lambda\nabla^\lambda\bar{h}_{\mu\nu} + g_{\mu\nu}\nabla_\lambda\nabla_\sigma\bar{h}^{\lambda\sigma} - \nabla_\mu\nabla_\lambda\bar{h}^\lambda{}_\nu - \nabla_\nu\nabla_\lambda\bar{h}^\lambda{}_\mu \;, \tag{5.13}$$

for a background metric satisfying $R_{\mu\nu} = 0$, and up to total derivatives. Here $\nabla_\mu$ is the covariant derivative with respect to the background metric $g_{\mu\nu}$, and $\bar{h}_{\mu\nu}$ the trace-reversed metric perturbation,

$$\bar{h}_{\mu\nu} \;=\; h_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}\,h^\mu{}_\nu \;. \tag{5.14}$$

There is no surface contribution in $\hat{I}_2[h]$, since it would have to be of the form $h\nabla h$ with $h_{\mu\nu}$, thus vanishing on the boundary.

The path integral over the $h_{\mu\nu}$ variables in Eq. (5.11) suffers from the usual problem of configuration over-counting due to the gauge freedom in the metric $h$. Specifically, the action $\hat{I}_2[h]$ is invariant under local gauge variations of $h_{\mu\nu}$

$$h_{\mu\nu} \;\to h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu \;, \tag{5.15}$$

such that the gauge function $\xi_\mu$ vanishes on the boundary. In order to avoid a divergence in the integration over the quantum fluctuations $h_{\mu\nu}$, one needs to restrict the functional integration over physically distinct metrics. One way of doing this is to introduce a gauge-fixing term, and the associated Faddeev-Popov determinant. A possible gauge condition would be

$$f^\nu \;=\; \nabla_\mu\left(h^{\mu\nu} - \beta\,g^{\mu\nu}\,h^\lambda{}_\lambda\right) \;, \tag{5.16}$$

with $f^\nu$ some prescribed vector function on the background manifold. The corresponding Faddeev-Popov determinant would then involve a differential operator, determined by the derivative of the gauge condition with respect to the gauge parameter. For the gauge condition in Eq. (5.16) the relevant operator $C_{\mu\nu}$ is

$$C_{\mu\nu}(h)\,\xi^\nu \;=\; -\nabla^\lambda\nabla_\lambda\,\xi_\mu - R_{\mu\nu}\,\xi^\nu + (2\beta - 1)\,\nabla_\mu\nabla_\nu\,\xi^\nu \;. \tag{5.17}$$

The above procedure works well with manifolds without boundaries (Gibbons and Perry, 1978). But it runs into technical problems when boundaries are present, which makes the procedure less transparent.

An alternative way of doing the calculation is to use a hybrid Hamiltonian formalism with an explicit choice of gauge, and later integrate (with some suitable functional measure) over the physical degrees only (Schleich, 1985). The procedure relies on establishing a correspondence between the covariant path integral approach and some aspects of the machinery, notably the constraint equations, of canonical quantum gravity. Such a procedure is possible because the Feynman path integral, already for a non-relativistic particle, can be written in the equivalent form

$$\begin{aligned} G(\mathbf{q}_f,t_f;\mathbf{q}_i,t_i) &= \langle\, \mathbf{q}_f,t_f \,|\, \mathbf{q}_i,t_i \,\rangle \\ &= \int_{\mathbf{q}_i(t_i)}^{\mathbf{q}_f(t_f)} \left[ \frac{d\mathbf{q}(t)\, d\mathbf{p}(t)}{2\pi\hbar} \right] \exp\left\{ \frac{i}{\hbar} \int_{t_i}^{t_f} dt \, [\, \mathbf{p}\,\dot{\mathbf{q}} - H(\mathbf{p},\mathbf{q})] \right\} \;, \end{aligned} \tag{5.18}$$

which gives the expression in Eq. (5.4) after integrating out the $p$'s (see, for example, Abers and Lee, 1973). Some of the earliest discussions on the connection between the canonical and functional-integral approaches to quantum gravity, and of some of the subtleties that arise in such a correspondence, can be found in (Leutwyler, 1964; Faddeev and Popov, 1973; Fradkin and Vilkovisky, 1973).

## 5.2 Semiclassical Expansion

To proceed further it will help to be a bit more specific about the boundary geometry in Eq. (5.6). If one wants to investigate the quantum mechanical behavior of closed cosmologies in the vicinity of classical singularities, one is naturally led to consider the behavior of the wave functional $\Psi$ at small volume, where corrections to the classical solution can be large and new effects can arise. In the semiclassical expansion of quantum gravity it is therefore natural to consider as saddle points of the Euclidean action solutions whose boundary geometry has the shape of a three-sphere. Indeed in the case of a positive cosmological constant $\lambda$ any regular Euclidean solution of the field equations is necessarily compact, with the solution of greatest symmetry corresponding to the four-sphere or radius $\sqrt{3/\lambda}$.

In the following it will be assumed (Schleich, 1985) that the boundary geometry is a three-sphere with radius $a$, such that

$$ds^2 \;=\; a^2\, \hat{g}_{ij}\, d\phi^i\, d\phi^j \;, \tag{5.19}$$

where $a$ is the radius of the three-sphere, and $\hat{g}_{ij}$ is the metric on the unit three-sphere. With such a boundary condition, the compact extrema of the action in Eq. (5.7) are sections of Euclidean de Sitter space, with

$$ds^2 \;=\; d\theta^2 + r_0^2 \sin^2(\theta/r_0)\, \hat{g}_{ij}\, d\phi^i\, d\phi^j \;, \tag{5.20}$$

with $\theta$ ranging from 0 to $\pi$, $r_0 = \sqrt{3/\lambda}$ and $a < r_0$. For sufficiently small boundary three-geometries, $a \ll r_0$, the metric interior to the boundary approaches that of Euclidean flat space (with $\theta \to t$),

$$ds^2 \;=\; dt^2 + t^2 \, \hat{g}_{ij} \, d\phi^i \, d\phi^j \;\; . \tag{5.21}$$

In this limit one would expect therefore that a calculation of the quantum corrections could be carried out in a *flat* background with $\lambda = 0$, with higher order corrections then involving the ratio between the two radii in question, $O(a^2/r_0^2)$. Then in the expansion of the metric of Eq. (5.10) one just has $g_{\mu\nu} = \eta_{\mu\nu}$, the flat metric.

The first step in such a program is therefore a re-writing of the expansion of the Euclidean action to quadratic order in the $h$ field, and specifically of $\hat{I}_2[h]$ in Eqs. (5.12) and (5.13), in terms of the quantities $t$ and $\hat{g}_{ij}$ appearing in the metric in Eq. (5.21). One finds

$$\hat{I}_2[h] \;=\; -\frac{1}{16\pi G} \int d^4x \left\{ \sqrt{g}\, p^{ij}\, \partial_t h_{ij} - \mathscr{H}_T - \mathscr{H}_L - h_{00}\,\mathscr{C} - h_{0i}\,\mathscr{C}^I \right\} \;\; , \tag{5.22}$$

with background three-metric $g_{ij} = t^2 \hat{g}_{ij}$, $\sqrt{g}$ the square root of the determinant of this $g_{ij}$, $\pi^{ij} = \sqrt{g} p^{ij}$ the momentum conjugate to the quantum fluctuation $h_{ij}$, and

$$\begin{aligned}
\mathscr{H}_T &= \sqrt{g} \left[ p^{ij} p_{ij} + 2t^{-1} p^{ij} h_{ij} - \tfrac{1}{4} \left( \nabla_k h_{ij} \nabla^k h^{ij} + 2t^{-2} h^{ij} h_{ij} \right) \right] \\
\mathscr{H}_L &= \tfrac{1}{2} \sqrt{g} \left[ p^i{}_i \, p^j{}_j + t^{-1} p^i{}_i h^j{}_j \right. \\
&\qquad \left. + \tfrac{1}{2} \left( \nabla_k h^j{}_j \nabla^k h^j{}_j + 2 \nabla_i h^{ij} (\nabla_k h^k{}_j - \nabla_j h^k{}_k) - \tfrac{1}{2}\, 7 t^{-2} (h^i{}_i)^2 \right) \right] \\
\mathscr{C} &= \tfrac{1}{2} \sqrt{g} \left[ \nabla_i \nabla_j h^{ij} - \nabla^k \nabla_k h^i{}_i + 2t^{-1} p^i{}_i - \tfrac{1}{2}\, 5 t^{-2} h^i{}_i \right] \\
\mathscr{C}^i &= -\sqrt{g} \left[ 2 \nabla_j p^{ji} - t^{-1} \nabla^i h^j{}_j \right] \;\; .
\end{aligned} \tag{5.23}$$

The metric components $h_{00}$ and $h_{0i}$ act as Lagrange multiplier, giving four constraints

$$\mathscr{C} \;=\; 0 \qquad \mathscr{C}^i \;=\; 0 \;\; . \tag{5.24}$$

The physical subspace is obtained by imposing a gauge condition, in the case at hand one that leads to a substantial simplification of the problem. The gauge condition is

$$\nabla^i \, h_{ij} \;=\; 0 \qquad h^i{}_i \;=\; 0 \;\; , \tag{5.25}$$

and restricts the functional integration over transverse-traceless ($TT$) modes only. Due the decomposition of $\hat{I}_2$ in Eq. (5.22) into transverse and longitudinal contributions, $\mathscr{H}_L$ only contains longitudinal and trace parts, and therefore vanishes.

In the physical phase space spanned by the $p$ ($\pi^{ij}_{TT} \equiv \sqrt{g}\, p^{ij}_{TT}$) and $q$ ($h^{TT}_{ij}$) variables one has

$$P[a] \;=\; \int [d\pi^{ij}_{TT}] \, [dh^{TT}_{ij}] \, \exp \left\{ -\hat{I}_2(\pi_{TT}, h^{TT}) \right\} \;\; , \tag{5.26}$$

with $TT$ quadratic action

$$\hat{I}_2(\pi_{TT}, h^{TT}) \;=\; -\frac{1}{16\pi G}\int d^4x \left(\pi^{ij}_{TT}\partial_t h^{TT}_{ij} \;-\; \mathscr{H}_T[\pi_{TT}, h^{TT}]\right) \;. \tag{5.27}$$

After assuming a canonical measure $\prod(dp\,dq/2\pi\hbar)$ over the conjugate variables $\pi^{ij}_{TT}$ and $h^{TT}_{ij}$, and integrating out the momenta $\pi^{ij}_{TT}$, one finally obtains the simple result

$$P[a] \;=\; \int [dh^{TT}_{\mu\nu}]\,\exp\left\{-\frac{1}{32\pi G}\int d^4x\sqrt{g}\,h^{TT\,\mu\nu}\,\nabla_\lambda\nabla^\lambda\,h^{TT}_{\mu\nu}\right\} \;. \tag{5.28}$$

Note that the last expression has been re-written in terms of purely $TT$ components $h^{TT\,\mu\nu}$ of the original four-metric perturbation in Eq. (5.10). Since physical perturbations of the metric are known to be transverse-traceless (associated with a particle of zero mass and spin two) the result is not surprising, and in fact has wider applicability to the semiclassical expansion, beyond the simple choice for background metric implicit in Eq. (5.21).

Formally, the Gaussian integration over the $h^{TT}$ variables gives

$$P[a] \;=\; \det{}^{-1/2}\left(\frac{-\nabla^\lambda\nabla_\lambda}{4\pi\,l_P^2\,\mu^2}\right) \;, \tag{5.29}$$

where $\mu$ is a parameter with units of inverse length, and $l_P^2 = 16\pi G$. The specific power of $\mu$ appearing in this last expression actually depends on the details of the measure $[dh^{TT}_{\mu\nu}]$, which will be discussed further below.

The determinant in Eq. (5.29) is fomally defined through an infinite product of eigenvalues $\lambda_n$ of the Laplacian $\nabla^\lambda\nabla_\lambda$ satisfying Dirichlet boundary conditions,

$$-\nabla^\lambda\nabla_\lambda\;\phi^{(n)}_{\mu\nu} \;=\; \lambda_n\,\phi^{(n)}_{\mu\nu} \tag{5.30}$$

$$\det\left(-\mu^2\,\nabla^\lambda\nabla_\lambda\right) \;\to\; \prod_n\,(\mu^2\,\lambda_n) \;. \tag{5.31}$$

But the product is expected to be divergent and needs to be regularized. One way of doing it is to define a zeta-function sum over eigenvalues (Hawking, 1976)

$$\zeta(s) \;\equiv\; \sum_n \frac{1}{\lambda_n^s} \;, \tag{5.32}$$

so that formally on obtains

$$-\tfrac{1}{2}\log\det\left(-\nabla^\lambda\nabla_\lambda\right) = -\tfrac{1}{2}\,\mathrm{tr}\log\left(-\nabla^\lambda\nabla_\lambda\right) = -\tfrac{1}{2}\sum_n \log\lambda_n = \tfrac{1}{2}\,\zeta'(0) \;, \tag{5.33}$$

with $\zeta'(0) \equiv d\zeta/ds|_{s=0}$. For the wave function $P[a]$ itself one then has

$$\log P[a] \;=\; \tfrac{1}{2}\,\zeta'(0) + \zeta(0)\,\log(4\pi\, l_P^2\, \mu^2) \;\; . \tag{5.34}$$

The dependence of $P[a]$ on $a$ can be determined by a scaling argument. When rescaling the original radius $a$ appearing in the background metric $g_{\mu\nu}$ by $a \to ka$ the Laplacian $\nabla^\lambda \nabla_\lambda$ scales like $1/k^2$, and therefore $\zeta(s)$ by $k^{2s}$. The first term in Eq. (5.34) therefore gives a contribution to $P[a]$ of the form $a^{\zeta(0)}$.

The second contribution in Eq. (5.34) depends on $\mu$, and therefore on the specific details of the functional measure $[dh_{\mu\nu}^{TT}]$. The measure factor comes in through a local weight $h^{p/2}$,

$$\int [dh_{\mu\nu}^{TT}] \;=\; \int \prod_x \, [\det h(x)]^{p/2} \, dh_{\mu\nu}^{TT}(x) \;\; , \tag{5.35}$$

[see Eqs. (2.18) and (2.22)]. If the above measure is chosen to be scale invariant (as in Faddeev and Popov, 1973), then the second contribution in Eq. (5.34) is scale independent, i.e. $k$- or $a$-independent. Then only the first contribution matters, and one has simply, to this order in the semiclassical expansion,

$$P[a] \;=\; \mathcal{N}\; a^{\zeta(0)} \;\; , \tag{5.36}$$

where $\mathcal{N}$ is an $a$-independent normalization constant. In general though one would expect $P[a]$, as defined here, to be sensitive to short-distance details of the theory, and to contain some dependence on the details of the regularization procedure of the measure. The general problem in these types of calculations seems to be the difficulty in decoupling the short distance details of the ultraviolet cutoff $\mu$, which is required to make the product $\prod_x$ in Eq. (5.35) well defined, from the other short distance quantity appearing in this problem, namely the spatial scale for the three-metric $a$.

The last point that needs to be addressed therefore is a determination of $\zeta(s)$ from the eigenvalues of the Laplacian $\nabla^2$, acting on transverse traceless tensors vanishing on the boundary of the three-sphere. Several sub-steps are involved in this calculation, which we will summarize here. The first is to establish a relationship between the function $\zeta(s)$ and the short time expansion for the kernel of the heat equation (for an accessible elementary introduction to the methods of zeta function regularization for functional determinants see for example Ramond, 1990). One can write for $\zeta(s)$ the integral representation

$$\zeta(s) \;=\; \frac{1}{\Gamma(s)} \int_0^\infty d\tau \; \tau^{s-1} \, \mathcal{G}(\tau) \;\; , \tag{5.37}$$

with $\mathcal{G}(\tau)$ defined as

$$\mathcal{G}(\tau) \;=\; \int d^4x \; \mathcal{G}^{\mu\nu}{}_{\mu\nu}(x,x;\tau) \;=\; \sum_n \exp(-\lambda_n \tau) \;\; . \tag{5.38}$$

Here $\mathcal{G}_{\mu\nu,\rho\sigma}(x,x';\tau)$ is a transverse-traceless Green's function for the heat equation

$$\left[ \frac{d}{d\tau} - \nabla^{\lambda} \nabla_{\lambda} \right] \mathscr{G}_{\mu\nu,\rho\sigma}(x,x';\tau) \; = \; \delta_{\mu\nu,\rho\sigma}(x,x')\,\delta(\tau) \ , \tag{5.39}$$

whose expansion in a complete orthonormal basis of $TT$ eigenfunction of the Laplacian $\nabla^2$ reads

$$\mathscr{G}_{\mu\nu,\rho\sigma}(x,x';\tau) \; = \; \sum_n \phi^{(n)*}_{\mu\nu}(x)\,\phi^{(n)}_{\rho\sigma}(x')\,\exp(-\lambda_n \tau) \ , \tag{5.40}$$

(similar to the usual quantum-mechanical completeness, here in imaginary time $\tau$). Using completeness it is then easy to see that the above relations indeed hold, and that $\mathscr{G}(\tau)$ vanishes exponentially for large $\tau$, so that the integral in Eq. (5.37) is convergent for large argument.

If $\mathscr{G}(\tau)$ has a short time asymptotic expansion of the form

$$\mathscr{G}(\tau) \; = \; \frac{1}{\tau^2} \sum_{i=0} g_{i/2}\, \tau^{i/2} \ , \tag{5.41}$$

then a determination of $\zeta(0)$ in Eq. (5.37) requires (since $1/\Gamma(s)$ has a simple zero at $s = 0$) the knowledge of the simple pole term at $s = 0$ in the $\tau$ integral, which comes from the $g_2$ (constant) term.

After transforming from the imaginary time variable $\tau$ to the Laplace transform "energy" variable $E$

$$\left[ E^2 - \nabla^{\lambda} \nabla_{\lambda} \right] \mathscr{G}_{\mu\nu,\rho\sigma}(x,x') \; = \; \delta_{\mu\nu,\rho\sigma}(x,x') \ , \tag{5.42}$$

and expanding the solution out in a complete set of transverse-traceless hyperspherical harmonics, one can solve the radial equation (in the original metric coordinate $t$) and extract form it (by inverse Laplace transform) the small $\tau$ behavior for $\mathscr{G}(\tau)$. In the end one finds (Schleich, 1985) $g_2 = -\gamma$ with $\gamma = 278/45 \approx 6.18$ in Eq. (5.41), and thus $\zeta(0) = -\gamma$ in Eq. (5.36), and finally

$$P[a] \; = \; \mathscr{N}\, a^{-\gamma} \ . \tag{5.43}$$

One concludes therefore that at least in the semi-classical approximation the amplitude $\Psi$ diverges at small volume. In general the probability $\mathscr{P}$ is related to the square of the amplitude,

$$d\,\mathscr{P}(a) \; = \; |\Psi(a)|^2\, d\mu(a) \ , \tag{5.44}$$

where now $d\mu(a)$ is a suitable measure, induced from a functional measure on the original space on which $\Psi(h)$ in Eq. (5.2) is defined, and here $\Psi(a) \equiv P[a]$.

We conclude this section with a brief discussion of regularization issues. The problems that arise in attempting to regulate the determinant in Eq. (5.29) can, in our opinion, be illustrated through the following simple example. Consider the prototype integral over a spinless field $h(x)$

$$Q[a] \;=\; \int [dh] \exp\left(-\frac{1}{4a_0^2}\int_V d^4x\, h\,(-\nabla^2)\,h\right) \;\;, \tag{5.45}$$

where $V \simeq a^d$, and we have set $l_P^2 = a_0^2$ with $a_0$ a short distance cutoff (this could have been set to a value much smaller than the Planck length, but here for simplicity we will identify the two scales). Furthermore we will assume in the following that we are in the vicinity of flat space, so that momentum space is still a useful tool. A discretized form, specifying a bit more what is meant in Eq. (5.45), is

$$\int \prod_x dh(x) \exp\left(-\frac{1}{4a_0^2}\sum_x a_0^4\, h(x)\,(-\nabla^2)_x\, h(x)\right) \;\;, \tag{5.46}$$

with $(-\nabla^2)_x$ some discretized lattice version of the original operator. After integrating over the $h$ variables one has

$$\det{}^{-1/2}\left(\frac{-a_0^2\,\nabla^2}{4\pi}\right) \;=\; \exp\left\{-\tfrac{1}{2}\sum_{\mathbf{k}} \log\left(\frac{a_0^2\,k^2}{4\pi}\right)\right\} \;\;. \tag{5.47}$$

To define the quantity in the exponent, one needs to supply an infrared cutoff ($1/a$ here) and an ultraviolet one ($1/a_0$). Using the usual form for the flat space density of states

$$\sum_{\mathbf{k}} \;\rightarrow\; \frac{V}{(2\pi)^d}\int d^dk \;=\; \frac{V\,\Omega_d}{(2\pi)^d}\int k^{d-1}dk \;\;, \tag{5.48}$$

one has

$$\begin{aligned} Q[a] &= \exp\left\{-\tfrac{1}{2}a^d \frac{\Omega_d}{(2\pi)^d}\int_{1/a}^{1/a_0} dk\, k^{d-1}\log\left(\frac{k^2 a_0^2}{4\pi}\right)\right\} \;\;, \\ &\simeq \exp\left\{-c\,(a/a_0)^d \,\log(a/a_0)\right\} \end{aligned} \tag{5.49}$$

where $c$ is some $d$-dependent numerical constant. One should perhaps not commit oneself at this point to a specific form of the functional measure for the field $h(x)$. A slightly more general measure over $h(x)$ could have been used in Eq. (5.45), such as

$$\int_{-\infty}^{+\infty} dh\,|h|^{\alpha}\, e^{-ah^2} \;=\; a^{-(1+\alpha)/2}\,\Gamma\left(\tfrac{1}{2}(1+\alpha)\right) \;\;, \tag{5.50}$$

which would have given instead

$$Q[a] \;\simeq\; e^{-c\,(1+\alpha)\,(a/a_0)^d\,\log(a/a_0)} \;\;. \tag{5.51}$$

Clearly the case of a scale invariant measure, $\alpha = -1$, which gives $Q[a] = \text{const.}$, is interesting but singular: the integral in Eq. (5.50) diverges, and needs to be somehow regulated at small $h$. Furthermore, in general one has a power law behavior for $Q[a]$ only if the $k$-measure in Eq. (5.49) is scale invariant $\sim dk/k$, corresponding here to the case $d = 0$. In this case one has

$$Q[a] \simeq \left( \frac{a}{a_0} \right)^{c\,(1+\alpha)} \quad . \tag{5.52}$$

It is unclear if this simple model is sophisticated enough to capture the essence of the more complete calculation, but it does suggest that the behavior at small $a$ could be strongly influenced by the detailed nature of the short distance cutoff, the form of the effective action as small distances, and the choice of measure over the $h$ variables.

## 5.3 Pair Creation in Constant Electric Fields

It is known that if an atom is subject to a sufficiently strong external uniform electric field it is possible for the field to create pairs. To be specific, if the atomic potential is approximated by a function with a minimum $-V_0$ and going to zero at infinity, then an external electric field, with potential $eEx$, will give an ionization probability for a bound electron

$$\mathscr{P} \simeq \exp - \int_0^{V_0/|eE|} dx \sqrt{2m(V_0 - |eE|x)} = \exp\left( -\frac{4}{3} \sqrt{2mV_0}\, \frac{V_0}{|eE|} \right) \quad , \tag{5.53}$$

using the $WKB$ formula, with $V_0/|eE|$ the distance between the two classical turning points. If one thinks of a negative energy electron as trapped in a potential of depth $V_0 \approx -2m$, then one obtains for the probability of pair production in a strong electric field the semiclassical estimate

$$\mathscr{P} \simeq \exp\left( -\frac{16\,m^2}{3\,|eE|} \right) \quad . \tag{5.54}$$

Clearly in this last instance one is dealing with a relativistic process involving strong fields, so a fully relativistic treatment would seem more appropriate.

To this end one considers the external field-dependent Feynman amplitude $Z[A]$ describing the vacuum to vacuum amplitude in the presence of an external electric field, expressed as a path integral over fermion fields $\psi$ and $\bar\psi$ in a fixed classical background $A$,

$$\begin{aligned} Z[A] &= \int [d\psi][d\bar\psi]\, \exp\left[ \int d^4x\, \bar\psi (i\,\partial\!\!\!/ + e A\!\!\!/ + m - i\varepsilon)\,\psi \right] \\ &= \det [\, i\,\partial\!\!\!/ + e A\!\!\!/ + m - i\varepsilon \,] \quad . \end{aligned} \tag{5.55}$$

Using the formal identity $\det M(A) = \exp \mathrm{tr} \log M(A)$, the corresponding effective action $S(A)$ is then given by minus the exponent of the above expression

$$S(A) \;=\; -\operatorname{tr}\log\det\left(i\,\partial\!\!\!/ + e A\!\!\!/ + m - i\varepsilon\right) + \;\operatorname{tr}\log\det\left(i\,\partial\!\!\!/ + m - i\varepsilon\right) \;, \tag{5.56}$$

where the zero external field, $A = 0$, contribution has been subtracted out in order to avoid spurious divergences. Due to charge conjugation invariance one can re-write the above traces as expressions involving the square of the Dirac operator,

$$S(A) \;=\; \tfrac{1}{2}\;\operatorname{tr}\log\det\left\{ \frac{(i\,\partial\!\!\!/ + e A\!\!\!/)^2 + m^2 - i\varepsilon}{-\partial^2 + m^2 - i\varepsilon} \right\} \;. \tag{5.57}$$

The pair creation probability at the point $x$ will be denoted by $\mathscr{P}(x)$, and is given by

$$|\,Z[A]\,|^2 \;=\; e^{-2S(A)} \;=\; e^{-\int d^4x\,\mathscr{P}(x)} \;, \tag{5.58}$$

so it is this $\mathscr{P}(x)$ that one needs to extract from the trace in Eq. (5.57) (Schwinger, 1951; Brezin and Itzykson, 1970). After making use of the identity

$$\log\frac{x}{y} \;=\; \int_0^\infty \frac{ds}{s}\left( e^{is(y+i\varepsilon)} \;-\; e^{is(x+i\varepsilon)} \right) \;, \tag{5.59}$$

and the explicit form for the square of the Dirac operator (which is the Klein-Gordon operator, plus the spin part $\sigma_{\mu\nu}\,F^{\mu\nu}$), one obtains for the probability of pair creation at $x$

$$\begin{aligned}\mathscr{P}(x) = \operatorname{Re}\operatorname{tr}\int_0^\infty \frac{ds}{s}\, e^{-is(m^2 - i\varepsilon)} \\ \times\, \langle x|\left(\exp\left(is\left[(\,i\nabla_\mu + eA_\mu)^2 + \tfrac{1}{2}\,e\sigma_{\mu\nu}\,F^{\mu\nu}\,\right]\right) - \exp\left(is\,(\,i\nabla_\mu\,)^2\right)\right)|x\rangle > \end{aligned} \tag{5.60}$$

where now the trace is over spinor indices only.

In general, for arbitrary $x$-dependent fields $A(x)$, this is still a rather formidable expression. To simplify things a bit, one can consider a static uniform electric field along the $z$ axis, for which $A^3 = -E\,t$ with $E$ constant. In this gauge the $\langle x|(\ldots)|x\rangle >$ matrix element can be evaluated by inserting a complete set of momentum eigenfunctions, which reduces the problem to computing the trace of the time evolution operator for a harmonic oscillator with imaginary frequency $\omega_0 = 2ieE$. Since the energy levels for such a system are well known, one can easily compute the trace, which then reduces the problem to the evaluation of a single $s$ integral. This last integral can then be evaluated by residues, and one finds for the probability of creating a pair the exact result, for constant uniform field $E$,

$$\mathscr{P} \;=\; \frac{2\,e^2\,E^2}{(2\pi)^3}\,\sum_{n=1}^{\infty}\frac{1}{n^2}\,\exp\left(-\,\frac{n\,\pi\,m^2}{|eE|}\right) \;. \tag{5.61}$$

Since the scale for the external field is set by the electron mass squared ($m^2$), the effect is generally very small in atoms. It is important to note that the result is essentially non-perturbative, and non-analytic in the external field perturbation $E$. Furthermore, a comparison of the exact answer in Eq. (5.61) with the $WKB$ result of Eq. (5.54) shows that the latter only gives the leading term in an infinite series of progressively smaller contribution.

The result obtained for electrons and positrons in strong uniform electric fields are clearly not transferable as is to the gravitational case. For once, there is no notion of oppositely charged particles in gravity. Thus the naive replacement $|eE| \to m/4MG$ for a particle-antiparticle pair, say close to the horizon of a black hole, does not seem to make much sense. Yet the strong electric field $QED$ calculation shows that quantum mechanical tunneling events can take place, and that their effects can be computed to first order using the $WKB$ approximation. A second lesson from $QED$ is that in general higher order corrections should be expected.

## 5.4 Black Hole Particle Emission

Originally quantum gravitational effects for black holes were ignored, since the radius of curvature outside the black hole is much larger than the Planck length, the length scale on which one would expect quantum fluctuations of the metric to become important. If the gravitational field is able to create locally virtual pairs, the local energy density associated with such a pair would be much smaller than the energy scale associated with the local curvature. It can be shown though that in the vicinity of a black hole horizon particle production is possible, due to vacuum fluctuations and tunneling. The resulting effects add up over time, are therefore macroscopic and could in principle be observable.

Normally when describing a stationary non-rotating black hole one uses the Schwarzschild metric in standard form

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2MG}{r}\right)^{-1} dr^2 + r^2 d\Omega_2^2 \ , \tag{5.62}$$

with $d\Omega_2^2 \equiv d\theta^2 + \sin^2\theta d\phi^2$. The metric shows a singularity at $r = 2MG$. Since none of the curvature invariants are singular on the horizon $r = 2MG$, one would expect the singularity to be perhaps an artifact of the coordinate system, here most suitably describing the viewpoint of an observer stationary at infinity. On the other hand a freely falling observer is expected to pass initially unscathed through the black hole horizon, and indeed a singularity-free coordinate system can be found describing such an observer (Kruskal, 1960). In these new coordinates, defined by the transformation from the original Schwarzschild coordinates $(r, \theta, \phi, t) \to (r', \theta, \phi, t')$

$$\begin{aligned} r'^2 - t'^2 &= T^2 \left(\frac{r}{2GM} - 1\right) \exp\left(\frac{r}{2GM}\right) \\ \frac{2r't'}{r'^2 + t'^2} &= \tanh\left(\frac{t}{2GM}\right) \ , \end{aligned} \tag{5.63}$$

the line element takes on the form

$$ds^2 = -\left(\frac{32G^3M^3}{rT^2}\right)\exp\left(\frac{-r}{2GM}\right)(dt'^2 - dr'^2) + r^2 d\Omega_2^2 \ , \tag{5.64}$$

with $T$ an arbitrary parameter, usually taken to be equal to one (often the Kruskal metric is described in terms of equivalent coordinates $x = r' - t'$ and $y = r' + t'$). The original singularity at $r = 2GM$ has disappeared entirely, at the price of a rather unusual topology, since now there are *two* solutions $r', t'$ for every $r, t$, joined together at $r' = 0$. Of course the curvature singularity at the origin $r = 0$ is still present in both coordinate systems.

Hawking has suggested a detailed mechanism by which classical black holes can radiate particles (Hawking, 1975). The basic idea invokes the process of vacuum fluctuations of quantum matter fields, by which virtual particle-antiparticle pairs can be created out of the vacuum as long as their fleeting lifetime is consistent with the uncertainty principle, $\Delta t \sim \hbar/2m$. When a virtual particle pair is created just inside the horizon, the positive energy particle can escape outside the horizon by a process similar to quantum mechanical tunneling, whereas the negative energy antiparticles continues to stay inside the horizon. Conversely, as one would expect from particle-antiparticle symmetry, if a pair is created just outside the horizon, the negative energy antiparticle will tunnel inward, while the positive energy one will eventually escape to infinity. Since particles can be described as positive energy solutions of the wave equation, and antiparticles as negative energy ones, one expects the black hole to accrete a small negative energy through this process, in other words decrease its mass. As a result a classical black hole is far from static, constantly cerating and emitting a steady stream of particles, whose existence, were it not for the very small integrated flux (gravitation is weak, and the relevant quantum amplitudes are very small), could be detected someday.

While the above intuitive picture is easy to convey, the actual original calculation, which will be only outlined below, is technically challenging. The problem is first of all finding a suitable well behaved coordinate system where the singularity at the horizon is not in the way. Secondly, to define a quantum amplitude for particle production one needs to define suitable "in" and "out" states, specifying the Fock space of the system at initial and final times, in a background that is not flat, and for which the notion of frequencies (and therefore energies) and of a vacuum state is frame dependent. Finally one needs to work out in detail the (Bogoliubov) transformation relating these two equivalent Fock spaces, in the case of the Schwarzschild solution or modification thereof. The resulting amplitudes can then be shown to lead to a *thermal* emission of particles, essentially identical to a black-body radiation spectrum, with temperature

$$kT = \frac{\hbar c^3}{8\pi MG} \ . \tag{5.65}$$

Recently a new derivation of the particle emission rate has been obtained, based on a relatively concise calculation involving quantum mechanical tunneling and the $WKB$ approximation (Parikh and Wilczek, 2000). It hinges on the key idea of energy conservation: the mass of the black hole $M$ needs to be decreased suitably when the virtual particle is emitted, thus leading to a non-zero real tunneling amplitude, which can then be shown to agree with the original Hawking calculation.

The computation is most easily carried out in Painlevé coordinates for a static, non-rotating black hole. The Painlevé line element (Painlevé, 1921; Gullstrand, 1922) reads

$$ds^2 = -\left(1 - \frac{2MG}{r}\right) dt^2 + 2\sqrt{\frac{2MG}{r}}\, dt\, dr + dr^2 + r^2\, d\Omega_2^2 \ . \tag{5.66}$$

The corresponding metric describes the same physics, but has several attractive features when compared to the Schwarzschild metric: none of the metric (or inverse metric) components diverge on the horizon $r = 2MG$; furthermore it still covers the inside and outside of the black hole, and constant time slices simply correspond to flat Euclidean space. One can show that the Painlevé time is related to the original Schwarzschild time $t_s$ by

$$t = t_s + 2\sqrt{2MGr} + 2MG \ln \frac{\sqrt{r} - \sqrt{2MG}}{\sqrt{r} + \sqrt{2MG}} \ . \tag{5.67}$$

From $d\tau/dt = 1$ it follows that in these coordinates the time $t$ is linearly related to proper time, $\tau = t + c$, for a radially infalling observer.

In this metric the radial null geodesics have a rather simple form,

$$\frac{dr}{dt} = \pm 1 - \sqrt{\frac{2MG}{r}} \ , \tag{5.68}$$

where the choice of signs depends on whether the rays go towards infinity $(+)$, or away from it $(-)$. One can view the above geodesic equation as arising from a classical mechanics effective potential $V_{eff}(r) = \sqrt{2MG/r} - GM/r$, shown in Fig. (5.1), with a total energy fixed at $1/2$. Note that the maximum of this function is precisely at $r = 2GM$, and that the peak is at the total energy value $1/2$, which seems to make the two classical turning points coincide with the peak.

The fact that the coordinate system is stationary and non-singular allows one to define what is meant by a vacuum: a state whose quantum fields will annihilate modes which carry negative frequency with respect to the Painlevé time $t$. But it is important to note that modifications arise when the particle's self-gravitation is taken into account, and which are crucial in obtaining the correct result. For a non-rotating self-gravitating shell of energy $E$ (visualized as an $s$-wave state) one can show (Kraus and Wilczek, 1995) that the shell moves on a geodesic still described by the line element in Eq. (5.66), but with mass $M \to M + E$, whereas if the total mass is fixed and the black hole mass is allowed to vary, then the shell moves on a geodesic with mass $M \to M - E$.

In order to compute a tunneling amplitude one would like to use the semi-classical or $WKB$ approximation, which assumes point particles. One might worry that a point particle description might not be adequate since the wavelengths involved could be comparable to the black hole size, but this is not so due to the blueshift of frequencies in the vicinity of the horizon. In the following it will therefore be assumed that such particles can be described by point-like objects.

In the *WKB* approximation, the imaginary part of the amplitude for an *s*-wave outgoing positive energy particle which crosses the horizon outward from $r_{in}$ to $r_{out}$ is given by

$$\mathrm{Im}\,S \;=\; \mathrm{Im}\int_{r_{in}}^{r_{out}} p_r\,dr \;=\; \mathrm{Im}\int_{r_{in}}^{r_{out}}\int_0^{p_r} dp_r'\,dr \ , \tag{5.69}$$

where the actual emission rate is the square of the amplitude, $\Gamma \sim \exp(-2\,\mathrm{Im}\,S)$. Using Hamilton's equation for the classical trajectory $\dot{q} = \partial H/\partial p$, here in the form

$$dp_r \;=\; \left(\frac{dr}{dt}\right)^{-1} dH \ , \tag{5.70}$$

with $H = M - E$ and thus $dH = -dE$, and inserting the radial geodesic $dr/dt$ given by Eq. (5.68), one obtains

$$\mathrm{Im}\,S \;=\; -\,\mathrm{Im}\int_{r_{in}}^{r_{out}}\int_0^{E} \frac{dr\,dE'}{1 - \sqrt{\frac{2G(M-E')}{r}}} \ . \tag{5.71}$$

The $r$-integral can now be done by residues, first by transforming to the variable $z = \sqrt{r}$, and then by adding a Feynman $i\varepsilon$ to the energy, which slightly displaces the pole to the upper half-plane,

$$\mathrm{Im}\,S \;=\; -\,\mathrm{Im}\int_0^{E} dE' \int_{z_{in}}^{z_{out}} \frac{2z^2\,dz}{z - \sqrt{2G(M - E' + i\varepsilon)}} \ . \tag{5.72}$$

After closing the contour in the upper half plane and keeping only the imaginary part of the amplitude (the real part contributes an irrelevant phase) one has

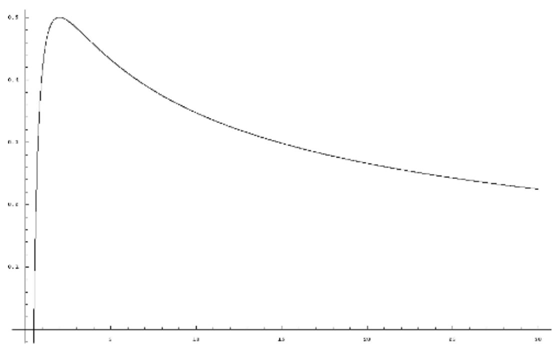


**Fig. 5.1** The effective potential $V_{eff}(r)$, obtained from the geodesic equation in Painlevé coordinates (here shown for $G = M = 1$). The maximum occurs on the horizon $r = 2MG$.

$$\mathrm{Im}\, S \;=\; \mathrm{Im}\, 2\pi i \cdot \tfrac{1}{2} \int_0^E dE' \, 4G(M-E') \;=\; 4\pi GME \left( 1 - \frac{E}{2M} \right) \;. \tag{5.73}$$

Note the sign change due to the fact that $z_i > z_f$. This last result, essential in getting the right sign for the tunneling amplitude, can be visualized by noting that the outgoing particle starts at $r = 2MG - \varepsilon$ (which is just barely inside the initial location of the horizon) and then traverses the contraction horizon (through the classically forbidden region) to materialize at $r = 2G(M-E') + \varepsilon$, which is just outside of the final location for the horizon!

The quantum-mechanical tunneling rate is then given, in the $WKB$ approximation, by

$$\Gamma \;\sim\; \exp(-2\,\mathrm{Im}\, S) \;=\; \exp\left\{ -8\pi GME \left( 1 - \frac{E}{2M} \right) \right\} \;. \tag{5.74}$$

When the quadratic correction in $E$ is neglected (small energy change in the black hole mass), one obtains the Boltzmann weight for a particle with energy $E$ and inverse Hawking temperature $T^{-1} = 8\pi GM$.

It is tempting to pursue for a while the thermodynamic analogy. Since one can associate with the black hole a temperature $T = 1/8\pi MG$, one can also define an entropy for it, using the thermodynamic relation $dE = T dS$, here with $dE = dM$. This gives

$$S \;=\; 8\pi G \int_{M_0}^{M} M' \, dM' \;=\; 4\pi GM^2 \;, \tag{5.75}$$

assuming the integration constant is zero ("a zero mass black hole has zero entropy"). Note then that the expression in the exponent of Eq. (5.74) is precisely the change in the Hawking-Beckenstein (Beckenstein, 1973; 1974; Hawking, 1976; t'Hooft, 1985) black hole entropy,

$$\Delta S \;=\; 4\pi GM^2 - 4\pi G(M-E)^2 \;=\; 8\pi G(ME - \tfrac{1}{2}E^2) \;. \tag{5.76}$$

One largely unresolved puzzle in the context of the semiclassical picture is: what microstates are being counted when one assigns to the black hole an entropy $S = k \log N$.

## 5.5 Method of In and Out Vacua

The original derivation by Hawking of black hole radiance relies on a slightly different set of arguments (Hawking, 1975). Here we will only outline the main steps of the argument. One start by considering a massless real scalar field with wave equation

$$g_{\mu\nu} \nabla^\mu \nabla^\nu \, \phi(x) \;=\; 0 \;, \tag{5.77}$$

where the covariant derivative $\nabla_\mu$ is defined for an asymptotically flat spacetime, describing initially the gravitationally collapsed object that gave origin to the black hole. The quantum operator $\phi$ can be expanded

$$\phi(x) = \sum_k \left[ f_k(x)\, a_k + \bar{f}_k(x)\, a_k^\dagger \right] \;, \tag{5.78}$$

in terms of a complete set $\{f_k\}$ of (generally complex) c-number solutions of the original wave equation,

$$g_{\mu\nu} \nabla^\mu \nabla^\nu \, f_k(x) = 0 \;. \tag{5.79}$$

The $f_k$'s contain asymptotically only *ingoing*, and positive frequency components: they only contain positive frequency particles on past null infinity, here denoted by $I^-$ (we follow here the original notation, where a state of positive frequency is assumed to have a time dependence $e^{i\omega t}$). The position independent operators $a_k^\dagger$ and $a_k$ are therefore interpreted as creation and destruction operators for these incoming particles. Such operators define in the usual way a vacuum, here denoted by $|0_-\rangle$, which is devoid of $a$ quanta

$$a_k |0_-\rangle = 0 \;, \tag{5.80}$$

and from it a corresponding Fock space.

The field operator $\phi$ can also be equivalently expanded in a different, but still complete, set of c-number solutions of the original wave equation. These will now be denoted by $p_k$ and $q_k$, with their complex conjugate counterparts $\bar{p}_k$ and $\bar{q}_k$. The $p_k$'s are chosen to be asymptotically *outgoing*, positive frequency, solutions of the wave equation, subject to the condition that they be zero on the horizon: they will only contain a positive frequency part on the future null horizon $I^+$. The $p_k$'s do not form a complete set, and that is where the $q_k$'s come in: they represent solutions which contain no outgoing component, and are zero on the future null horizon $I^+$. No restriction is needed on the frequency part of the $q_k$'s.

In this second basis the quantum operator $\phi$ has the expansion

$$\phi(x) = \sum_k \left[ p_k(x)\, b_k + \bar{p}_k(x)\, b_k^\dagger + q_k(x)\, c_k + \bar{q}_k(x)\, c_k^\dagger \right] \;, \tag{5.81}$$

with $p_k$ and $q_k$ c-number solutions of the original wave equation,

$$g_{\mu\nu} \nabla^\mu \nabla^\nu \, p_k(x) = g_{\mu\nu} \nabla^\mu \nabla^\nu \, q_k(x) = 0 \;, \tag{5.82}$$

and the $b_k^\dagger$, $b_k$, $c_k^\dagger$, $c_k$ the corresponding creation and destruction operators for particles in the corresponding mode.

Since the two sets of c-number solutions both individually form a complete set, and are equivalent, they should be related to each other by a linear transformation,

$$p_k = \sum_{k'} \left[ \alpha_{kk'}\, f_{k'} + \beta_{kk'}\, \bar{f}_{k'} \right] \;, \tag{5.83}$$

with a similar expressions for $q_k$. It is easy to see that the mixing between the $f_k$'s and the $p_k$'s will in general involve complex coefficients due to the mixing of positive and negative frequencies taking place during the collapse, as a consequence of the time dependence of the metric (which can be visualized as a time dependent external quantum mechanical perturbation).

After substituting Eq. (5.83) into Eq. (5.81) and equating it to the expansion in Eq. (5.78), one obtains a direct expression for the second set of creation and destruction operators $b_k^\dagger$, $b_k$, $c_k^\dagger$, $c_k$ in terms of the original set $a_k^\dagger$, $a_k$, with

$$b_k = \sum_{k'} \left[ \bar{\alpha}_{kk'}\, a_{k'} - \bar{\beta}_{kk'}\, a_{k'}^\dagger \right] \ , \tag{5.84}$$

and similarly for the $c_k$ operators.

The last equality allows one to compute the expectation value of the number operator $N_k^{(b)} = b_k^\dagger b_k$, in a state which does not contain any *incoming* particles (and therefore, as stated at the beginning, is $|0_-\rangle$ with $a_k|0_-\rangle = 0$), giving therefore for the mode $k$

$$\langle 0_- |\, b_k^\dagger b_k \,| 0_- \rangle = \sum_{k'} |\beta_{kk'}|^2 \ . \tag{5.85}$$

Therefore the problem of computing the number of particles created and emitted to infinity has been reduced to computing the complex expansion coefficients $\beta_{kk'}$ in Eq. (5.83).

To derive an expression for the coefficients $\beta_{kk'}$ it is sufficient to consider a metric with spherical symmetry, where the solutions to the wave equation $\nabla^2 f_k = 0$ can be written as a product of spherical harmonics $Y_{lm}( \ theta, \phi)$ times a time dependent radial wave function. The latter can be written in terms of advanced and retarded solutions with frequency dependence

$$\frac{1}{\sqrt{2\pi\omega}}\, e^{i\omega u} \ , \tag{5.86}$$

with $u$ and $v$ retarded and advanced coordinates

$$\begin{aligned} u &= t - r - 2MG \log \left| \frac{r}{2MG} - 1 \right| \\ v &= t + r + 2MG \log \left| \frac{r}{2MG} - 1 \right| \ , \end{aligned} \tag{5.87}$$

and invariant distance

$$ds^2 = -\frac{2MG}{r}\, e^{-r/2MG}\, e^{(v-u)/4MG}\, du\, dv + r^2\, d\Omega_2^2 \ , \tag{5.88}$$

with $r = r(u,v)$. Then the sum over modes $\sum_k$ gets replaced by a sum over frequencies $\sum_{\omega l m}$, so that for each partial wave $(l,m)$

$$f_{\omega' lm} = \frac{1}{\sqrt{2\pi\,\omega'}}\, f_{\omega'}(r)\, Y_{lm}(\theta,\phi)\, \frac{e^{i\omega' v}}{r} \tag{5.89}$$

$$p_{\omega lm} = \frac{1}{\sqrt{2\pi\,\omega}}\, p_{\omega}(r)\, Y_{lm}(\theta,\phi)\, \frac{e^{i\omega u}}{r} \ , \tag{5.90}$$

one has an expansion of the outgoing solution $p_{lm}$ in terms of the incoming $f_{lm}$ and $\bar{f}_{lm}$ solutions,

$$p_\omega = \int d\omega' \left[ \alpha_{\omega\,\omega'} f_{\omega'} + \beta_{\omega\,\omega'} \bar{f}_{\omega'} \right] \ . \tag{5.91}$$

The conventions used are such that on past null infinity $I^-$ one has for the $f$ solution

$$f_\omega \to e^{i\omega u} \ . \tag{5.92}$$

In order to compute the amplitudes $\alpha_{\omega\,\omega'}$ and $\beta_{\omega\,\omega'}$ one considers the following process: a wave which has a positive frequency $\omega$ on the future null infinity $I^+$,

$$p_\omega \to e^{i\omega u} \ , \tag{5.93}$$

and propagates backward through spacetime. Part of this wave will be scattered elastically by the curvature of the black hole and will end up at past null infinity $I^-$, with the same frequency it started out with. Its contribution to $\alpha_{\omega\,\omega'}$ is therefore be proportional to $\delta(\omega - \omega')$. A second part of the wave will propagate backwards into the black hole, through the origin at $r = 0$ and out to past null infinity $I^-$. These waves will experience a large blueshift due to the presence of the horizon, and will reach $I^-$ with an asymptotic form

$$p_\omega \sim C \frac{1}{\sqrt{2\pi\omega}} e^{-i\,\omega\,\kappa\,\log(v_0 - v)} \ , \tag{5.94}$$

for $v < v_0$, and zero for $v > v_0$. Here $\kappa = 4MG/c^4$ (the inverse of the surface gravity of the black hole), and $v_0$ is the last advanced time at which the particle can leave $I^-$ and reach $I^+$ by passing through the origin.

Then the complex amplitude $\alpha_{\omega\,\omega'}$ is computed from Eq. (5.74) by Fourier transform, and in the limit of large $\omega'$,

$$\alpha_{\omega\,\omega'} = (f_{\omega'}, p_\omega)_{I^-} \ . \tag{5.95}$$

With the help of the $v$ integral

$$\frac{1}{\sqrt{2\pi\omega}} \frac{1}{\sqrt{2\pi\omega'}} \int_{-\infty}^{0} dv \left( \kappa \frac{\omega}{v} - \omega' \right) e^{-i\,\omega'\,v}\, e^{-i\,\omega\,\kappa\,\log(-v)}$$
$$= -\frac{i}{\sqrt{\omega\,\omega'}} (-i\omega')^{i\omega\kappa}\, \Gamma(1 - i\,\omega\kappa) \ , \tag{5.96}$$

one can evaluate the Fourier transforms, and for large values of $\omega'$ one finds

$$\begin{aligned} \alpha_{\omega\,\omega'} &\approx C \exp\left[i(\omega - \omega')v_0\right] (\omega'/\omega)^{1/2}\, \Gamma(1 - i\,\omega\kappa) \left[-i(\omega - \omega')\right]^{-1+i\omega\kappa} \\ \beta_{\omega\,\omega'} &\approx C \exp\left[i(\omega + \omega')v_0\right] (\omega'/\omega)^{1/2}\, \Gamma(1 - i\,\omega\kappa) \left[-i(\omega + \omega')\right]^{-1+i\omega\kappa} \ , \end{aligned} \tag{5.97}$$

and $\beta_{\omega\,\omega'} \approx -i\,\alpha_{\omega,-\omega'}$.

One can then compute the total number of outgoing particles created in the frequency range $d\omega$ around $\omega$. It is given by

$$d\omega \cdot \int d\omega' \, |\beta_{\omega\,\omega'}|^2 \ , \tag{5.98}$$

which is infinite since $\beta_{\omega\,\omega'}$ goes like $1/\sqrt{\omega'}$ for large $\omega'$. The infinity in the total number of created particles indicates a steady rate of particle emission from the black hole, continuing for an infinite amount of time.

A precise relationship between the magnitudes of $\alpha_{\omega\,\omega'}$ and $\beta_{\omega\,\omega'}$ can be computed by the following argument based on the analytic properties of the amplitude. The $p_\omega$ solutions just constructed is zero on past null infinity $I^-$ for large values of $v$. Its Fourier transform will therefore be analytic in the upper half $\omega'$ plane, but the complex amplitude $\alpha_{\omega\,\omega'}$ contains a factor $(-i\omega')^{i\omega\kappa}$ with a logarithmic branch point at $\omega' = 0$. To obtain $\beta_{\omega\,\omega'}$ from $\alpha_{\omega\,\omega'}$ one needs to analytically continue $\alpha_{\omega\,\omega'}$ around the singularity, which implies for large $\omega'$ the relationship

$$|\,\alpha_{\omega\,\omega'}\,| \;=\; e^{\pi\omega\kappa} |\,\beta_{\omega\,\omega'}\,| \ . \tag{5.99}$$

This last equality then implies, for a given mode of frequency $\omega$ and for a fixed $\omega' \equiv \omega_0$, the following frequency distribution of outgoing particles

$$\frac{|\,\beta_{\omega\,\omega_0}\,|^2}{|\,\alpha_{\omega\,\omega_0}\,|^2 - |\,\beta_{\omega\,\omega_0}\,|^2} \;=\; \frac{1}{e^{\pi\omega\kappa} - 1} \ , \tag{5.100}$$

where the denominator $|\,\alpha_{\omega\,\omega_0}\,|^2 - |\,\beta_{\omega\,\omega_0}\,|^2$ describes the fraction of particles that enters the collapsing body. It leads therefore to an emission probability

$$\mathcal{N}(\omega) \;=\; \frac{1}{e^{\pi\omega\kappa} - 1} \ . \tag{5.101}$$

But this is precisely the relationship one would expect between absorption and emission cross sections for a body with temperature $T = 1/2\pi\kappa = 1/8\pi MG$.

Other types of fields can be considered, such as the electromagnetic field or the linearized gravitational field, and one finds also in these cases a thermal radiation spectrum. If the particles are fermions, such as neutrinos, then the spectrum can be shown to be of the Fermi-Dirac type $(e^{\pi\omega\kappa} + 1)^{-1}$. If the fermions have mass $m$, then $\omega$ will contain a rest mass contribution. The thermal emission will then be very small unless $T = 1/8\pi GM$ is greater than $m$. Finally there is the very interesting and difficult general issue of the extent to which these results are universal, that is independent of the detailed nature, history and composition of the black hole (Fredenhagen and Haag, 1990).

## 5.6 Complex Periodic Time

There is another, and perhaps more direct, way by which a temperature can be associated with a Schwarzschild black hole: through a periodicity in imaginary time of the original static isotropic Lorentzian solution of Eq. (5.62). One performs the usual Wick rotation $t = -i\tau$ on this metric, and obtains

$$ds^2 = \left(1 - \frac{2GM}{r}\right) d\tau^2 + \left(1 - \frac{2MG}{r}\right)^{-1} dr^2 + r^2 d\Omega_2^2 \ , \tag{5.102}$$

which has positive signature for $r > 2MG$. One can now define a new radial coordinate $\rho$ through

$$\rho = 4MG \left(1 - \frac{2MG}{r}\right)^{1/2} \ . \tag{5.103}$$

In these new coordinates the Euclidean Schwarzschild metric becomes

$$ds^2 = \rho^2 \left(\frac{d\tau}{4MG}\right) + \left(\frac{r^2}{4M^2 G^2}\right) d\rho^2 + r^2 d\Omega_2^2 \ . \tag{5.104}$$

Near $r = 2MG$ one can compare the first two terms with a flat Euclidean two-dimensional metric in polar coordinates ($ds^2 = dr^2 + r^2 d\theta^2$). In order to keep the metric regular at $\rho = 0$, the variable $\tau/4MG$ has to be identified with period $2\pi$. This in turn implies that the Euclidean Schwarzschild solution has to be periodic in imaginary time, with period $8\pi MG$. As will be shown below, this implies the result $\beta = 8\pi MG$ of Eq. (5.65).

The Schwarzschild solution is not the only case where one finds a periodicity in imaginary time, and therefore an associated natural temperature. In the case of a uniformly accelerated observer in the $x$ direction the natural metric is the Rindler one

$$ds^2 = -a^2 \rho^2 d\tau^2 + d\rho^2 \ , \tag{5.105}$$

which is related to the flat Minkowski metric by $x = \rho \cosh(a\tau)$ and $t = \rho \sinh(a\tau)$ (the $y$ and $z$ coordinates are left unchanged). Therefore Rindler coordinates correspond to a flat metric written in polar coordinates, and describe a detector moving with uniform acceleration $a$ along a path of constant $\rho$. Since Rindler coordinates are periodic in complex time with period $2\pi/a$, they can be associated with an inverse temperature $\beta = 2\pi/a$, which implies

$$kT = \frac{\hbar a}{2\pi c} \ , \tag{5.106}$$

for the temperature seen by the uniformly accelerating particle.

This leads to one of the simplest examples for the Unruh effect (Unruh, 1976), whereby an accelerated observer observes a black-body radiation spectrum when an inertial observer under otherwise identical conditions sees none, implying that the notion of a vacuum depends on the state of motion of the observer. A space that looks empty for one observer (the inertial one), is filled with a thermal bath of particles for a different observer (the uniformly accelerating one). It is easy to see that in fact Rindler coordinates can be used to approximate the Schawrzschild solution close to the horizon, with a surface gravity for the black hole given by $a = 1/4MG$.

From rather general arguments one can relate the real time evolution of a quantum state to the imaginary time evolution of quantum statistical mechanics, at a fixed finite temperature $\beta = 1/T$, via the correspondence

$$U(t) \equiv e^{-\frac{i}{\hbar}Ht} \underset{t \to -i\hbar\tau}{\sim} e^{-H\tau} \underset{\tau \to \beta}{\sim} e^{-\beta H} , \tag{5.107}$$

with a periodic complex time $\tau$, with period $\beta$. This last statement is proven as follows. One notices that this rather well known chain of connections between quantum and statistical mechanics is particularly evident in the Feynman path integral approach. There one writes for the quantum-mechanical amplitude of Eq. (5.4)

$$\langle \mathbf{x}_f \,|\, \exp\left(-\frac{i}{\hbar} H\,(t_f - t_i)\right) |\, \mathbf{x}_i \rangle = \int_{\mathbf{x}_i(t_i)}^{\mathbf{x}_f(t_f)} [d\mathbf{x}(t)] \, \exp\left\{\frac{i}{\hbar}\, I[\mathbf{x}(t)]\right\} . \tag{5.108}$$

In statistical mechanics on the other hand one considers the canonical partition function $Z$ at inverse temperature $\beta$,

$$Z(\beta) = \sum_{\text{states } n} \langle \phi_n \,|\, \exp(-\beta H) \,|\, \phi_n \rangle = \int [d\phi] \, \exp\{-\beta \, \hat{I}[\phi]\} , \tag{5.109}$$

where the sum over $n$ is over a complete set of states $\phi_n$. In the last expression the quantity $\hat{I}[\phi]$ is essentially the same as the statistical mechanics Hamiltonian $H$, and in the last integral one is summing over all $\phi$ field configurations. Thus effectively one is considering, in the statistical mechanics case, an Euclidean functional integral ("sum over all states") over all fields or degrees of freedom $\phi$, on a space which is periodic (due to the trace over $n$) in the imaginary time direction, with period $\beta = 1/T$. The connection between the two theories generally holds between quantum mechanics in $(d-1)+1$ dimensions and statistical mechanics in $d$ dimensions.

In the Feynman path integral approach it is possible to introduce a finite temperature from the start (Feynman and Hibbs, 1965; Bernard, 1974; Dolan and Jackiw, 1974). One writes, in analogy to the non-relativistic single particle derivation of the Feynman path integral,

$$\begin{aligned} \operatorname{tr} e^{-\beta H} &= \sum_{\phi} \langle \phi \,|\, e^{-\beta H} \,|\, \phi \rangle \\ &= \mathcal{N} \int [d\pi] \int_{\text{periodic}} [d\phi] \exp\left\{\frac{1}{\hbar}\int_0^\beta d\tau \int d^3x \left[i\pi\dot{\phi} - \mathcal{H}(\pi,\phi)\right]\right\} , \end{aligned} \tag{5.110}$$

with $\dot{\phi} = i\partial\phi/\partial\tau$, $\mathscr{N}$ a normalization factor, and complex time $t = -i\tau$. The functional integration $[d\phi]$ only contain those paths which are periodic in the complex time $\tau$, i.e. $\phi(\tau = \beta) = \phi(\tau)$ (for Fermions they need to be anti-periodic). Integrating over the momenta $\pi$ gives

$$\mathrm{tr}\, e^{-\beta H} = \mathscr{N}' \int_{\text{periodic}} [d\phi] \exp\left\{ \frac{1}{\hbar} \int_0^\beta d\tau \int d^3x \mathscr{L}(\phi, i\dot{\phi}) \right\} \tag{5.111}$$

where $\mathscr{N}'(\beta)$ is a second $\beta$-dependent constant that comes from the Gaussian $\pi$ integration. The latter has to be defined, as in the original non-relativistic path integral, by introducing a lattice spacing and doing the $\pi$ integration carefully. Note that the finite temperature formalism has automatically achieved a Wick rotation to the Euclidean theory, and the Feynman $i\varepsilon$ prescription is no longer needed. Furthermore, because of the periodicity in complex time, all energy integrals are converted into finite frequency sums [see, for example, Abrikosov, Gorkov and Dzyaloshinski, 1963; Fetter and Walecka, 1971].

In the gravitational case, a similar functional integral needs to be evaluated at finite temperatures. Formally it is given by

$$Z(\beta) = \mathrm{tr}\, e^{-\beta H} = \mathscr{N}' \int_{\text{periodic}} [dg_{\mu\nu}] \exp\left\{ \frac{1}{\hbar} \int_0^\beta d\tau \int d^3x \mathscr{L}\left(g_{\mu\nu}, i\dot{g}_{\mu\nu}\right) \right\} . \tag{5.112}$$

Here the path integral is over *all* gravitational fields $g_{\mu\nu}$ which are periodic in imaginary time $\tau$, with period $\beta$; only in the semi classical limit these are restricted to the saddle points of $\mathscr{L}$, corresponding to suitable solutions of the classical field equation of general relativity. But the introduction of a finite temperature $\beta$ does not of course alleviate the short distance problem of ultraviolet divergences, which still remains and needs to somehow be addressed in order to obtain finite quantum corrections to $Z(\beta)$.

The periodicity in imaginary time of the Schwarzschild solution has in fact been used to provide an alternative path integral derivation of black hole radiance (Hartle and Hawking, 1976). There the amplitude for a black hole to emit a scalar particle in a particular mode is expressed as a sum over paths connecting the future singularity and infinity. By analytic continuation in the complexified Schwarzschild space this amplitude is then related to that for a particle to propagate from the past singularity to infinity and hence, by time reversal, to the amplitude for the black hole to absorb a particle in the same mode. The form of the connection between the emission and absorption probabilities then shows that the black hole will emit scalar particles with a thermal spectrum characterized by the temperature of Eq. (5.65).

The thermodynamic analogy, applied to a black hole with a temperature $T = \beta^{-1} = 1/8\pi M$ with (average) energy $E = M$ and entropy $S = 4\pi G M^2$, gives a free energy $F \equiv E - TS = \frac{1}{2}M$ and therefore a partition function $Z$ (for a single state!)

$$Z = \sum_n e^{-\beta E_n} = e^{-\beta F} \to e^{-\frac{\beta^2}{16\pi G}} . \tag{5.113}$$

But the last contribution is all there is to it only if one assumes that the Schwarzschild solution is the only one contributing to the thermodynamic state sum at fixed inverse temperature $\beta$. This is indeed true in the saddle point approximation, but in general many more configurations, which are not necessarily saddle points, will contribute to the finite temperature Euclidean Feynman path integral.

## 5.7 Black Hole Evaporation

If black holes emit particles by quantum mechanical tunneling, then one would expect them to gradually loose mass and eventually disappear. Thus one of the most important effects of Hawking radiation is that black holes, given enough time, can evaporate. One would expect that the back reaction, i.e. the adjustment of the space-time metric to the process of thermal emission, would be rather small provided the Schwarzschild radius is much larger than the Planck length, $MG \gg \sqrt{G}$, or $M \gg 1/\sqrt{G}$, in which case one could continue to use the previous semi-classical results.

It is of interest to provide a quantitative framework and determine an approximate lifetime for the black hole. Since the thermal emission of particles is a quantum effect, one would expect the rate of mass decrease to be proportional to $\hbar$. Thus by purely dimensional arguments one is lead to an equation of the type

$$\frac{dM(t)}{dt} = -\frac{\hbar\,\alpha}{G^2 M^2(t)} \quad , \tag{5.114}$$

with $\alpha$ a dimensionless constant of order one, which could in principle be estimated by more detailed calculations. Integrating this equation with initial condition $M(t=0) = M_0$ gives

$$M(t) = M_0 \left(1 - \frac{3\hbar\,\alpha}{G^2 M_0^3}\, t\right)^{1/3} \quad , \tag{5.115}$$

which shows that the black hole will eventually evaporate, with a characteristic time scale $\tau \simeq G^2 M_0^3 / 3\hbar\,\alpha$. In terms of actual numbers the above result implies that today, i.e. about $1.37 \times 10^{10}$ years after the big bang, one would expect to see the final explosive stages in the evaporation of relatively small black holes of mass ca. $10^{15} g$ (the mass of a small mountain), expected to have been created at the beginning of the universe, the so-called primordial black holes. The Hawking temperature of such mini black holes must have been originally therefore about $10^{11}$ degrees Kelvin. On the other hand, for astronomically large black holes, the evaporation time is expected to be extremely long, about $10^{61}$ times the age of the universe for a 30 solar mass black hole.

Since black holes radiate energy in the form of black body radiation, one expects their total luminosity $\mathscr{L}$ to be described by the Stefan-Boltzmann formula (essentially the frequency integral of the Planck distribution),

$$\mathscr{L} = A \sigma T^4 \,, \tag{5.116}$$

with $A = 4\pi(2MG)^2$ the surface area of the black hole, and $\sigma = \pi^2 k^4/60c^2\hbar^3$, the Stefan-Boltzmann constant. Depending on the value of the Hawking temperature relative to the rest mass of the particle, the black body will then either radiate mostly photons (at low temperatures), or photons as well as other particles (at high temperatures). The effect of these additional particles will then be to increase slightly the above quoted luminosity, by an additional factor which will take into account the number of particle species and their spin multiplicities.

## 5.8 Quantum Gravity Corrections

The effects of black hole radiation are derived within a semi classical picture of gravity, where relativistic matter is treated quantum-mechanically but the gravity background is classical. Thus the back reaction of the particles on the metric, and quantum fluctuations of the metric itself, are not taken into account. How do quantum fluctuations in the metric affect the semi classical picture? One might worry, for example, to what extent a scale-dependent gravitational constant $G(r)$ would affect the classical background metric and therefore lead to a modification of the semi classical tunneling results, in a way that is similar to how the QED Uehling correction affects the nature of the static Coulomb potential and therefore energy levels in atomic physics. The problem is of course largely unresolved to this day, but a few interesting proposals have been put forward.

In a quantum theory of gravity one would expect radiative corrections to involve the Planck length $l_P$, which can be regarded in some ways as a natural ultraviolet cutoff. In particular a (largely Newtonian) argument has been given in (Sorkin, 1996; see also Casher et al., 1997) suggesting that quantum fluctuations in the metric endow the horizon at short distances with a non-zero quantum mechanical width whose size is related to $l_P$ (for an illustration see Fig. 5.2). One can understand such arguments based on the intuitive picture of quantum fluctuations one gathers from the Feynman path integral approach. In such a theory of gravity one is supposed to sum over all field configurations, weighted by $\exp(iS)$.

In a semi-classical expansion for quantum gravity one would expect the dominant contribution to the path integral to come from field configurations close to the stationary points of the action, i.e. solutions to the classical field equations. The semi classical corrections would then arise from a Gaussian integral over small fluctuations in the vicinity of the saddle point. Up to ultraviolet divergences, which would lead to a possibly scale-dependent renormalizations of Newton's constant $G$, of the cosmological constant $\lambda$ and other short distance corrections coming from higher derivative terms, one would expect the relevant scale close to the horizon to be the Planck length $l_P$, or some combination of the Planck scale and the black hole mass $M$. The quoted author suggests that the spectrum of metric fluctuations around the

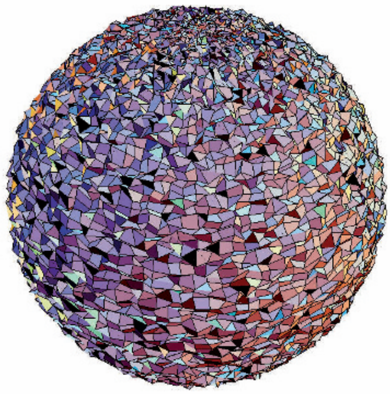


**Fig. 5.2** Pictorial representation of the wrinkled surface of a quantum-mechanical black hole. The geometry of the horizon is rendered rough by strong short distance quantum fluctuations in the metric. Since these are described by a massless particle, one would expect the microscopic geometry to be self-similar, perhaps best described by a gravitational Hausdorff dimension, in analogy with the quantum-mechanical Feynman path of a non-relativistic particle.

horizon then changes to a scale-invariant spectrum for length scales smaller than some critical value.

This would imply that, in a quantum theory of gravity, the classical notion of a sharp black hole horizon at $r = 2MG$ would be superseded at shorter distances by a geometrically more complex object, an entity whose shape is far from smooth on small scales. The horizon would appear to be essentially classical up to some distance scale, below which its self-similar (or fractal) nature would start to emerge (in the Newtonian approximation this scale is of the order of $(MG^2)^{1/3}$). Thus its surface area would no longer be given by the Euclidean result; instead it would depend, just like the overall size of a random walk or a Wiener path, on the scale at which it is measured.

Since these short distance fluctuations are essentially described by a massless particle, one would expect equally the microscopic geometry to be self-similar or fractal, and therefore to be best described by a suitable Hausdorff dimension, in analogy with the quantum-mechanical Feynman path of a non-relativistic particle. Furthermore, one would expect that the difference in surface area on microscopic scales, comparable to the ultraviolet cutoff, and on the larger classical scale would be described by an (ultraviolet divergent) renormalization factor, $A_R = Z_A(\Lambda) \cdot A_0(\Lambda)$, in analogy to charge and wavefunction renormalization in QED.

# Chapter 6
# Lattice Regularized Quantum Gravity

## 6.1 The Lattice Theory

The following sections are based on the lattice discretized description of gravity known as Regge calculus, where the Einstein theory is expressed in terms of a simplicial decomposition of space-time manifolds. Its use in quantum gravity is prompted by the desire to make use of techniques developed in lattice gauge theories (Wilson, 1973),[1] but with a lattice which reflects the structure of space-time rather than just providing a flat passive background (Regge, 1961). It also allows one to use powerful nonperturbative analytical techniques of statistical mechanics as well as numerical methods. A regularized lattice version of the continuum field theory is also usually perceived as a necessary prerequisite for a rigorous study of the latter.

In Regge gravity the infinite number of degrees of freedom in the continuum is restricted by considering Riemannian spaces described by only a finite number of variables, the geodesic distances between neighboring points. Such spaces are taken to be flat almost everywhere and are called *piecewise linear* (Singer and Thorpe, 1967). The elementary building blocks for $d$-dimensional space-time are *simplices* of dimension $d$. A 0-simplex is a point, a 1-simplex is an edge, a 2-simplex is a triangle, a 3-simplex is a tetrahedron. A $d$-simplex is a $d$-dimensional object with $d+1$ vertices and $d(d+1)/2$ edges connecting them. It has the important property that the values of its edge lengths specify the shape, and therefore the relative angles, uniquely.

A simplicial complex can be viewed as a set of simplices glued together in such a way that either two simplices are disjoint or they touch at a common face. The relative position of points on the lattice is thus completely specified by the *incidence matrix* (it tells which point is next to which) and the *edge lengths*, and this in turn induces a metric structure on the piecewise linear space. Finally the polyhedron constituting the union of all the simplices of dimension $d$ is called a geometrical

[1] As an example of a state-of-the-art calculation of hadron properties in the lattice formulation of $SU(3)$ *QCD* see (Aoki et al, 2003).

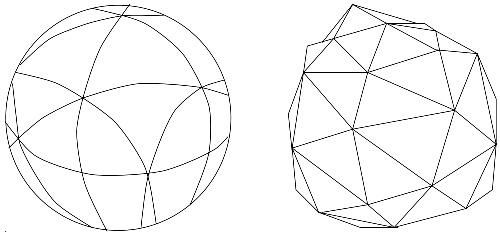


**Fig. 6.1** Polydedral approximation to a sphere.

complex or *skeleton*. The transition from a smooth triangulation of a sphere to the corresponding secant approximation is illustrated in Fig. 6.1.

A manifold can then be defined by its relationship to a piecewise linear space: a topological space is called a closed $d$-dimensional manifold if it is homeomorphic to a connected polyhedron, and furthermore, if its points possess neighborhoods which are homeomorphic to the interior of the d-dimensional sphere.

## 6.2 General Formulation

We will consider here a general simplicial lattice in $d$ dimensions, made out of a collection of flat $d$-simplices glued together at their common faces so as to constitute a triangulation of a smooth continuum manifold, such as the $d$-torus or the surface of a sphere. If we focus on one such $d$-simplex, it will itself contain sub-simplices of smaller dimensions; as an example in four dimensions a given 4-simplex will contain 5 tetrahedra, 10 triangles (also referred to as hinges in four dimensions), 10 edges and 5 vertices. In general, an $n$-simplex will contain $\binom{n+1}{k+1}$ $k$-simplices in its boundary. It will be natural in the following to label simplices by the letter $s$, faces by $f$ and hinges by $h$. A general connected, oriented simplicial manifold consisting of $N_s$ $d-$simplices will also be characterized by an incidence matrix $I_{s,s'}$, whose matrix element $I_{s,s'}$ is chosen to be equal to one if the two simplices labeled by $s$ and $s'$ share a common face, and zero otherwise.

The geometry of the interior of a $d$-simplex is assumed to be flat, and is therefore completely specified by the lengths of its $d(d+1)/2$ edges. Let $x^\mu(i)$ be the $\mu$-th coordinate of the $i$-th site. For each pair of neighboring sites $i$ and $j$ the link length squared is given by the usual expression

$$l_{ij}^2 \;=\; \eta_{\mu\nu}\,[x(i)-x(j)]^\mu\,[x(i)-x(j)]^\nu \quad , \tag{6.1}$$

with $\eta_{\mu\nu}$ the flat metric. It is therefore natural to associate, within a given simplex $s$, an edge vector $l_{ij}^\mu(s)$ with the edge connecting site $i$ to site $j$.

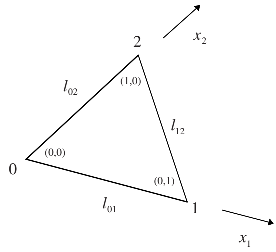



**Fig. 6.2** Coordinates chosen along edges of a simplex, here a triangle.

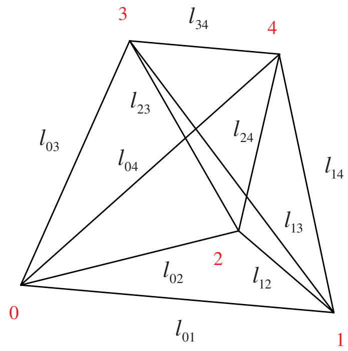



**Fig. 6.3** A four-simplex, the four-dimensional analog of a tetrahedron. It contains five vertices, ten edges, ten triangles and five tetrahedra.

## 6.3 Volumes and Angles

When focusing on one such $n$-simplex it will be convenient to label the vertices by $0, 1, 2, 3, \ldots, n$ and denote the square edge lengths by $l_{01}^2 = l_{10}^2$, ... , $l_{0n}^2$. The simplex can then be spanned by the set of $n$ vectors $e_1$, ... $e_n$ connecting the vertex 0 to the other vertices. To the remaining edges within the simplex one then assigns vectors $e_{ij} = e_i - e_j$ with $1 \leq i < j \leq n$. One has therefore $n$ independent vectors, but $\frac{1}{2}n(n+1)$ invariants given by all the edge lengths squared within $s$ (see Figs. 6.2 and 6.3).

In the interior of a given $n-$simplex one can also assign a second, orthonormal (Lorentz) frame, which we will denote in the following by $\Sigma(s)$. The expansion coefficients relating this orthonormal frame to the one specified by the $n$ directed edges of the simplex associated with the vectors $e_i$ is the lattice analogue of the $n$-bein or tetrad $e_\mu^a$.

Within each $n$-simplex one can define a metric

$$g_{ij}(s) = e_i \cdot e_j \ , \tag{6.2}$$

with $1 \leq i, j \leq n$, and which in the Euclidean case is positive definite. In components one has $g_{ij} = \eta_{ab} e_i^a e_j^b$. In terms of the edge lengths $l_{ij} = |e_i - e_j|$, the metric is given by

$$g_{ij}(s) \;=\; \tfrac{1}{2}\left(l_{0i}^2 + l_{0j}^2 - l_{ij}^2\right) \quad . \tag{6.3}$$

Comparison with the standard expression for the invariant interval $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$ confirms that for the metric in Eq. (6.3) coordinates have been chosen along the $n$ $e_i$ directions.

The volume of a general $n$-simplex is given by the $n$-dimensional generalization of the well-known formula for a tetrahedron, namely

$$V_n(s) \;=\; \frac{1}{n!}\sqrt{\det g_{ij}(s)} \quad . \tag{6.4}$$

An equivalent, but more symmetric, form for the volume of an $n$-simplex can be given in terms of the bordered determinant of an $(n+2) \times (n+2)$ matrix (Wheeler, 1964)

$$V_n(s) \;=\; \frac{(-1)^{\frac{n+1}{2}}}{n!\,2^{n/2}} \begin{vmatrix} 0 & 1 & 1 & \ldots \\ 1 & 0 & l_{01}^2 & \ldots \\ 1 & l_{10}^2 & 0 & \ldots \\ 1 & l_{20}^2 & l_{21}^2 & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ 1 & l_{n,0}^2 & l_{n,1}^2 & \ldots \end{vmatrix}^{1/2} \quad . \tag{6.5}$$

It is possible to associate $p$-forms with lower dimensional objects within a simplex, which will become useful later (Hartle, 1984). With each face $f$ of an $n$-simplex (in the shape of a tetrahedron in four dimensions) one can associate a vector perpendicular to the face

$$\omega(f)_\alpha \;=\; \varepsilon_{\alpha\beta_1\ldots\beta_{n-1}}\, e_{(1)}^{\beta_1} \cdots e_{(n-1)}^{\beta_{n-1}} \quad , \tag{6.6}$$

where $e_{(1)} \ldots e_{(n-1)}$ are a set of oriented edges belonging to the face $f$, and $\varepsilon_{\alpha_1\ldots\alpha_n}$ is the sign of the permutation $(\alpha_1 \ldots \alpha_n)$.

The volume of the face $f$ is then given by

$$V_{n-1}(f) \;=\; \left(\sum_{\alpha=1}^{n} \omega_\alpha^2(f)\right)^{1/2} \quad . \tag{6.7}$$

Similarly, one can consider a hinge (a triangle in four dimensions) spanned by edges $e_{(1)}, \ldots, e_{(n-2)}$. One defines the (un-normalized) hinge bivector

$$\omega(h)_{\alpha\beta} \;=\; \varepsilon_{\alpha\beta\gamma_1\ldots\gamma_{n-2}}\, e_{(1)}^{\gamma_1} \cdots e_{(n-2)}^{\gamma_{n-2}} \quad , \tag{6.8}$$

with the area of the hinge then given by

$$V_{n-2}(h) \;=\; \frac{1}{(n-2)!}\left(\sum_{\alpha<\beta} \omega_{\alpha\beta}^2(h)\right)^{1/2} \quad . \tag{6.9}$$

Next, in order to introduce curvature, one needs to define the *dihedral angle* between faces in an $n$-simplex. In an $n$-simplex $s$ two $n-1$-simplices $f$ and $f'$ will intersect on a common $n-2$-simplex $h$, and the dihedral angle at the specified hinge $h$ is defined as

$$\cos\theta(f,f') = \frac{\omega(f)_{n-1}\cdot\omega(f')_{n-1}}{V_{n-1}(f)\,V_{n-1}(f')} \ , \tag{6.10}$$

where the scalar product appearing on the r.h.s. can be re-written in terms of squared edge lengths using

$$\omega_n\cdot\omega'_n = \frac{1}{(n!)^2}\det(e_i\cdot e'_j) \ , \tag{6.11}$$

and $e_i\cdot e'_j$ in turn expressed in terms of squared edge lengths by the use of Eq. (6.3). (Note that the dihedral angle $\theta$ would have to be defined as $\pi$ minus the arccosine of the expression on the r.h.s. if the orientation for the $e$'s had been chosen in such a way that the $\omega$'s would all point from the face $f$ inward into the simplex $s$). As an example, in two dimensions and within a given triangle, two edges will intersect at a vertex, giving $\theta$ as the angle between the two edges. In three dimensions within a given simplex two triangles will intersect at a given edge, while in four dimension two tetrahedra will meet at a triangle. For the special case of an equilateral $n$-simplex, one has simply $\theta = \arccos\frac{1}{n}$. A related and often used formula for the sine of the dihedral angle $\theta$ is

$$\sin\theta(f,f') = \frac{n}{n-1}\,\frac{V_n(s)\,V_{n-2}(h)}{V_{n-1}(f)\,V_{n-1}(f')} \ , \tag{6.12}$$

but is less useful for practical calculations, as the sine of the angle does not unambiguously determine the angle itself, which is needed in order to compute the local curvature.

In a piecewise linear space curvature is detected by going around elementary loops which are dual to a $(d-2)$-dimensional subspace. From the dihedral angles associated with the faces of the simplices meeting at a given hinge $h$ one can compute the *deficit angle* $\delta(h)$, defined as

$$\delta(h) = 2\pi - \sum_{s\supset h}\theta(s,h) \ , \tag{6.13}$$

where the sum extends over all simplices $s$ meeting on $h$. It then follows that the deficit angle $\delta$ is a measure of the curvature at $h$. The two-dimensional case is illustrated in Fig. 6.4, while the three- and four-dimensional cases are shown in Fig. 6.5.

## 6.4 Rotations, Parallel Transports and Voronoi Loops

Since the interior of each simplex $s$ is assumed to be flat, one can assign to it a Lorentz frame $\Sigma(s)$. Furthermore inside $s$ one can define a $d$-component vector

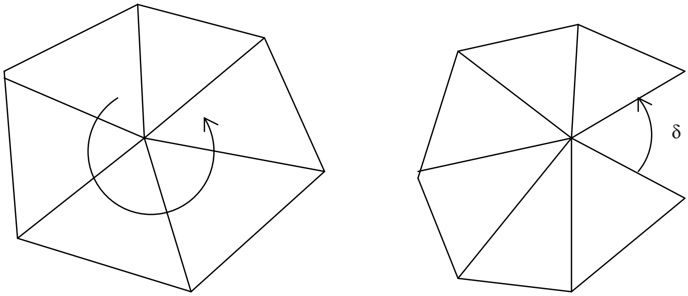



**Fig. 6.4** Illustration of the deficit angle $\delta$ in two dimensions, where several flat triangles meet at a vertex.

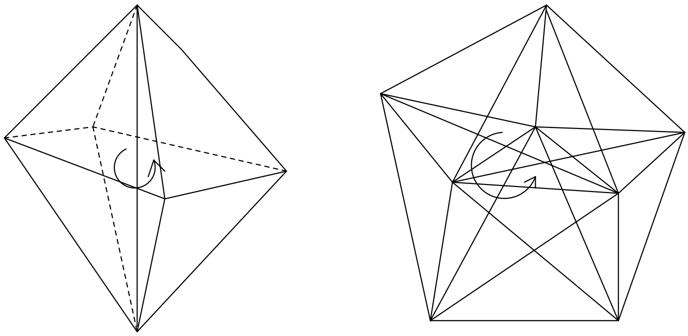


**Fig. 6.5** Deficit angle in three dimensions where several flat tetrahedra meet at an edge, and in four dimensions where several flat four-simplices meet at a triangle.

$\phi(s) = (\phi_0 \dots \phi_{d-1})$. Under a Lorentz transformation of $\Sigma(s)$, described by the $d \times d$ matrix $\Lambda(s)$ satisfying the usual relation for Lorentz transformation matrices

$$\Lambda^T \eta \Lambda = \eta \ , \tag{6.14}$$

the vector $\phi(s)$ will rotate to

$$\phi'(s) = \Lambda(s) \phi(s) \ . \tag{6.15}$$

The base edge vectors $e_i^\mu = l_{0i}^\mu(s)$ themselves are of course an example of such a vector.

Next consider two $d$-simplices, individually labeled by $s$ and $s'$, sharing a common face $f(s,s')$ of dimensionality $d-1$. It will be convenient to label the $d$ edges residing in the common face $f$ by indices $i, j = 1 \dots d$. Within the first simplex $s$ one can then assign a Lorentz frame $\Sigma(s)$, and similarly within the second $s'$ one can assign the frame $\Sigma(s')$. The $\frac{1}{2}d(d-1)$ edge vectors on the common interface $f(s,s')$ (corresponding physically to the same edges, viewed from two different coordinate systems) are expected to be related to each other by a Lorentz rotation $\mathbf{R}$,

$$l_{ij}^{\mu}(s') \;=\; R^{\mu}{}_{\nu}(s',s)\, l_{ij}^{\nu}(s) \;\; . \tag{6.16}$$

Under individual Lorentz rotations in $s$ and $s'$ one has of course a corresponding change in $\mathbf{R}$, namely $\mathbf{R} \rightarrow \Lambda(s')\,\mathbf{R}(s',s)\,\Lambda(s)$. In the Euclidean $d$-dimensional case $\mathbf{R}$ is an orthogonal matrix, element of the group $SO(d)$.

In the absence of torsion, one can use the matrix $\mathbf{R}(s',s)$ to describes the parallel transport of any vector $\phi^{\mu}$ from simplex $s$ to a neighboring simplex $s'$,

$$\phi^{\mu}(s') \;=\; R^{\mu}{}_{\nu}(s',s)\,\phi^{\nu}(s) \;\; . \tag{6.17}$$

$\mathbf{R}$ therefore describes a lattice version of the connection (Lee, 1983). Indeed in the continuum such a rotation would be described by the matrix

$$R^{\mu}{}_{\nu} \;=\; \left(e^{\Gamma \cdot dx}\right)^{\mu}{}_{\nu} \;\; , \tag{6.18}$$

with $\Gamma^{\lambda}_{\mu\nu}$ the affine connection. The coordinate increment $dx$ is interpreted as joining the center of $s$ to the center of $s'$, thereby intersecting the face $f(s,s')$. On the other hand, in terms of the Lorentz frames $\Sigma(s)$ and $\Sigma(s')$ defined within the two adjacent simplices, the rotation matrix is given instead by

$$R^{a}{}_{b}(s',s) \;=\; e^{a}{}_{\mu}(s')\, e^{\nu}{}_{b}(s)\, R^{\mu}{}_{\nu}(s',s) \;\; , \tag{6.19}$$

(this last matrix reduces to the identity if the two orthonormal bases $\Sigma(s)$ and $\Sigma(s')$ are chosen to be the same, in which case the connection is simply given by $R(s',s)_{\mu}{}^{\nu} = e_{\mu}{}^{a}\, e^{\nu}{}_{a}$). Note that it is possible to choose coordinates so that $\mathbf{R}(s,s')$ is the unit matrix for one pair of simplices, but it will not then be unity for all other pairs if curvature is present.

This last set of results will be useful later when discussing lattice Fermions. Let us consider here briefly the problem of how to introduce lattice *spin rotations*. Given in $d$ dimensions the above rotation matrix $\mathbf{R}(s',s)$, the spin connection $\mathbf{S}(s,s')$ between two neighboring simplices $s$ and $s'$ is defined as follows. Consider $\mathbf{S}$ to be an element of the $2^{\nu}$-dimensional representation of the covering group of $SO(d)$, $Spin(d)$, with $d = 2\nu$ or $d = 2\nu + 1$, and for which $S$ is a matrix of dimension $2^{\nu} \times 2^{\nu}$. Then $\mathbf{R}$ can be written in general as

$$\mathbf{R} \;=\; \exp\left[\tfrac{1}{2}\,\sigma^{\alpha\beta}\,\theta_{\alpha\beta}\right] \;\; , \tag{6.20}$$

where $\theta_{\alpha\beta}$ is an antisymmetric matrix The $\sigma$'s are $\frac{1}{2}d(d-1)$ $d \times d$ matrices, generators of the Lorentz group ($SO(d)$ in the Euclidean case, and $SO(d-1,1)$ in the Lorentzian case), whose explicit form is

$$\left[\sigma_{\alpha\beta}\right]^{\gamma}{}_{\delta} \;=\; \delta^{\gamma}{}_{\alpha}\,\eta_{\beta\delta} \;-\; \delta^{\gamma}{}_{\beta}\,\eta_{\alpha\delta} \;\; , \tag{6.21}$$

so that, for example,

$$\sigma_{12} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} . \tag{6.22}$$

For fermions the corresponding spin rotation matrix is then obtained from

$$\mathbf{S} = \exp\left[ \tfrac{i}{4} \gamma^{\alpha\beta} \theta_{\alpha\beta} \right] , \tag{6.23}$$

with generators $\gamma^{\alpha\beta} = \frac{1}{2i}[\gamma^\alpha, \gamma^\beta]$, and with the Dirac matrices $\gamma^\alpha$ satisfying as usual $\gamma^\alpha \gamma^\beta + \gamma^\beta \gamma^\alpha = 2\,\eta^{\alpha\beta}$. Taking appropriate traces, one can obtain a direct relationship between the original rotation matrix $\mathbf{R}(s,s')$ and the corresponding spin rotation matrix $\mathbf{S}(s,s')$

$$R_{\alpha\beta} = \operatorname{tr}\left( \mathbf{S}^\dagger \gamma_\alpha \mathbf{S} \gamma_\beta \right) / \operatorname{tr} \mathbf{1} , \tag{6.24}$$

which determines the spin rotation matrix up to a sign.

One can consider a sequence of rotations along an arbitrary path $P(s_1, \ldots, s_{n+1})$ going through simplices $s_1 \ldots s_{n+1}$, whose combined rotation matrix is given by

$$\mathbf{R}(P) = \mathbf{R}(s_{n+1}, s_n) \cdots \mathbf{R}(s_2, s_1) , \tag{6.25}$$

and which describes the parallel transport of an arbitrary vector from the interior of simplex $s_1$ to the interior of simplex $s_{n+1}$,

$$\phi^\mu(s_{n+1}) = R^\mu{}_\nu(P)\, \phi^\nu(s_1) . \tag{6.26}$$

If the initial and final simplices $s_{n+1}$ and $s_1$ coincide, one obtains a closed path $C(s_1, \ldots, s_n)$, for which the associated expectation value can be considered as the gravitational analog of the Wilson loop. Its combined rotation is given by

$$\mathbf{R}(C) = \mathbf{R}(s_1, s_n) \cdots \mathbf{R}(s_2, s_1) . \tag{6.27}$$

Under Lorentz transformations within each simplex $s_i$ along the path one has a pairwise cancellation of the $\Lambda(s_i)$ matrices except at the endpoints, giving in the closed loop case

$$\mathbf{R}(C) \to \Lambda(s_1)\, \mathbf{R}(C)\, \Lambda^T(s_1) . \tag{6.28}$$

Clearly the deviation of the matrix $\mathbf{R}(C)$ from unity is a measure of curvature. Also, the trace $\operatorname{tr}\mathbf{R}(C)$ is independent of the choice of Lorentz frames.

Of particular interest is the elementary loop associated with the smallest non-trivial, segmented parallel transport path one can build on the lattice. One such polygonal path in four dimensions is shown in Fig. 6.6. In general consider a $(d-2)$-dimensional simplex (hinge) $h$, which will be shared by a certain number $m$ of $d$-simplices, sequentially labeled by $s_1 \ldots s_m$, and whose common faces $f(s_1, s_2) \ldots f(s_{m-1}, s_m)$ will also contain the hinge $h$. Thus in four dimensions several four-simplices will contain, and therefore encircle, a given triangle (hinge).

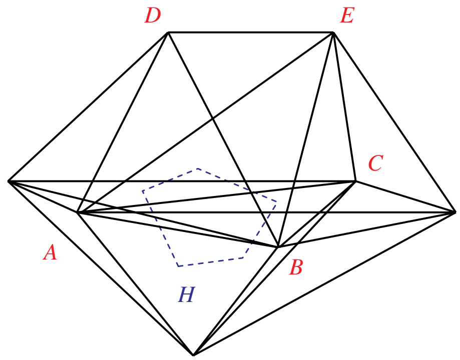



**Fig. 6.6** Elementary polygonal path around a hinge (*triangle*) in four dimensions. The hinge *ABC*, contained in the simplex *ABCDE*, is encircled by the polygonal path *H* connecting the surrounding vertices, which reside in the dual lattice. One such vertex is contained within the simplex *ABCDE*.

In three dimensions the path will encircle an edge, while in two dimensions it will encircle a site. Thus for each hinge $h$ there is a unique elementary closed path $C_h$ for which one again can define the ordered product

$$\mathbf{R}(C_h) \;=\; \mathbf{R}(s_1, s_m) \cdots \mathbf{R}(s_2, s_1) \;. \tag{6.29}$$

The hinge $h$, being geometrically an object of dimension $(d-2)$, is naturally represented by a tensor of rank $(d-2)$, referred to a coordinate system in $h$: an edge vector $l_h^\mu$ in $d=3$, and an area bi-vector $\frac{1}{2}(l_h^\mu l_h'^\nu - l_h^\nu l_h'^\mu)$ in $d=4$ etc. Following Eq. (6.8) it will therefore be convenient to define a hinge bi-vector $U$ in any dimension as

$$U_{\mu\nu}(h) \;=\; \mathscr{N}\, \varepsilon_{\mu\nu\alpha_1 \alpha_{d-2}}\, l_{(1)}^{\alpha_1} \cdots l_{(d-2)}^{\alpha_{d-2}} \;, \tag{6.30}$$

normalized, by the choice of the constant $\mathscr{N}$, in such a way that $U_{\mu\nu}U^{\mu\nu} = 2$. In four dimensions

$$U_{\mu\nu}(h) \;=\; \frac{1}{2A_h}\, \varepsilon_{\mu\nu\alpha\beta}\, l_1^\alpha\, l_2^\beta \;, \tag{6.31}$$

where $l_1(h)$ and $l_2(h)$ two independent edge vectors associated with the hinge $h$, and $A_h$ the area of the hinge.

An important aspect related to the rotation of an arbitrary vector, when parallel transported around a hinge $h$, is the fact that, due to the hinge's intrinsic orientation, only components of the vector in the plane perpendicular to the hinge are affected. Since the direction of the hinge $h$ is specified locally by the bivector $U_{\mu\nu}$ of Eq. (6.31), one can write for the loop rotation matrix $\mathbf{R}$

$$R^\mu{}_\nu(C) \;=\; \left(e^{\delta U}\right)^\mu{}_\nu \;, \tag{6.32}$$

where $C$ is now the small polygonal loop entangling the hinge $h$, and $\delta$ the deficit angle at $h$, previously defined in Eq. (6.13). One particularly noteworthy aspect of this last result is the fact that the area of the loop $C$ does not enter in the expression for the rotation matrix, only the deficit angle and the hinge direction.

At the same time, in the continuum a vector $V$ carried around an infinitesimal loop of area $A_C$ will change by

$$\Delta V^{\mu} \;=\; \tfrac{1}{2} R^{\mu}{}_{\nu\lambda\sigma} A^{\lambda\sigma} V^{\nu} \;\; , \tag{6.33}$$

where $A^{\lambda\sigma}$ is an area bivector in the plane of $C$, with squared magnitude $A_{\lambda\sigma}A^{\lambda\sigma} = 2A_C^2$. Since the change in the vector $V$ is given by $\delta V^{\alpha} = (\mathbf{R}-\mathbf{1})^{\alpha}{}_{\beta} V^{\beta}$ one is led to the identification

$$\tfrac{1}{2} \, R^{\alpha}{}_{\beta\mu\nu} A^{\mu\nu} \;=\; (\mathbf{R}-\mathbf{1})^{\alpha}{}_{\beta} \;\; . \tag{6.34}$$

Thus the above change in $V$ can equivalently be re-written in terms of the infinitesimal rotation matrix

$$R^{\mu}{}_{\nu}(C) \;=\; \left( e^{\frac{1}{2} R \cdot A} \right)^{\mu}{}_{\nu} \;\; , \tag{6.35}$$

(where the Riemann tensor appearing in the exponent on the r.h.s. should not be confused with the rotation matrix $\mathbf{R}$ on the l.h.s.).

It is then immediate to see that the two expressions for the rotation matrix $\mathbf{R}$ in Eqs. (6.32) and (6.35) will be compatible provided one uses for the Riemann tensor at a hinge $h$ the expression

$$R_{\mu\nu\lambda\sigma}(h) \;=\; \frac{\delta(h)}{A_C(h)} \, U_{\mu\nu}(h) \, U_{\lambda\sigma}(h) \;\; , \tag{6.36}$$

expected to be valid in the limit of small curvatures, with $A_C(h)$ the area of the loop entangling the hinge $h$. Here use has been made of the geometric relationship $U_{\mu\nu}A^{\mu\nu} = 2A_C$. Note that the bivector $U$ has been defined to be perpendicular to the $(d-2)$ edge vectors spanning the hinge $h$, and lies therefore in the same plane as the loop $C$. Furthermore, the expression of Eq. (6.36) for the *Riemann tensor at a hinge* has the correct algebraic symmetry properties, such as the antisymmetry in the first and second pair of indices, as well as the swap symmetry between first and second pair, and is linear in the curvature, with the correct dimensions of one over length squared.

The area $A_C$ is most suitably defined by introducing the notion of a *dual lattice*, i.e. a lattice constructed by assigning centers to the simplices, with the polygonal curve $C$ connecting these centers sequentially, and then assigning an area to the interior of this curve. One possible way of assigning such centers is by introducing perpendicular bisectors to the faces of a simplex, and locate the vertices of the dual lattice at their common intersection, a construction originally discussed in (Voronoi, 1908) and in (Meijering, 1953). Another, and perhaps even simpler, possibility is to use a barycentric subdivision (Singer and Thorpe, 1967).

## 6.5 Invariant Lattice Action

The first step in writing down an invariant lattice action, analogous to the continuum Einstein-Hilbert action, is to find the lattice analogue of the Ricci scalar. From the expression for the Riemann tensor at a hinge given in Eq. (6.36) one obtains by contraction

$$R(h) \;=\; 2\,\frac{\delta(h)}{A_C(h)} \quad . \tag{6.37}$$

The continuum expression $\sqrt{g}R$ is then obtained by multiplication with the volume element $V(h)$ associated with a hinge. The latter is defined by first joining the vertices of the polyhedron $C$, whose vertices lie in the dual lattice, with the vertices of the hinge $h$, and then computing its volume.

By *defining* the polygonal area $A_C$ as $A_C(h) = d\,V(h)/V^{(d-2)}(h)$, where $V^{(d-2)}(h)$ is the volume of the hinge (an area in four dimensions), one finally obtains for the Euclidean lattice action for pure gravity

$$I_R(l^2) \;=\; -\,k \sum_{\text{hinges h}} \delta(h)\,V^{(d-2)}(h) \quad , \tag{6.38}$$

with the constant $k = 1/(8\pi G)$. One would have obtained the same result for the single-hinge contribution to the lattice action if one had contracted the infinitesimal form of the rotation matrix $R(h)$ in Eq. (6.32) with the hinge bivector $\omega_{\alpha\beta}$ of Eq. (6.8) (or equivalently with the bivector $U_{\alpha\beta}$ of Eq. (6.31) which differs from $\omega_{\alpha\beta}$ by a constant). The fact that the lattice action only involves the content of the hinge $V^{(d-2)}(h)$ (the area of a triangle in four dimensions) is quite natural in view of the fact that the rotation matrix at a hinge in Eq. (6.32) only involves the deficit angle, and not the polygonal area $A_C(h)$.

An alternative form for the lattice action (Fröhlich, 1981) can be obtained instead by contracting the elementary rotation matrix $\mathbf{R}(C)$ of Eq. (6.32), and not just its infinitesimal form, with the hinge bivector of Eq. (6.8),

$$I_{\text{com}}(l^2) \;=\; -\,k \sum_{\text{hinges h}} \tfrac{1}{2}\,\omega_{\alpha\beta}(h)\,R^{\alpha\beta}(h) \quad . \tag{6.39}$$

The above construction can be regarded as analogous to Wilson's lattice gauge theory, for which the action also involves traces of products of $SU(N)$ color rotation matrices (Wilson, 1973). For small deficit angles one can of course use $\omega_{\alpha\beta} = (d-2)!\,V^{(d-2)}\,U_{\alpha\beta}$ to show the equivalence of the two lattice actions.

But in general, away from a situation of small curvatures, the two lattice action are not equivalent, as can be seen already in two dimensions. Writing the rotation matrix at a hinge as $\mathbf{R}(h) = \begin{pmatrix} \cos\delta & \sin\delta \\ -\sin\delta & \cos\delta \end{pmatrix}$, expressed for example in terms of Pauli matrices, and taking the appropriate trace ($\omega_{\alpha\beta} = \varepsilon_{\alpha\beta}$ in two dimensions) one finds

$$\operatorname{tr}\left[\tfrac{1}{2}(-i\sigma_y)(\cos\delta_p + i\sigma_y\sin\delta_p)\right] \;=\; \sin\delta_p \quad , \tag{6.40}$$

and therefore $I_{\rm com} = -k \sum_p \sin \delta_p$. In general one can show that the compact action $I_{com}$ in $d$ dimensions involves the sine of the deficit angle, instead of just the angle itself as in the Regge case. In the weak field limit the two actions should lead to similar expansions, while away from the weak field limit one would have to verify that the same universal long distance properties are recovered.

The preceding observations can in fact be developed into a consistent first order (Palatini) formulation of Regge gravity, with suitably chosen independent transformation matrices and metrics, related to each other by a set of appropriate lattice equations of motion (Caselle, D'Adda and Magnea, 1989). Ultimately one would expect the first and second order formulations to describe the same quantum theory, with common universal long-distance properties. How to consistently define finite rotations, frames and connections in Regge gravity was first discussed systematically in (Fröhlich, 1981).

One important result that should be mentioned in this context is the rigorous proof of convergence in the sense of measures of the Regge lattice action towards the continuum Einstein-Hilbert action (Cheeger, Müller and Schrader, 1984). Some general aspects of this result have recently been reviewed from a mathematical point of view in (Lafontaine, 1986). A derivation of the Regge action from its continuum counterpart was later presented in (Lee, Feinberg and Friedberg and Reu, 1984).

Other terms can be added to the lattice action. Consider for example a cosmological constant term, which in the continuum theory takes the form $\lambda_0 \int d^d x \sqrt{g}$. The expression for the cosmological constant term on the lattice involves the total volume of the simplicial complex. This may be written as

$$V_{\textbf{total}} = \sum_{\text{simplices s}} V_s \ , \tag{6.41}$$

or equivalently as

$$V_{\textbf{total}} = \sum_{\text{hinges h}} V_h \ , \tag{6.42}$$

where $V_h$ is the volume associated with each hinge via the construction of a dual lattice, as described above. Thus one may regard the local volume element $\sqrt{g}\, d^d x$ as being represented by either $V_h$ (centered on $h$) or $V_s$ (centered on $s$).

The Regge and cosmological constant term then lead to the combined action

$$I_{\rm latt}(l^2) \ = \ \lambda_0 \sum_{\text{simplices s}} V_s^{(d)} - \sum_{\text{hinges h}} \delta_h V_h^{(d-2)} \ . \tag{6.43}$$

One would then write for the lattice regularized version of the Euclidean Feynman path integral

$$Z_{\rm latt}(\lambda_0, k) \ = \ \int [d\, l^2] \exp\left(-I_{\rm latt}(l^2)\right) \ , \tag{6.44}$$

where $[d\, l^2]$ is an appropriate functional integration measure over the edge lengths, to be discussed later.

The structure of the gravitational action of Eq. (6.43) leads naturally to some rather general observations, which we will pursue here. The first, cosmological constant, term represents the total four-volume of space-time. As such, it does not contain any derivatives (or finite differences) of the metric and is completely local; it does not contribute to the propagation of gravitational degrees of freedom and is more akin to a mass term (as is already clear from the weak field expansion of $\int \sqrt{g}$ in the continuum). In an ensemble in which the total four-volume is fixed in the thermodynamic limit (number of simplices tending to infinity) one might in fact take the lattice coupling $\lambda_0 = 1$, since different values of $\lambda_0$ just correspond to a trivial rescaling of the overall four-volume (of course in a traditional renormalization group approach to field theory, the overall four-volume is always kept fixed while the scale or $q^2$ dependence of the action and couplings are investigated). Alternatively, one might even want to choose directly an ensemble for which the probability distribution in the total four-volume $V$ is

$$\mathscr{P}(V) \propto \delta(V - V_0) \ , \tag{6.45}$$

in analogy with the microcanonical ensemble of statistical mechanics.

The second, curvature contribution to the action contains, as in the continuum, the proper kinetic term. This should already be clear from the derivation of the lattice action given above, and will be made even more explicit in the section dedicated to the lattice weak field expansion. Such a term now provides the necessary coupling between neighboring lattice metrics, but the coupling still remains local. Geometrically, it can be described as a sum of elementary loop contributions, as it contains as its primary ingredient the deficit angle associated with an elementary parallel transport loop around the hinge $h$. When $k = 0$ one resides in the extreme strong coupling regime There the fluctuations in the metric are completely unconstrained by the action, insofar as only the total four-volume of the manifold is kept constant.

At this point it might be useful to examine some specific cases with regards to the overall dimensionality of the simplicial complex. In *two dimensions* the Regge action reduces to a sum over lattice sites $p$ of the $2-d$ deficit angle, giving the discrete analog of the Gauss-Bonnet theorem

$$\sum_{\text{sites p}} \delta_p = 2\pi\chi \ , \tag{6.46}$$

where $\chi = 2 - 2g$ is the Euler characteristic of the surface, and $g$ the genus (the number of handles). In this case the action is therefore a topological invariant, and the above lattice expression is therefore completely analogous to the well known continuum result

$$\tfrac{1}{2}\int d^2x\sqrt{g}R = 2\pi\chi \ . \tag{6.47}$$

This remarkable identity ensures that two-dimensional lattice $R$-gravity is as trivial as in the continuum, since the variation of the local action density under a small variation of an edge length $l_{ij}$ is still zero. Of course there is a much simpler formula for the Euler characteristic of a simplicial complex, namely

$$\chi = \sum_{i=0}^{d} (-1)^i N_i \ , \tag{6.48}$$

where $N_i$ is the number of simplices of dimension $i$. Also it should be noted that in two dimensions the compact action $I_{com} = \sum_p \sin\delta_p$ does not satisfy the Gauss-Bonnet relation.

In *three dimensions* the Regge lattice action reads

$$I_R = -k \sum_{\text{edges h}} l_h \, \delta_h \ , \tag{6.49}$$

where $\delta_h$ is the deficit angle around the edge labeled by $h$. Variation with respect to an edge length $l_h$ gives two terms, of which only the term involving the variation of the edge is non-zero

$$\delta I_R = -k \sum_{\text{edges h}} \delta l_h \cdot \delta_h \ . \tag{6.50}$$

In fact it was shown by Regge that for any $d > 2$ the term involving the variation of the deficit angle does not contribute to the equations of motion (just as in the continuum the variation of the Ricci tensor does not contribute to the equations of motion either). Therefore in three dimensions the lattice equations of motion, in the absence of sources and cosmological constant term, reduce to

$$\delta_h \ = \ 0 \ , \tag{6.51}$$

implying that all deficit angles have to vanish, i.e. a flat space.

In four dimensions variation of $I_R$ with respect to the edge lengths gives the simplicial analogue of Einstein's field equations, whose derivation is again, as mentioned, simplified by the fact that the contribution from the variation of the deficit angle is zero

$$\delta I_R = \sum_{\text{triangles h}} \delta(A_h) \cdot \delta_h \ . \tag{6.52}$$

In the discrete case the field equations reduce therefore to

$$\lambda_0 \sum_s \frac{\partial V_s}{\partial l_{ij}} - k \sum_h \delta_h \frac{\partial A_h}{\partial l_{ij}} \ = \ 0 \ , \tag{6.53}$$

and the derivatives can then be worked out for example from Eq. (6.5). Alternatively, a rather convenient and compact expression can be given (Hartle, 1984) for the derivative of the squared volume $V_n^2$ of an arbitrary $n$-simplex with respect to one of its squared edge lengths

$$\frac{\partial V_n^2}{\partial l_{ij}^2} \ = \ \frac{1}{n^2} \, \omega_{n-1} \cdot \omega'_{n-1} \ , \tag{6.54}$$

where the $\omega^{\alpha}_{(n-1)}$'s (here referring to two $(n-1)$ simplices part of the same $n$-simplex) are given in Eqs. (6.6) and (6.11). In the above expression the $\omega$'s are meant to be associated with vertex labels $0,\ldots,i-1,i+1,\ldots,n$ for $\omega_{n-1}$, and $0,\ldots,j-1,j+1,\ldots,n$ for $\omega'_{n-1}$ respectively.

Then in the absence of a cosmological term one finds the remarkably simple expression for the lattice field equations

$$\tfrac{1}{2}\, l_p \sum_{h \supset l_p} \delta_h \cot\theta_{ph} \;=\; 0 \;, \tag{6.55}$$

where the sum is over hinges (triangles) labeled by $h$ meeting on the common edge $l_p$, and $\theta_{ph}$ is the angle in the hinge $h$ opposite to the edge $l_p$. This is illustrated in Fig. 6.7.

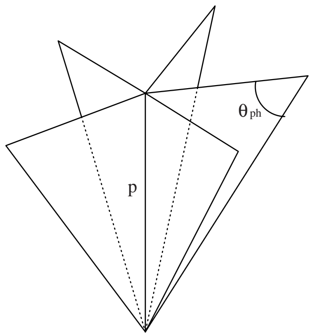



**Fig. 6.7** Angles appearing in the Regge equations.

The discrete equations given above represent the lattice analogs of the Einstein field equations in a vacuum, for which suitable solutions can be searched for by adjusting the edge lengths. Since the equations are in general non-linear, the existence of multiple solutions cannot in general be ruled out (Misner, Thorne and Wheeler, 1973). A number of papers have addressed the general issue of convergence to the continuum in the framework of the classical formulation (Brewin and Gentle, 2001). Several authors have discussed non-trivial applications of the Regge equations to problems in classical general relativity such as the Schwarzschild and Reissner-Nordstrom geometries (Wong, 1971), the Friedmann and Tolman universes (Collins and Williams, 1973; 1974), and the problem of radial motion and circular (actually polygonal) orbits (Williams and Ellis, 1981; 1984). Spherically symmetric, as well as more generally inhomogeneous, vacuum spacetimes were studied using a discrete $3+1$ formulation with a variety of time-slicing prescriptions in (Porter, 1987a,b,c), and later extended (Dubal, 1989a,b) to a systematic investigation of the axis-symmetric non-rotating vacuum solutions and to the problem of relativistic spherical collapse for polytropic perfect fluids.

In classical gravity the general time evolution problem plays of course a central role. The $3+1$ time evolution problem in Regge gravity was discussed originally in (Sorkin, 1975a,b) and later re-examined from a numerical, practical prespective in

(Barrett et al, 1997) using a discrete time step formulation, whereas in (Piran and Williams, 1986) a continuous time fomalism was proposed. The choice of lapse and shift functions in Regge gravity were discussed further in (Tuckey, 1989; Galassi, 1993) and in (Gentle and Miller, 1998), and applied to the Kasner cosmology in the last reference. An alternative so-called null-strut approach was proposed in (Miller and Wheeler, 1985) which builds up a spacelike-foliated spacetime with a maximal number of null edges, but seems difficult to implement in practice. Finally in (Khatsymovsky, 1991) and (Immirzi, 1996) a continuous time Regge gravity formalism in the tetrad-connection variables was developed, in part targeted towards quantum gravity calculations. A recent comprehensive review of classical applications of Regge gravity can be found for example in (Gentle, 2002), as well as a more complete set of references.

## 6.6 Lattice Diffeomorphism Invariance

Consider the two-dimensional flat skeleton shown in Fig. 6.8. It is clear that one can move around a point on the surface, keeping all the neighbors fixed, without violating the triangle inequalities and leave all curvature invariants unchanged.

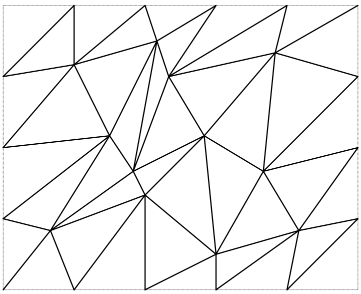


**Fig. 6.8** On a random simplicial lattice there are in general no preferred directions.

In $d$ dimensions this transformation has $d$ parameters and is an exact invariance of the action. When space is slightly curved, the invariance is in general only an approximate one, even though for piecewise linear spaces piecewise diffeomorphisms can still be defined as the set of local motions of points that leave the local contribution to the action, the measure and the lattice analogues of the continuum curvature invariants unchanged (Hamber and Williams, 1998). Note that in general the gauge deformations of the edges are still constrained by the triangle inequalities. The general situation is illustrated in Figs. 6.8, 6.9 and 6.10. In the limit when the number of edges becomes very large, the full continuum diffeomorphism group should be recovered.

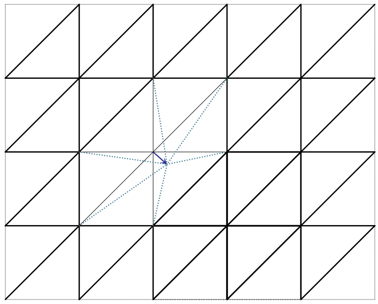


**Fig. 6.9** Example of a lattice diffeomorphism, the local gauge transform of a flat lattice, corresponding in $d$ dimensions to a $d$-parameter local deformation of the edges.

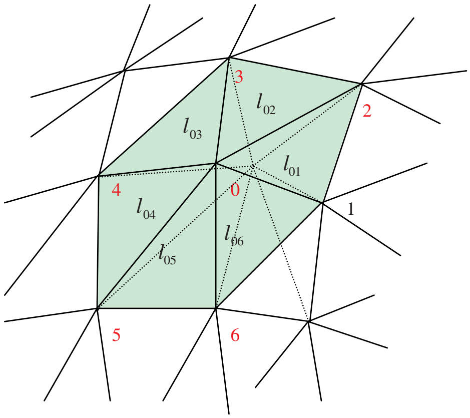



**Fig. 6.10** Another example of a lattice diffeomorphism, the gauge deformation of a lattice around a vertex 0, leaving the local action contribution from that vertex unchanged.

In general the structure of lattice local gauge transformations is rather complicated and will not be given here; it can be found in the above quoted reference. These are defined as transformations acting locally on a given set of edges which leave the local lattice curvature invariant. The simplest context in which this local invariance can be exhibited explicitly is the lattice weak field expansion, which will be discussed later in Sect. 7.2. The local gauge invariance corresponding to continuum diffeomorphism is given there in Eq. (7.9). From the transformation properties of the edge lengths it is clear that their transformation properties are related to those of the local metric, as already suggested for example by the identification of Eqs. (6.3) and (6.68). In the quantum theory, a local gauge invariance implies the existence of Ward identities for $n$-point functions.

## 6.7 Lattice Bianchi Identities

Consider therefore a closed path $C_h$ encircling a hinge $h$ and passing through each of the simplices that meet at that hinge. In particular one may take $C_h$ to be the boundary of the polyhedral dual (or Voronoi) area surrounding the hinge. We recall that the Voronoi polyhedron dual to a vertex $P$ is the set of all points on the lattice which are closer to $P$ than any other vertex; the corresponding new vertices then represent the sites on the dual lattice. A unique closed parallel transport path can then be assigned to each hinge, by suitably connecting sites in the dual lattice.

With each neighboring pair of simplices $s, s+1$ one associates a Lorentz transformation $\mathbf{R}^{\alpha}{}_{\beta}(s, s+1)$, which describes how a given vector $V_\mu$ transforms between the local coordinate systems in these two simplices As discussed previously, the above transformation is directly related to the continuum path-ordered ($P$) exponential of the integral of the local affine connection $\Gamma^{\lambda}_{\mu\nu}$ via

$$R^{\mu}{}_{\nu} = \left[ \mathscr{P} \, e^{\int_{\textbf{between simplices}}^{\textbf{path}} \Gamma^{\lambda} dx_{\lambda}} \right]^{\mu}_{\nu} . \tag{6.56}$$

The connection here has support only on the common interface between the two simplices.

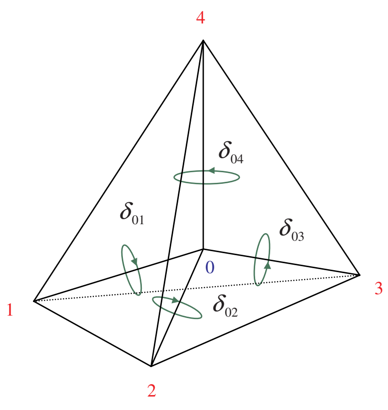
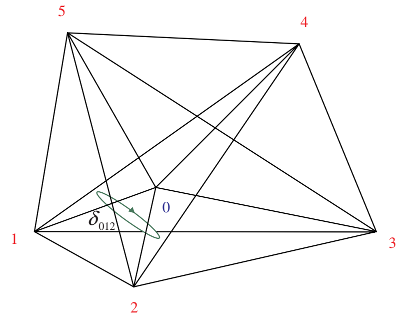



**Fig. 6.11** Illustration of the exact lattice Bianchi identity in the case of three dimensions, where several hinges (*edges*) meet on a vertex. The combined rotation for a path that sequentially encircles several hinges and which can be shrunk to a point is given by the identity matrix.

Just as in the continuum, where the affine connection and therefore the infinitesimal rotation matrix is determined by the metric and its first derivatives, on the lattice one expects that the elementary rotation matrix between simplices $\mathbf{R}_{s,s+1}$ is fixed by the difference between the $g_{ij}$'s of Eq. (6.3) within neighboring simplices.

For a vector $V$ transported once around a Voronoi loop, i.e. a loop formed by Voronoi edges surrounding a chosen hinge, the change in the vector $V$ is given by

**Fig. 6.12** A second illustration of the exact lattice Bianchi identity, now in four dimensions. Here several hinges (*triangles*) meet at the vertex labelled by 0. Around each hinge one has a corresponding rotation and therefore a deficit angle $\delta$. The product of rotation matrices that sequentially encircle several hinges and is topologically trivial gives the identity matrix.




$$\delta V^{\alpha} = (\mathbf{R} - \mathbf{1})^{\alpha}_{\ \beta} V^{\beta} \quad , \tag{6.57}$$

where $\mathbf{R} \equiv \prod_s \mathbf{R}_{s,s+1}$ is now the total rotation matrix associated with the given hinge, given by

$$\left[ \prod_s \mathbf{R}_{s,s+1} \right]^{\mu}_{\ \nu} = \left[ e^{\delta(h) U(h)} \right]^{\mu}_{\ \nu} \quad . \tag{6.58}$$

It is these lattice parallel transporters around closed elementary loops that satisfy the lattice analogues of the Bianchi identities. These are derived by considering paths which encircle more than one hinge and yet are topologically trivial, in the sense that they can be shrunk to a point without entangling any hinge (Regge, 1961; Roček and Williams, 1985).

Thus, for example, the ordered product of rotation matrices associated with the triangles meeting on a given edge has to give one, since a single path can be constructed which sequentially encircles all the triangles and is topologically trivial

$$\prod_{\substack{\textbf{hinges h} \\ \textbf{meeting on edge p}}} \left[ e^{\delta(h) U(h)} \right]^{\mu}_{\ \nu} = 1 \quad . \tag{6.59}$$

Other identities might be derived by considering paths that encircle several hinges meeting on one point. Regge has shown that the above lattice relations correspond precisely to the continuum Bianchi identities. One can therefore explicitly construct exact lattice analogues of the continuum uncontracted, partially contracted, and fully contracted Bianchi identities. The lattice Bianchi identities are illustrated in Fig. 6.11 for the three-dimensional case, and Fig. 6.12 for the four-dimensional case.

The resulting lattice equations are quite similar in structure to the Bianchi identities in $SU(N)$ lattice gauge theories, where one considers identities arising from the multiplication of group elements associated with the square faces of a single cube part of a hypercubic lattice (Wilson, 1973). The motivation there was the possible replacement of the integration over the group elements by an integration over the "plaquette variables" associated with an elementary square (thereby involving the ordered product of four group elements), provided the Bianchi identity constraint is included as well in the lattice path integral.

## 6.8 Gravitational Wilson Loop

We have seen that with each neighboring pair of simplices $s, s+1$ one can associate a Lorentz transformation $\mathbf{R}^{\mu}{}_{\nu}(s, s+1)$, which describes how a given vector $V^{\mu}$ transforms between the local coordinate systems in these two simplices, and that the above transformation is directly related to the continuum path-ordered ($P$) exponential of the integral of the local affine connection $\Gamma^{\lambda}_{\mu\nu}(x)$ via

$$R^{\mu}{}_{\nu} \;=\; \left[ P\, e^{\int_{\substack{\textbf{path}\\ \textbf{between simplices}}} \Gamma_{\lambda}\, dx^{\lambda}} \right]^{\mu}_{\;\nu} \,, \tag{6.60}$$

with the connection having support only on the common interface between the two simplices. Also, for a closed elementary path $C_h$ encircling a hinge $h$ and passing through each of the simplices that meet at that hinge one has for the total rotation matrix $\mathbf{R} \equiv \prod_s \mathbf{R}_{s,s+1}$ associated with the given hinge

$$\left[ \prod_s \mathbf{R}_{s,s+1} \right]^{\mu}_{\;\nu} \;=\; \left[ e^{\delta(h) U(h)} \right]^{\mu}_{\;\nu} \,. \tag{6.61}$$

Equivalently, this last expression can be re-written in terms of a surface integral of the Riemann tensor, projected along the surface area element bivector $A^{\alpha\beta}(C_h)$ associated with the loop,

$$\left[ \prod_s \mathbf{R}_{s,s+1} \right]^{\mu}_{\;\nu} \;\approx\; \left[ e^{\frac{1}{2} \int_S R^{\cdot}{}_{\cdot\alpha\beta} A^{\alpha\beta}(C_h)} \right]^{\mu}_{\;\nu} \,. \tag{6.62}$$

More generally one might want to consider a near-planar, but non-infinitesimal, closed loop $C$, as shown in Fig. 7.9. Along this closed loop the overall rotation matrix will still be given by

$$R^{\mu}{}_{\nu}(C) \;=\; \left[ \prod_{s \subset C} \mathbf{R}_{s,s+1} \right]^{\mu}_{\;\nu} \,. \tag{6.63}$$

In analogy with the infinitesimal loop case, one would like to state that for the overall rotation matrix one has

$$R^{\mu}{}_{\nu}(C) \;\approx\; \left[ e^{\delta(C) U(C))} \right]^{\mu}_{\;\nu} \,, \tag{6.64}$$

where $U_{\mu\nu}(C)$ is now an area bivector perpendicular to the loop - which will work only if the loop is close to planar so that $U_{\mu\nu}$ can be taken to be approximately constant along the path $C$.

If that is true, then one can define, again in analogy with the infinitesimal loop case, an appropriate coordinate scalar by contracting the above rotation matrix $\mathbf{R}(C)$ with the bivector of Eq. (6.8), namely

$$W(C) \; = \; \omega_{\alpha\beta}(C) \, R^{\alpha\beta}(C) \;\; , \tag{6.65}$$

where the loop bivector, $\omega_{\alpha\beta}(C) = (d-2)! \, V^{(d-2)} \, U_{\alpha\beta} = 2 A_C \, U_{\alpha\beta}(C)$ in four dimensions, is now intended as being representative of the overall geometric features of the loop. For example, it can be taken as an average of the hinge bivector $\omega_{\alpha\beta}(h)$ along the loop.

In the quantum theory one is of course interested in the average of the above loop operator $W(C)$, as in Eq. (3.144). The previous construction is indeed quite analogous to the Wilson loop definition in ordinary lattice gauge theories (Wilson, 1973), where it is defined via the trace of path ordered products of $SU(N)$ color rotation matrices. In gravity though the Wilson loop does not give any information about the static potential (Modanese, 1993; Hamber, 1994). It seems that the Wilson loop in gravity provides instead some insight into the large-scale curvature of the manifold, just as the infinitesimal loop contribution entering the lattice action of Eqs. (6.38) and (6.39) provides, through its averages, insight into the very short distance, local curvature. Of course for any continuum manifold one can define locally the parallel transport of a vector around a near-planar loop $C$. Indeed parallel transporting a vector around a closed loop represents a suitable operational way of detecting curvature locally. If the curvature of the manifold is small, one can treat the larger loop the same way as the small one; then the expression of Eq. (6.64) for the rotation matrix $\mathbf{R}(C)$ associated with a near-planar loop can be re-written in terms of a surface integral of the large-scale Riemann tensor, projected along the surface area element bivector $A^{\alpha\beta}(C)$ associated with the loop,

$$R^{\mu}{}_{\nu}(C) \; \approx \; \Big[ e^{\frac{1}{2} \int_S R^{\cdot}{}_{\cdot\,\alpha\beta} \, A^{\alpha\beta}(C)} \Big]^{\mu}{}_{\nu} \;\; . \tag{6.66}$$

Thus a direct calculation of the Wilson loop provides a way of determining the *effective* curvature at large distance scales, even in the case where short distance fluctuations in the metric may be significant. Conversely, the rotation matrix appearing in the elementary Wilson loop of Eqs. (6.29) and (6.32) only provides information about the parallel transport of vectors around *infinitesimal* loops, with size comparable to the ultraviolet cutoff.

One would expect that for a geometry fluctuating strongly at short distances (corresponding therefore to the small $k$ limit) the infinitesimal parallel transport matrices $\mathbf{R}(s,s')$ should be distributed close to randomly, with a measure close to the uniform Haar measure, and with little correlation between neighboring hinges. In such instance one would have for the local quantum averages of the infinitesimal lattice parallel transports $< \mathbf{R} >= 0$, but $< \mathbf{R}\mathbf{R}^{-1} > \neq 0$, which would require, for a non-vanishing lowest order contribution to the Wilson loop, that the loop at least be tiled by elementary action contributions from Eqs. (6.38) or (6.39), thus forming a minimal surface spanning the loop. Then, in close analogy to the Yang-Mills case of Eq. (3.145) (see for example Peskin and Schroeder, 1995), the leading contribution to the gravitational Wilson loop would be expected to follow an area law,

$$W(C) \sim \text{const.}\, k^{A(C)} \sim \exp(-A(C)/\xi^2) \ , \tag{6.67}$$

where $A(C)$ is the minimal physical area spanned by the near-planar loop $C$, and $\xi = 1/\sqrt{|\ln k|}$ the gravitational correlation length at small $k$. Thus for a close-to-circular loop of perimeter $P$ one would use $A(C) \approx P^2/4\pi$.

The rapid decay of the gravitational Wilson loop as a function of the area is seen here simply as a general and direct consequence of the disorder in the fluctuations of the parallel transport matrices $\mathbf{R}(s,s')$ at strong coupling. It should then be clear from the above discussion that the gravitational Wilson loop provides in a sense a measure of the magnitude of the large-scale, averaged curvature, where the latter is most suitably defined by the process of parallel-transporting test vectors around very large loops, and is therefore, from the above expression, computed to be of the order $R \sim 1/\xi^2$, at least in the small $k$ limit. A direct calculation of the Wilson loop should therefore Provide, among other things, a direct insight into whether the manifold is de Sitter or anti-de Sitter at large distances. More details on the lattice construction of the gravitational Wilson loop, the various issues that arise in its precise definition, and a sample calculation in the strong coupling limit of lattice gravity, can be found in (Hamber and Williams, 2007).

Finally we note that the definition of the gravitational Wilson loop is based on a surface with a given boundary $C$, in the simplest case the minimal surface spanning the loop. It is possible though to consider other surfaces built out of elementary parallel transport loops. An example of such a generic closed surface tiled with elementary parallel transport polygons (here chosen for illustrative purposes to be triangles) will be given later in Fig. 7.13.

Later similar surfaces will arise naturally in the context of the strong coupling (small $k$) expansion for gravity, as well as in the high dimension (large $d$) expansion.

## 6.9 Lattice Regularized Path Integral

As the edge lengths $l_{ij}$ play the role of the continuum metric $g_{\mu\nu}(x)$, one would expect the discrete measure to involve an integration over the squared edge lengths (Hamber, 1984; Hartle, 1984; 1986; Hamber and Williams, 1999). Indeed the induced metric at a simplex is related to the squared edge lengths within that simplex, via the expression for the invariant line element $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$. After choosing coordinates along the edges emanating from a vertex, the relation between metric perturbations and squared edge length variations for a given simplex based at 0 in $d$

dimensions is

$$\delta g_{ij}(l^2) \;=\; \tfrac{1}{2}\left(\delta l_{0i}^2 + \delta l_{0j}^2 - \delta l_{ij}^2\right) \;. \tag{6.68}$$

For one $d$-dimensional simplex labeled by $s$ the integration over the metric is thus equivalent to an integration over the edge lengths, and one has the identity

$$\left(\frac{1}{d!}\sqrt{\det g_{ij}(s)}\right)^{\sigma}\prod_{i\geq j} dg_{ij}(s) = \left(-\tfrac{1}{2}\right)^{\frac{d(d-1)}{2}}\left[V_d(l^2)\right]^{\sigma}\prod_{k=1}^{d(d+1)/2} dl_k^2 \;. \tag{6.69}$$

There are $d(d+1)/2$ edges for each simplex, just as there are $d(d+1)/2$ independent components for the metric tensor in $d$ dimensions (Cheeger, Müller and Schrader, 1982). Here one is ignoring temporarily the triangle inequality constraints, which will further require all sub-determinants of $g_{ij}$ to be positive, including the obvious restriction $l_k^2 > 0$.

Let us discuss here briefly the simplicial inequalities which need to be imposed on the edge lengths (Wheeler, 1964). These are conditions on the edge lengths $l_{ij}$ such that the sites $i$ can be considered the vertices of a $d$-simplex embedded in flat $d$-dimensional Euclidean space. In one dimension, $d = 1$, one requires trivially for all edge lengths

$$l_{ij}^2 > 0 \;. \tag{6.70}$$

In two dimensions, $d = 2$, the conditions on the edge lengths are again $l_{ij}^2 > 0$ as in one dimensions, as well as

$$A_\Delta^2 = \left(\frac{1}{2!}\right)^2 \det g_{ij}^{(2)}(s) > 0 \;, \tag{6.71}$$

which is equivalent, by virtue of Heron's formula for the area of a triangle $A_\Delta^2 = s(s-l_{ij})(s-l_{jk})(s-l_{ki})$ where $s$ is the semi-perimeter $s = \frac{1}{2}(l_{ij}+l_{jk}+l_{ki})$, to the requirement that the area of the triangle be positive. In turn Eq. (6.71) implies that the triangle inequalities must be satisfied for all three edges,

$$\begin{aligned} l_{ij} + l_{jk} &> l_{ik} \\ l_{jk} + l_{ki} &> l_{ji} \\ l_{ki} + l_{ij} &> l_{kj} \;. \end{aligned} \tag{6.72}$$

In three dimensions, $d = 3$, the conditions on the edge lengths are again such that one recovers a physical tetrahedron. One therefore requires for the individual edge lengths the condition of Eq. (6.70), the reality and positivity of all four triangle areas as in Eq. (6.71), as well as the requirement that the volume of the tetrahedron be real and positive,

$$V_{\text{tetrahedron}}^2 = \left(\frac{1}{3!}\right)^2 \det g_{ij}^{(3)}(s) > 0 \;. \tag{6.73}$$

The generalization to higher dimensions is such that one requires all triangle inequalities and their higher dimensional analogs to be satisfied,

$$\begin{aligned} l_{ij}^2 &> 0 \\ V_k^2 &= \left(\frac{1}{k!}\right)^2 \det g_{ij}^{(k)}(s) > 0 \ , \end{aligned} \tag{6.74}$$

with $k = 2 \dots d$ for every possible choice of sub-simplex (and therefore sub-determinant) within the original simplex $s$.

The extension of the measure to many simplices glued together at their common faces is then immediate. For this purpose one first needs to identify edges $l_k(s)$ and $l_{k'}(s')$ which are shared between simplices $s$ and $s'$,

$$\int_0^\infty dl_k^2(s) \int_0^\infty dl_{k'}^2(s') \ \delta\left[l_k^2(s) - l_{k'}^2(s')\right] = \int_0^\infty dl_k^2(s) \ . \tag{6.75}$$

After summing over all simplices one derives, up to an irrelevant numerical constant, the unique functional measure for simplicial geometries

$$\int [dl^2] = \int_\varepsilon^\infty \prod_s \left[V_d(s)\right]^\sigma \prod_{ij} dl_{ij}^2 \, \Theta[l_{ij}^2] \ . \tag{6.76}$$

Here $\Theta[l_{ij}^2]$ is a (step) function of the edge lengths, with the property that it is equal to one whenever the triangle inequalities and their higher dimensional analogs are satisfied, and zero otherwise. The quantity $\varepsilon$ has been introduced as a cutoff at small edge lengths. If the measure is non-singular for small edges, one can safely take the limit $\varepsilon \to 0$. In four dimensions the lattice analog of the DeWitt measure ($\sigma = 0$) takes on a particularly simple form, namely

$$\int [dl^2] = \int_0^\infty \prod_{ij} dl_{ij}^2 \, \Theta[l_{ij}^2] \ . \tag{6.77}$$

Lattice measures over the space of squared edge lengths have been used extensively in numerical simulations of simplicial quantum gravity (Hamber and Williams, 1984; Hamber, 1984; Berg, 1985; 1986). The derivation of the above lattice measure closely parallels the analogous procedure in the continuum.

There is no obstacle in defining a discrete analog of the supermetric, as a way of introducing an invariant notion of distance between simplicial manifolds, as proposed in (Cheeger, Müller and Schrader, 1984). It leads to an alternative way of deriving the lattice measure in Eq. (6.77), by considering the discretized distance between induced metrics $g_{ij}(s)$

$$\| \delta g(s) \|^2 = \sum_s G^{ijkl}\left[g(s)\right] \ \delta g_{ij}(s) \, \delta g_{kl}(s) \ , \tag{6.78}$$

with the inverse of the lattice DeWitt supermetric now given by the expression

$$G^{ijkl}[g(s)] = \tfrac{1}{2}\sqrt{g(s)} \ \left[g^{ik}(s) g^{jl}(s) + g^{il}(s) g^{jk}(s) + \lambda \, g^{ij}(s) g^{kl}(s)\right] \ , \tag{6.79}$$

and with again $\lambda \neq -2/d$. This procedure defines a metric on the tangent space of positive real symmetric matrices $g_{ij}(s)$. After computing the determinant of $G$, the resulting functional measure is

$$\int d\mu[l^2] = \int \prod_s \left[ \det G[g(s)] \right]^{\frac{1}{2}} \prod_{i \geq j} dg_{ij}(s) \ , \tag{6.80}$$

with the determinant of the super-metric $G^{ijkl}[g(s)]$ given by the local expression

$$\det G[g(s)] \ \propto \ (1 + \tfrac{1}{2} d\lambda) \ [g(s)]^{(d-4)(d+1)/4} \ . \tag{6.81}$$

Using Eq. (6.69), and up to irrelevant constants, one obtains again the standard lattice measure of Eq. (6.76). Of course the same procedure can be followed for the Misner-like measure, leading to a similar result for the lattice measure, but with a different power $\sigma$.

One might be tempted to try to find alternative lattice measures by looking directly at the discrete form for the supermetric, written as a quadratic form in the squared edge lengths (instead of the metric components), and then evaluating the resulting determinant. The main idea, inspired by work described in a paper (Lund and Regge, 1974) on the $3+1$ formulation of simplicial gravity, is as follows. First one considers a lattice analog of the DeWitt supermetric by writing

$$\| \delta l^2 \|^2 \ = \ \sum_{ij} G^{ij}(l^2) \, \delta l_i^2 \, \delta l_j^2 \ , \tag{6.82}$$

with $G^{ij}(l^2)$ now defined on the space of squared edge lengths (Hartle, Miller and Williams, 1997). The next step is to find an appropriate form for $G_{ij}(l^2)$ expressed in terms of known geometric objects. One simple way of constructing the explicit form for $G_{ij}(l^2)$, in any dimension, is to first focus on one simplex, and write the squared volume of a given simplex in terms of the induced metric components within the *same* simplex $s$,

$$V^2(s) \ = \ \left(\tfrac{1}{d!}\right)^2 \det g_{ij}\left[l^2(s)\right] \ . \tag{6.83}$$

One computes to linear order

$$\frac{1}{V(l^2)} \sum_i \frac{\partial V^2(l^2)}{\partial l_i^2} \, \delta l_i^2 \ = \ \tfrac{1}{d!} \sqrt{\det(g_{ij})} \, g^{ij} \, \delta g_{ij} \ , \tag{6.84}$$

and to quadratic order

$$\frac{1}{V(l^2)} \sum_{ij} \frac{\partial^2 V^2(l^2)}{\partial l_i^2 \partial l_j^2} \, \delta l_i^2 \, \delta l_j^2 \ = \ \tfrac{1}{d!} \sqrt{\det(g_{ij})} \left[ g^{ij} g^{kl} \delta g_{ij} \delta g_{kl} - g^{ij} g^{kl} \delta g_{jk} \delta g_{li} \right] \ . \tag{6.85}$$

The r.h.s. of this equation contains precisely the expression appearing in the continuum supermetric of Eq. (2.14), for the specific choice of the parameter $\lambda = -2$. One is led therefore to the identification

$$G^{ij}(l^2) \;=\; -\,d!\,\sum_s \frac{1}{V(s)}\,\frac{\partial^2\,V^2(s)}{\partial l_i^2\,\partial l_j^2}\;, \tag{6.86}$$

and therefore for the norm

$$\|\delta l^2\|^2 \;=\; \sum_s V(s)\left\{-\,\frac{d!}{V^2(s)}\sum_{ij}\frac{\partial^2\,V^2(s)}{\partial l_i^2\,\partial l_j^2}\,\delta l_i^2\,\delta l_j^2\right\}\;. \tag{6.87}$$

One could be tempted at this point to write down a lattice measure, in parallel with Eq. (2.16), and write

$$\int [dl^2] \;=\; \int \prod_i \sqrt{\det G^{ij\,(\omega')}(l^2)}\,dl_i^2 \tag{6.88}$$

with

$$G^{ij\,(\omega')}(l^2) \;=\; -\,d!\,\sum_s \frac{1}{[V(s)]^{1+\omega'}}\,\frac{\partial^2\,V^2(s)}{\partial l_i^2\,\partial l_j^2}\;, \tag{6.89}$$

where one has allowed for a parameter $\omega'$, possibly different from zero, interpolating between apparently equally acceptable measures. The reasoning here is that, as in the continuum, different edge length measures, here parametrized by $\omega$', are obtained, depending on whether the local volume factor $V(s)$ is included in the supermetric or not.

One rather undesirable, and puzzling, feature of the lattice measure of Eq. (6.88) is that in general it is non-local, in spite of the fact that the original continuum measure of Eq. (2.18) is completely local (although it is clear that for some special choices of $\omega'$ and $d$, one does recover a local measure; thus in two dimensions and for $\omega' = -1$ one obtains again the simple result $\int [dl^2] = \int_0^\infty \prod_i dl_i^2$). Unfortunately irrespective of the value chosen for $\omega'$, one can show (Hamber and Williams, 1999) that the measure of Eq. (6.88) disagrees with the continuum measure of Eq. (2.18) already to lowest order in the weak field expansion, and does not therefore describe an acceptable lattice measure.

The lattice action for pure four-dimensional Euclidean gravity contains a cosmological constant and Regge scalar curvature term

$$I_{latt} \;=\; \lambda_0 \sum_h V_h(l^2) \,-\, k \sum_h \delta_h(l^2)\,A_h(l^2)\;, \tag{6.90}$$

with $k = 1/(8\pi G)$, as well as possibly higher derivative terms. It only couples edges which belong either to the same simplex or to a set of neighboring simplices, and can therefore be considered as *local*, just like the continuum action, and leads to the regularized lattice functional integral

$$Z_{latt} = \int [d\, l^2] \; e^{-\lambda_0 \sum_h V_h + k \sum_h \delta_h A_h} \quad , \tag{6.91}$$

where, as customary, the lattice ultraviolet cutoff is set equal to one (i.e. all length scales are measured in units of the lattice cutoff).

The lattice partition function $Z_{latt}$ should then be compared to the continuum Euclidean Feynman path integral of Eq. (2.34),

$$Z_{cont} = \int [d\, g_{\mu\nu}] \; e^{-\lambda_0 \int dx \sqrt{g} + \frac{1}{16\pi G} \int dx \sqrt{g} R} \quad . \tag{6.92}$$

Occasionally it can be convenient to include the $\lambda_0$-term in the measure. For this purpose one defines

$$d\mu(l^2) \equiv [d\, l^2] \, e^{-\lambda_0 \sum_h V_h} \quad . \tag{6.93}$$

It should be clear that this last expression represents a fairly non-trivial quantity, both in view of the relative complexity of the expression for the volume of a simplex, Eq. (6.5), and because of the generalized triangle inequality constraints already implicit in $[d\, l^2]$. But, like the continuum functional measure, it is certainly *local*, to the extent that each edge length appears only in the expression for the volume of those simplices which explicitly contain it. Furthermore, $\lambda_0$ sets the overall scale and can therefore be set equal to one without any loss of generality.

## 6.10 An Elementary Example

In the very simple case of one dimension ($d = 1$) one can work out explicitly a number of details, and see how potential problems with the functional measure arise, and how they are resolved.

In one dimension one discretizes the line by introducing $N$ points, with lengths $l_n$ associated with the edges, and periodic boundary conditions, $l_{N+1} = l_1$. Here $l_n$ is the distance between points $n$ and $n+1$. The only surviving invariant term in one dimension is then the overall length of a curve,

$$L(l) = \sum_{n=1}^{N} l_n \quad , \tag{6.94}$$

which corresponds to

$$\int dx \sqrt{g(x)} = \int dx \, e(x) \quad , \tag{6.95}$$

[with $g(x) \equiv g_{00}(x)$] in the continuum. Here $e(x)$ is the "einbein", and satisfies the obvious constraint $\sqrt{g(x)} = e(x) > 0$. In this context the discrete action is unique, preserving the geometric properties of the continuum definition. From the expression for the invariant line element, $ds^2 = g dx^2$, one associates $g(x)$ with $l_n^2$ (and therefore $e(x)$ with $l_n$). One can further take the view that distances can only be assigned between vertices which appear on some lattice in the ensemble, although this is not strictly necessary, as distances can also be defined for locations that do not coincide with any specific vertex.

The gravitational measure then contains an integration over the elementary lattice degrees of freedom, the lattice edge lengths. For the edges one writes the lattice integration measure as

$$\int d\mu[l] \;=\; \prod_{n=1}^{N} \int_0^\infty dl_n^2 \, l_n^\sigma \;\;, \tag{6.96}$$

where $\sigma$ is a parameter interpolating between different local measures. The positivity of the edge lengths is all that remains of the triangle inequality constraints in one dimension. The factor $l_n^\sigma$ plays a role analogous to the $g^{\sigma/2}$ which appears for continuum measures in the Euclidean functional integral.

The functional measure does not have compact support, and the cosmological term (with a coefficient $\lambda_0 > 0$) is therefore necessary to obtain convergence of the functional integral, as can be seen for example from the expression for the average edge length,

$$\langle L(l) \rangle \;=\; \langle \sum_{n=1}^{N} l_n \rangle \;=\; \mathscr{Z}_N^{-1} \prod_{n=1}^{N} \int_0^\infty dl_n^2 \, l_n^\sigma \exp\left( -\lambda_0 \sum_{n=1}^{N} l_n \right) \sum_{n=1}^{N} l_n \;=\; \frac{2+\sigma}{\lambda_0} \, N \tag{6.97}$$

with

$$\mathscr{Z}_N(\lambda_0) \;=\; \prod_{n=1}^{N} \int_0^\infty dl_n^2 \, l_n^\sigma \exp\left( -\lambda_0 \sum_{n=1}^{N} l_n \right) \;=\; \left[ \frac{2\,\Gamma(2+\sigma)}{\lambda_0^{2+\sigma}} \right]^N \;\;. \tag{6.98}$$

Similarly one finds for the fluctuation in the total length $\Delta L/L = 1/\sqrt{(2+\sigma)N}$, which requires $\sigma > -2$. Different choices for $\lambda_0$ then correspond to trivial rescalings of the average lattice spacing, $l_0 \equiv \langle l \rangle = (2+\sigma)/\lambda_0$.

In the continuum, the action of Eq. (6.95) is invariant under continuous reparametrizations

$$x \rightarrow x'(x) = x - \varepsilon(x) \tag{6.99}$$

$$g(x) \rightarrow g'(x') = \left( \frac{dx}{dx'} \right)^2 g(x) = g(x) + 2\,g(x) \left( \frac{d\varepsilon}{dx} \right) + O(\varepsilon^2) \;\;, \tag{6.100}$$

or equivalently

$$\delta g(x) \equiv g'(x') - g(x) = 2g\partial\varepsilon \;\;, \tag{6.101}$$

and we have set $\partial \equiv d/dx$. A gauge can then be chosen by imposing $g'(x') = 1$, which can be achieved by the choice of coordinates $x' = \int dx \sqrt{g(x)}$.

The discrete analog of the transformation rule is

$$\delta l_n = \varepsilon_{n+1} - \varepsilon_n \;\;, \tag{6.102}$$

where the $\varepsilon_n$'s represent continuous gauge transformations defined on the lattice vertices. In order for the edge lengths to remain positive, one needs to require $\varepsilon_n - \varepsilon_{n+1} < l_n$, which is certainly satisfied for sufficiently small $\varepsilon$'s. The above continuous symmetry is an exact invariance of the lattice action of Eq. (6.94), since

$$\delta L \;=\; \sum_{n=1}^{N} \delta l_n = \sum_{n=1}^{N} \varepsilon_{n+1} - \sum_{n=1}^{N} \varepsilon_n = 0 \;, \tag{6.103}$$

and we have used $\varepsilon_{N+1} = \varepsilon_1$. Moreover, it is the only local symmetry of the action of Eq. (6.94).

The infinitesimal local invariance property defined in Eq. (6.102) formally selects a unique measure over the edge lengths, corresponding to $\prod_n dl_n$ [$\sigma = -1$ in Eq. (6.96)], as long as we ignore the effects of the lower limit of integration. On the other hand for sufficiently large lattice diffeomorphisms, the lower limit of integration comes into play (since we require $l_n > 0$ always) and the measure is no longer invariant. Thus a measure $\int_{-\infty}^{\infty} \prod dl_n$ would not be acceptable on physical grounds; it would violate the constraint $\sqrt{g} > 0$ or $e > 0$.

The same functional measure can be obtained from the following physical consideration. Define the gauge invariant distance $d$ between two configurations of edge lengths $\{l_n\}$ and $\{l'_n\}$ by

$$d^2(l,l') \equiv \left[L(l) - L'(l')\right]^2 \;=\; \left( \sum_{n=1}^{N} l_n - \sum_{n'=1}^{N} l'_{n'} \right)^2 = \sum_{n=1}^{N} \sum_{n'=1}^{N} \delta l_n \, M_{n,n'} \, \delta l_{n'} \;, \tag{6.104}$$

with $M_{n,n'} = 1$. Since $M$ is independent of $l_n$ and $l'_{n'}$, the ensuing measure is again simply proportional to $\prod dl_n$. Note that the above metric over edge length deformations $\delta l$ is non-local.

In the continuum, the functional measure is usually determined by considering the following (local) norm in function space,

$$||\delta g||^2 \;=\; \int dx \, \sqrt{g(x)} \; \delta g(x) \; G(x) \; \delta g(x) \;, \tag{6.105}$$

and diffeomorphism invariance would seem to require $G(x) = 1/g^2(x)$. The volume element in function space is then an ultraviolet regulated version of $\prod_x \sqrt{G(x)}\, dg(x) = \prod_x dg(x)/g(x)$, which is the Misner measure in one dimension. Its naive discrete counterpart would be $\prod dl_n/l_n$, which is not invariant under the transformation of Eq. (6.102) (it is invariant under $\delta l_n = l_n(\varepsilon_{n+1} - \varepsilon_n)$, which is not an invariance of the action).

The point of the discussion of the one-dimensional case is to bring to the surface the several non-trivial issues that arise when defining a properly regulated version of the continuum Feynman functional measure $[dg_{\mu\nu}]$, and how they can be systematically resolved.

## 6.11 Lattice Higher Derivative Terms

So far only the gravitational Einstein-Hilbert contribution to the lattice action and the cosmological constant term have been considered. There are several motivations for extending the discussion to lattice higher derivative terms, which would include the fact that these terms a) might appear in the original microscopic action, or might have to be included to cure the classical unboundedness problem of the Euclidean Einstein-Hilbert action, b) that they are in any case generated by radiative corrections, and c) that on a more formal level they may shed new light on the relationship between the lattice and continuum expressions for curvature terms as well as quantities such as the Riemann tensor on a hinge, Eq. (6.36).

For these reasons we will discuss here a generalization of the Regge gravity equivalent of the Einstein action to curvature squared terms. When considering contributions quadratic in the curvature there are overall six possibilities, listed in Eq. (1.127). Among the two topological invariants, the Euler characteristic $\chi$ for a simplicial decomposition may be obtained from a particular case of the general formula for the analogue of the Lipschitz-Killing curvatures of smooth Riemannian manifolds for piecewise flat spaces. The formula of (Cheeger, Müller and Schrader, 1984) reduces in four dimensions to

$$\chi = \sum_{\sigma^0} \Big[ 1 - \sum_{\sigma^2 \supset \sigma^0} (0,2) - \sum_{\sigma^4 \supset \sigma^0} (0,4) + \sum_{\sigma^4 \supset \sigma^2 \supset \sigma^0} (0,2)(2,4) \Big] \ , \tag{6.106}$$

where $\sigma^i$ denotes an i-dimensional simplex and $(i,j)$ denotes the (internal) dihedral angle at an i-dimensional face of a j-dimensional simplex. Thus, for example, $(0,2)$ is the angle at the vertex of a triangle and $(2,4)$ is the dihedral angle at a triangle in a 4-simplex (The normalization of the angles is such that the volume of a sphere in any dimension is one; thus planar angles are divided by $2\pi$, 3-dimensional solid angles by $4\pi$ and so on).

Of course, as noted before, there is a much simpler formula for the Euler characteristic of a simplicial complex

$$\chi = \sum_{i=0}^{d} (-1)^i N_i \ , \tag{6.107}$$

where $N_i$ is the number of simplices of dimension $i$. However, it may turn out to be useful in quantum gravity calculations to have a formula for $\chi$ in terms of the angles, and hence of the edge lengths, of the simplicial decomposition. These expression are interesting and useful, but do not shed much light on how the other curvature squared terms should be constructed.

In a piecewise linear space curvature is detected by going around elementary loops which are dual to a $(d-2)$-dimensional subspace. The area of the loop itself is not well defined, since any loop inside the $d$-dimensional simplices bordering the hinge will give the same result for the deficit angle. On the other hand the hinge has a content (the length of the edge in $d=3$ and the area of the triangle in $d=4$), and there is a natural volume associated with each hinge, defined by dividing the volume of each simplex touching the hinge into a contribution belonging to that hinge, and other contributions belonging to the other hinges on that simplex (Hamber and Williams, 1984). The contribution belonging to that simplex will be called *dihedral volume* $V_d$ (for an illustration, see Fig. 6.13). The volume $V_h$ associated with the hinge $h$ is then naturally the sum of the dihedral volumes $V_d$ belonging to each simplex

$$V_h = \sum_{\substack{\mathbf{d-simplices} \\ \mathbf{meeting\ on\ h}}} V_d \ . \tag{6.108}$$

The dihedral volume associated with each hinge in a simplex can be defined using dual volumes, a barycentric subdivision, or some other natural way of dividing the volume of a $d$-simplex into $d(d+1)/2$ parts. If the theory has some reasonable continuum limit, then the final results should not depend on the detailed choice of volume type.

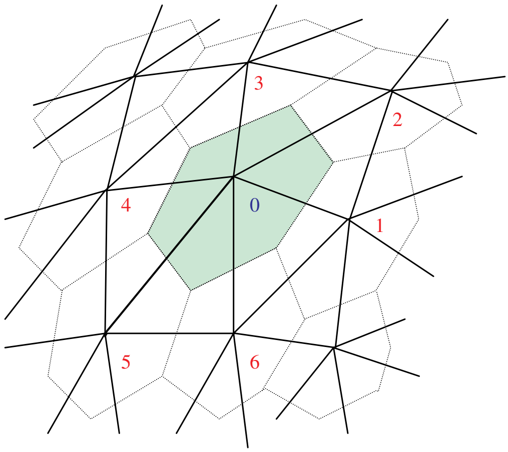



**Fig. 6.13** Illustration of dual volumes in two dimensions. The vertices of the polygons reside in the dual lattice. The shaded region describes the dual area associated with the vertex 0.

As mentioned previously, there is a well-established procedure for constructing a dual lattice for any given lattice. This involves constructing polyhedral cells, known in the literature as Voronoi polyhedra, around each vertex, in such a way that the cell around each particular vertex contains all points which are nearer to that vertex than to any other vertex. Thus the cell is made up from $(d-1)$-dimensional subspaces which are the perpendicular bisectors of the edges in the original lattice, $(d-2)$-dimensional subspaces which are orthogonal to the 2-dimensional subspaces of the original lattice, and so on. General formulas for dual volumes are given in (Hamber and Williams, 1986). In the case of the barycentric subdivision, the dihedral volume is just $2/d(d+1)$ times the volume of the simplex. This leads one to conclude that there is a natural area $A_{C_h}$ associated with each hinge

$$A_{C_h} = \frac{V_h}{A_h^{(d-2)}} , \tag{6.109}$$

obtained by dividing the volume per hinge (which is $d$-dimensional) by the volume of the hinge (which is $(d-2)$-dimensional).

The next step is to find terms equivalent to the continuum expression of Eq. (1.127), and the remainder of this section will be devoted to this problem. It may be objected that since in Regge calculus where the curvature is restricted to the hinges which are subspaces of dimension 2 less than that of the space considered, then the curvature tensor involves $\delta$-functions with support on the hinges, and so higher powers of the curvature tensor are not defined. But this argument clearly does not apply to the Euler characteristic and the Hirzebruch signature of Eq. (1.127), which are both integrals of 4-forms. However it is a common procedure in lattice field theory to take powers of fields defined at the same point, and there is no reason why one should not consider similar terms in lattice gravity. Of course one would like the expressions to correspond to the continuum ones as the edge lengths of the simplicial lattice become smaller and smaller.

Since the curvature is restricted to the hinges, it is natural that expressions for curvature integrals should involve sums over hinges as in Eq. (6.38). The curvature tensor, which involves second derivatives of the metric, is of dimension $L^{-2}$. Therefore $\frac{1}{4}\int d^dx\sqrt{g}R^n$ is of dimension $L^{d-2n}$. Thus if one postulates that an $R^2$-type term will involve the square of $A_h\delta_h$, which is of dimension $L^{2(d-2)}$, then one will need to divide by some $d$-dimensional volume to obtain the correct dimension for the extra term in the action. Now any hinge is surrounded by a number of $d$-dimensional simplices, so the procedure of dividing by a $d$-dimensional volume seems ambiguous. The crucial step is to realize that there *is* a unique $d$-dimensional volume associated with each hinge, as described above.

If one regards the invariant volume element $\sqrt{g}d^dx$ as being represented by $V_h$ of Eq. (6.108) when one performs the sum over hinges as in Eq. (6.38), then this means that one may regard the scalar curvature $R$ contribution as being represented at each hinge by $2A_h\delta_h/V_h$

$$\frac{1}{2}\int d^dx\sqrt{g}\,R \;\to\; \sum_{\text{hinges h}} V_h\,\frac{A_h\delta_h}{V_h} \equiv \sum_{\text{hinges h}} A_h\delta_h \; . \tag{6.110}$$

It is then straightforward to see that a candidate curvature squared term is

$$\sum_{\text{hinges h}} V_h\left(\frac{A_h\delta_h}{V_h}\right)^2 \equiv \sum_{\text{hinges h}} V_h\left(\frac{\delta_h}{A_{C_h}}\right)^2 , \tag{6.111}$$

where $A_{C_h}$ was defined in Eq. (6.109). Since the expression in Eq. (6.111) vanishes if and only if all deficit angles are zero, it is naturally identified with the continuum Riemann squared term,

$$\tfrac{1}{4}\int d^d x \sqrt{g}\, R_{\mu\nu\lambda\sigma} R^{\mu\nu\lambda\sigma} \;\;\rightarrow\;\; \sum_{\text{hinges h}} V_h \Big( \frac{\delta_h}{A_{C_h}} \Big)^2 \;. \tag{6.112}$$

The above construction then leaves open the question of how to construct the remaining curvature squared terms in four dimensions. If one takes the form given previously in Eq. (6.36) for the Riemann tensor on a hinge and contracts one obtains

$$R(h) = 2\,\frac{\delta_h}{A_{C_h}} \;\;, \tag{6.113}$$

which agrees with the form used in the Regge action for $R$. But one also finds readily that with this choice the higher derivative terms are all proportional to each other (Hamber and Williams, 1986),

$$\tfrac{1}{4} R_{\mu\nu\rho\sigma}(h) R^{\mu\nu\rho\sigma}(h) = \tfrac{1}{2} R_{\mu\nu}(h) R^{\mu\nu}(h) = \tfrac{1}{4} R(h)^2 = \Big( \frac{\delta_h}{A_{C_h}} \Big)^2 \;. \tag{6.114}$$

Furthermore if one uses the above expression for the Riemann tensor to evaluate the contribution to the Euler characteristic on each hinge one obtains zero, and becomes clear that at least in this case one needs cross terms involving contributions from different hinges.

The next step is therefore to embark on a slightly more sophisticated approach, and construct the full Riemann tensor by considering more than one hinge. Define the Riemann tensor for a simplex $s$ as a weighted sum of hinge contributions

$$\Big[ R_{\mu\nu\rho\sigma} \Big]_s = \sum_{h \subset s} \omega_h \Big[ \frac{\delta}{A_C} U_{\mu\nu} U_{\rho\sigma} \Big]_h \;\;, \tag{6.115}$$

where the $\omega_h$ are dimensionless weights, to be determined later. After squaring one obtains

$$\Big[ R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} \Big]_s = \sum_{h,h' \subset s} \omega_h \omega_{h'} \Big[ \frac{\delta}{A_C} U_{\mu\nu} U_{\rho\sigma} \Big]_h \Big[ \frac{\delta}{A_C} U^{\mu\nu} U^{\rho\sigma} \Big]_{h'} \;. \tag{6.116}$$

The question of the weights $\omega_h$ introduced in Eq. (6.115) will now be addressed. Consider the expression for the scalar curvature of a simplex defined as

$$\Big[ R \Big]_s = \sum_{h \subset s} \omega_h \Big[ 2\,\frac{\delta}{A_C} \Big]_h \;. \tag{6.117}$$

It is clear from the formulae given above for the lattice curvature invariants (constructed in a simplex by summing over hinge contributions) that there is again a natural volume associated with them : the sum of the volumes of the hinges in the simplex

$$V_s = \sum_{h \subset s} V_h \;\;, \tag{6.118}$$

where $V_h$ is the volume around the hinge, as defined in Eq. (6.108). Summing the scalar curvature over all simplices, one should recover Regge's expression

$$\sum_s V_s \left[ R \right]_s = \sum_s \sum_{h \subset s} \omega_h \left[ 2 \, \frac{\delta}{A_C} \right]_h = \sum_h \delta_h A_h \ , \tag{6.119}$$

which implies

$$N_{2,4} V_s \, \omega_h \frac{\delta_h}{A_{C_h}} \equiv N_{2,4} V_s \, \omega_h \frac{\delta_h A_h}{V_h} = \delta_h A_h \ , \tag{6.120}$$

where $N_{2,4}$ is the number of simplices meeting on that hinge. Therefore the correct choice for the weights is

$$\omega_h = \frac{V_h}{N_{2,4} \, V_s} = \frac{V_h}{N_{2,4} \sum_{h \subset s} V_h} \ . \tag{6.121}$$

Thus the weighting factors that reproduce Regge's formula for the Einstein action are just the volume fractions occupied by the various hinges in a simplex, which is not surprising (of course the above formulae are not quite unique, since one might have done the above construction of higher derivative terms by considering a point $p$ instead of a four-simplex $s$).

In particular the following form for the Weyl tensor squared was given in (Hamber and Williams, 1986)

$$\begin{aligned} & \int d^d x \sqrt{g} \, C_{\mu\nu\lambda\sigma} C^{\mu\nu\lambda\sigma} \sim \\ & \tfrac{2}{3} \sum_s V_s \sum_{h,h' \subset s} \varepsilon_{h,h'} \left( \omega_h \left[ \frac{\delta}{A_C} \right]_h - \omega_{h'} \left[ \frac{\delta}{A_C} \right]_{h'} \right)^2 \ , \end{aligned} \tag{6.122}$$

which introduces a short range coupling between deficit angles. The numerical factor $\varepsilon_{h,h'}$ is equal to 1 if the two hinges $h, h'$ have one edge in common and $-2$ if they do not. Note that this particular interaction term has the property that it requires neighboring deficit angles to have similar values, but it does not require them to be small, which is a key property one would expect from the Weyl curvature squared term.

In conclusion the formulas given above allow one to construct the remaining curvature squared terms in four dimensions, and in particular to write, for example, the lattice analog of the continuum curvature squared action of Eq. (1.137)

$$\begin{aligned} I = \int d^4 x \sqrt{g} \, \Big[ \ & \lambda_0 \ - k \, R - b \, R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} \\ & + \tfrac{1}{2} \ (a + 4b) \, C_{\mu\nu\rho\sigma} C^{\mu\nu\rho\sigma} \Big] \ . \end{aligned} \tag{6.123}$$

To compare with the form of Eq. (1.137) use has been made of $R^2 = 3R^2_{\mu\nu\rho\sigma} - 6C^2$ and $R^2_{\mu\nu} = R^2_{\mu\nu\rho\sigma} - \frac{3}{2} C^2$, up to additive constant contributions. A special case corresponds to $b = -a/4$, which gives a pure $R^2_{\mu\nu\rho\sigma}$ contribution. The latter vanishes if and only if the curvature is locally zero, which is not true of any of the other curvature squared terms.

## 6.12 Scalar Matter Fields

In the previous section we have discussed the construction and the invariance properties of a lattice action for pure gravity. Next a scalar field can be introduced as the simplest type of dynamical matter that can be coupled invariantly to gravity. In the continuum the scalar action for a single component field $\phi(x)$ is usually written as

$$I[g,\phi] = \tfrac{1}{2} \int dx \sqrt{g} \, [ g^{\mu\nu} \, \partial_\mu \phi \, \partial_\nu \phi + (m^2 + \xi R) \phi^2 ] + \ldots \tag{6.124}$$

where the dots denote scalar self-interaction terms. Thus for example a scalar field potential $U(\phi)$ could be added containing quartic field terms, whose effects could be of interest in the context of cosmological models where spontaneously broken symmetries play an important role. The dimensionless coupling $\xi$ is arbitrary; two special cases are the minimal ($\xi = 0$) and the conformal ($\xi = \frac{1}{6}$) coupling case. In the following we shall mostly consider the case $\xi = 0$. Also, it will be straightforward to extend later the treatment to the case of an $N_s$-component scalar field $\phi^a$ with $a = 1, ..., N_s$.

One way to proceed is to introduce a lattice scalar $\phi_i$ defined at the vertices of the simplices. The corresponding lattice action can then be obtained through a procedure by which the original continuum metric is replaced by the induced lattice metric, with the latter written in terms of squared edge lengths as in Eq. (6.3). For illustrative purposes only the two-dimensional case will be worked out explicitly here (Christ, Friedberg and Lee, 1982; Itzykson, 1983; Itzykson and Bander, 1983; Jevicki and Ninomiya, 1985). The generalization to higher dimensions is straightforward, and in the end the final answer for the lattice scalar action is almost identical to the two dimensional form. Furthermore in two dimensions it leads to a natural dicretization of the bosonic string action (Polyakov, 1989).

In two dimensions the simplicial lattice is built out of triangles. For a given triangle it will be convenient to use the notation of Fig. 6.14, which will display more readily the symmetries of the resulting scalar lattice action. Here coordinates will be picked in each triangle along the (1,2) and (1,3) directions.

To construct a lattice action for the scalar field, one performs in two dimensions the replacement

$$\begin{aligned}
g_{\mu\nu}(x) &\longrightarrow g_{ij}(\Delta) \\
\det g_{\mu\nu}(x) &\longrightarrow \det g_{ij}(\Delta) \\
g^{\mu\nu}(x) &\longrightarrow g^{ij}(\Delta) \\
\partial_\mu \phi \, \partial_\nu \phi &\longrightarrow \Delta_i \phi \, \Delta_j \phi \quad ,
\end{aligned} \tag{6.125}$$

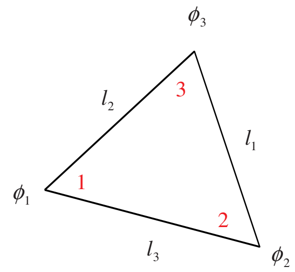



**Fig. 6.14** Labeling of edges and fields for the construction of the scalar field action.

with the following definitions

$$g_{ij}(\Delta) = \begin{pmatrix} l_3^2 & \frac{1}{2}(-l_1^2+l_2^2+l_3^2) \\ \frac{1}{2}(-l_1^2+l_2^2+l_3^2) & l_2^2 \end{pmatrix} \tag{6.126}$$

$$\det g_{ij}(\Delta) = \tfrac{1}{4}\left[2(l_1^2 l_2^2 + l_2^2 l_3^2 + l_3^2 l_1^2) - l_1^4 - l_2^4 - l_3^4\right] \equiv 4A_\Delta^2 \tag{6.127}$$

$$g^{ij}(\Delta) = \frac{1}{\det g(\Delta)} \begin{pmatrix} l_2^2 & \frac{1}{2}(l_1^2-l_2^2-l_3^2) \\ \frac{1}{2}(l_1^2-l_2^2-l_3^2) & l_3^2 \end{pmatrix} . \tag{6.128}$$

The scalar field derivatives get replaced as usual by finite differences

$$\partial_\mu \phi \longrightarrow (\Delta_\mu \phi)_i = \phi_{i+\mu} - \phi_i , \tag{6.129}$$

where the index $\mu$ labels the possible directions in which one can move away from a vertex within a given triangle. Then

$$\Delta_i \phi \, \Delta_j \phi = \begin{pmatrix} (\phi_2-\phi_1)^2 & (\phi_2-\phi_1)(\phi_3-\phi_1) \\ (\phi_2-\phi_1)(\phi_3-\phi_1) & (\phi_3-\phi_1)^2 \end{pmatrix} . \tag{6.130}$$

Then the discrete scalar field action takes the form

$$I = \tfrac{1}{16} \sum_\Delta \frac{1}{A_\Delta} \left[ l_1^2(\phi_2-\phi_1)(\phi_3-\phi_1) + l_2^2(\phi_3-\phi_2)(\phi_1-\phi_2) + l_3^2(\phi_1-\phi_3)(\phi_2-\phi_3) \right] , \tag{6.131}$$

where the sum is over all triangles on the lattice. Using the identity

$$(\phi_i-\phi_j)(\phi_i-\phi_k) = \tfrac{1}{2}\left[(\phi_i-\phi_j)^2 + (\phi_i-\phi_k)^2 - (\phi_j-\phi_k)^2\right] , \tag{6.132}$$

one obtains after some re-arrangements the slightly more appealing expression for the action of a massless scalar field (Itzykson and Bander, 1983)

$$I(l^2,\phi) = \tfrac{1}{2} \sum_{<ij>} A_{ij} \left( \frac{\phi_i - \phi_j}{l_{ij}} \right)^2 . \tag{6.133}$$

$A_{ij}$ is the dual (Voronoi) area associated with the edge $ij$, and the symbol $<ij>$ denotes a sum over nearest neighbor lattice vertices. It is immediate to generalize the action of Eq. (6.133) to higher dimensions, with the two-dimensional Voronoi volumes replaced by their higher dimensional analogs, leading to

$$I(l^2,\phi) = \frac{1}{2} \sum_{<ij>} V_{ij}^{(d)} \left( \frac{\phi_i - \phi_j}{l_{ij}} \right)^2 . \tag{6.134}$$

Here $V_{ij}^{(d)}$ is the dual (Voronoi) volume associated with the edge $ij$, and the sum is over all links on the lattice.

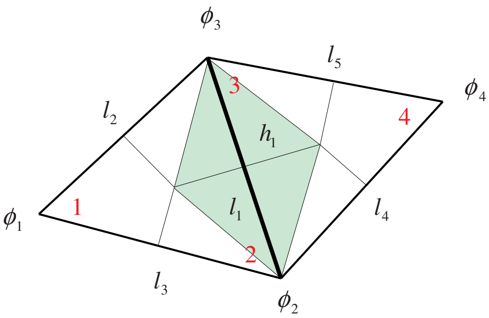



**Fig. 6.15** Dual area associated with the edge $l_1$ (*shaded area*), and the corresponding dual link $h_1$.

In two dimensions, in terms of the edge length $l_{ij}$ and the dual edge length $h_{ij}$, connecting neighboring vertices in the dual lattice, one has $A_{ij} = \frac{1}{2} h_{ij} l_{ij}$ (see Fig. 6.15). Other choices for the lattice subdivision will lead to a similar formula for the lattice action, with the Voronoi dual volumes replaced by their appropriate counterparts for the new lattice. Explicitly, for an edge of length $l_1$ the dihedral dual volume contribution is given by

$$A_{l_1} = \frac{l_1^2(l_2^2 + l_3^2 - l_1^2)}{16 A_{123}} + \frac{l_1^2(l_4^2 + l_5^2 - l_1^2)}{16 A_{234}} = \frac{1}{2}\, l_1 h_1 , \tag{6.135}$$

with $h_1$ is the length of the edge dual to $l_1$.

On the other hand the barycentric dihedral area for the same edge would be simply (see Fig 6.16)

$$A_{l_1} = (A_{123} + A_{234})/3 . \tag{6.136}$$

It is well known that one of the disadvantages of the Voronoi construction is the lack of positivity of the dual volumes, as pointed out in (Hamber and Williams, 1984). Thus some of the weights appearing in Eq. (6.133) can be negative for such an action. For the barycentric subdivision this problem does not arise, as the areas $A_{ij}$

are always positive due to the enforcement of the triangle inequalities. Thus from a practical point of view the barycentric volume subdivision is the simplest to deal with.

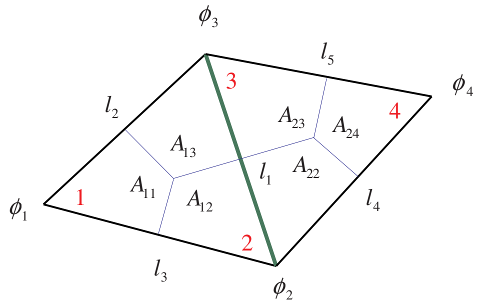



**Fig. 6.16** More dual areas appearing in the scalar field action.

The scalar action of Eq. (6.134) has a very natural form: it involves the squared difference of fields at neighboring points divided by their invariant distances $(\phi_i - \phi_j)/l_{ij}$, weighted by the appropriate space-time volume element $V_{ij}^{(d)}$ associated with the lattice link $ij$. This suggests that one could just as well define the scalar fields on the vertices of the dual lattice, and write

$$I(l^2,\phi) \;=\; \tfrac{1}{2} \sum_{<rs>} V_{rs}^{(d)} \left( \frac{\phi_r - \phi_s}{l_{rs}} \right)^2 \;, \tag{6.137}$$

with $l_{rs}$ the length of the edge connecting the dual lattice vertices $r$ and $s$, and consequently $V_{rs}^{(d)}$ the spacetime volume fraction associated with the dual lattice edge $rs$. One would expect both forms to be equivalent in the continuum limit.

Continuing on with the two-dimensional case, mass and curvature terms such as the ones appearing in Eq. (6.124) can be added to the action, so that the total scalar lattice action contribution becomes

$$I \;=\; \tfrac{1}{2} \sum_{<ij>} A_{ij} \left( \frac{\phi_i - \phi_j}{l_{ij}} \right)^2 + \tfrac{1}{2} \sum_i A_i \,(m^2 + \xi R_i)\, \phi_i^2 \;\; . \tag{6.138}$$

The term containing the discrete analog of the scalar curvature involves the quantity

$$A_i R_i \equiv \sum_{h \supset i} \delta_h \;\sim\; \sqrt{g}\, R \;\; . \tag{6.139}$$

In the above expression for the scalar action, $A_{ij}$ is the area associated with the edge $l_{ij}$, while $A_i$ is associated with the site $i$. Again there is more than one way to define the volume element $A_i$, (Hamber and Williams, 1986), but under reasonable assumptions, such as positivity, one expects to get equivalent results in the lattice continuum limit, if it exists.

In higher dimensions one would use

$$I = \tfrac{1}{2} \sum_{<ij>} V_{ij}^{(d)} \left( \frac{\phi_i - \phi_j}{l_{ij}} \right)^2 + \tfrac{1}{2} \sum_i V_i^{(d)} \, (m^2 + \xi R_i) \phi_i^2 \ , \tag{6.140}$$

where the term containing the discrete analog of the scalar curvature involves

$$V_i^{(d)} R_i \equiv \sum_{h \supset i} \delta_h V_h^{(d-2)} \sim \sqrt{g} R \ . \tag{6.141}$$

In the expression for the scalar action, $V_{ij}^{(d)}$ is the (dual) volume associated with the edge $l_{ij}$, while $V_i^{(d)}$ is the (dual) volume associated with the site $i$.

The lattice scalar action contains a mass parameter $m$, which has to be tuned to a small value in lattice units to achieve the lattice continuum limit for scalar correlations. The agreement between different lattice actions in the smooth limit can be shown explicitly in the lattice weak field expansion. But in general, as is already the case for the purely gravitational action, the correspondence between lattice and continuum operators is true classically only up to higher derivative corrections. But such higher derivative corrections in the scalar field action are expected to be irrelevant when looking at the large distance limit, and they will not be considered here any further.

As an extreme case one could even consider a situation in which the matter action by itself is the only action contribution, without any additional term for the gravitational field, but still with a non-trivial gravitational measure; integration over the scalar field would then give rise to an effective non-local gravitational action.

Finally let us take notice here of the fact that if an $N_s$-component scalar field is coupled to gravity, the power $\sigma$ appearing in the gravitational functional measure has to be modified to include an additional factor of $\prod_x (\sqrt{g})^{N_s/2}$. The additional measure factor insures that the integral

$$\int \prod_x \left[ d\phi \, (\sqrt{g})^{\frac{N_s}{2}} \right] \exp \left( -\tfrac{1}{2} m^2 \int \sqrt{g} \, \phi^2 \right) = \left( \frac{2\pi}{m^2} \right)^{\frac{N_s V}{2}} \ , \tag{6.142}$$

evaluates to a constant. Thus for large mass $m$ the scalar field completely decouples, leaving only the dynamics of the pure gravitational field.

The quadratic scalar field action of Eq. (6.134) can be written in terms of a matrix $\Delta_{ij}(l^2)$

$$I(l^2, \phi) \ = \ -\tfrac{1}{2} \sum_{<ij>} \phi_i \, \Delta_{ij}(l^2) \, \phi_j \ . \tag{6.143}$$

The matrix $\Delta_{ij}(l^2)$ can then be regarded as a lattice version of the continuum scalar Laplacian,

$$\Delta(g) = \frac{1}{\sqrt{g}} \partial_\mu \sqrt{g} g^{\mu\nu} \partial_\nu \quad , \tag{6.144}$$

for a given background metric. This then allows one to define the massless lattice scalar propagator as the inverse of the above matrix, $G_{ij}(l^2) = \Delta_{ij}^{-1}(l^2)$. The continuum scalar propagator for a finite scalar mass $m$ and in a given background geometry, evaluated for large separations $d(x,y) \gg m^{-1}$,

$$\begin{aligned} G(x,y|g) &= <x| \frac{1}{-\Delta(g) + m^2} |y> \\ &\underset{d(x,y)\to\infty}{\sim} d^{-(d-1)/2}(x,y) \exp\{-m\, d(x,y)\} \quad , \end{aligned} \tag{6.145}$$

involves the geodesic distance $d(x,y)$ between points $x$ and $y$,

$$d(x,y|g) = \int_{\tau(x)}^{\tau(y)} d\tau \sqrt{g_{\mu\nu}(\tau) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}} \quad . \tag{6.146}$$

Analogously, one can define the discrete massive lattice scalar propagator

$$\begin{aligned} G_{ij}(l^2) &= \left[ \frac{1}{-\Delta(l^2) + m^2} \right]_{ij} \\ &\underset{d(i,j)\to\infty}{\sim} d^{-(d-1)/2}(i,j) \exp\{-m\, d(i,j)\} \quad , \end{aligned} \tag{6.147}$$

where $d(i,j)$ is the lattice geodesic distance between vertex $i$ and vertex $j$. The inverse can be computed, for example, via the recursive expansion (valid for $m^2 > 0$ to avoid the zero eigenvalue of the Laplacian)

$$\frac{1}{-\Delta(l^2) + m^2} = \frac{1}{m^2} \sum_{n=0}^{\infty} \left( \frac{1}{m^2} \Delta(l^2) \right)^n \quad . \tag{6.148}$$

The large distance behavior of the Euclidean (flat space) massive free field propagator in $d$ dimensions is known in the statistical mechanics literature as the Ornstein-Zernike result.

As a consequence, the lattice propagator $G_{ij}(l^2)$ can be used to estimate the lattice geodesic distance $d(i,j|l^2)$ between any two lattice points $i$ and $j$ in a fixed background lattice geometry (provided again that their mutual separation is such that $d(i,j) \gg m^{-1}$).

$$d(i,j) \underset{d(i,j)\to\infty}{\sim} -\frac{1}{m} \ln G_{ij}(l^2) \quad . \tag{6.149}$$

## 6.13 Invariance Properties of the Scalar Action

In the very simple case of one dimension ($d = 1$) one can work out the details to any degree of accuracy, and see how potential problems arise and how they are resolved. Introduce a scalar field $\phi_n$ defined on the sites, with action

$$I(\phi) \;=\; \tfrac{1}{2} \sum_{n=1}^{N} V_1(l_n) \left( \frac{\phi_{n+1} - \phi_n}{l_n} \right)^2 + \tfrac{1}{2}\, \omega \sum_{n=1}^{N} V_0(l_n)\, \phi_n^2 \ , \tag{6.150}$$

with $\phi(N+1) = \phi(1)$. It is natural in one dimension to take for the "volume per edge" $V_1(l_n) = l_n$, and for the "volume per site" $V_0(l_n) = (l_n + l_{n-1})/2$. Here $\omega$ plays the role of a mass for the scalar field, $\omega = m^2$. In addition one needs a term

$$\lambda\, L(l) \;=\; \lambda_0 \sum_{n=1}^{N} l_n \ , \tag{6.151}$$

which is necessary in order to make the $dl_n$ integration convergent at large $l$. Varying the action with respect to $\phi_n$ gives

$$\frac{2}{l_{n-1} + l_n} \left[ \frac{\phi_{n+1} - \phi_n}{l_n} - \frac{\phi_n - \phi_{n-1}}{l_{n-1}} \right] \;=\; \omega\, \phi_n \ . \tag{6.152}$$

This is the discrete analog of the equation $g^{-1/2} \partial g^{-1/2} \partial \phi = \omega \phi$. The spectrum of the Laplacian of Eq. (6.152) corresponds to $\Omega \equiv -\omega > 0$. Variation with respect to $l_n$ gives instead

$$\frac{1}{2 l_n^2}\, (\phi_{n+1} - \phi_n)^2 \;=\; \lambda_0 + \frac{1}{4}\, \omega\, (\phi_n^2 + \phi_{n+1}^2) \ . \tag{6.153}$$

For $\omega = 0$ it suggests the well-known interpretation of the fields $\phi_n$ as coordinates in embedding space. In the following we shall only consider the case $\omega = 0$, corresponding to a massless scalar field.

It is instructive to look at the invariance properties of the scalar field action under the continuous lattice diffeomorphisms defined in Eq. (6.102). Physically, these local gauge transformations, which act on the vertices, correspond to re-assignments of edge lengths which leave the distance between two fixed points unchanged. In the simplest case, only two neighboring edge lengths are changed, leaving the total distance between the end points unchanged. On physical grounds one would like to maintain such an invariance also in the case of coupling to matter, just as is done in the continuum.

The scalar nature of the field requires that in the continuum under a change of coordinates $x \to x'$,

$$\phi'(x') = \phi(x) \ , \tag{6.154}$$

where $x$ and $x'$ refer to the same physical point in the two coordinate systems. On the lattice, as discussed previously, diffeomorphisms move the points around, and at the same vertex labeled by $n$ we expect

$$\phi_n \;\to\; \phi_n' \approx \phi_n + \left( \frac{\phi_{n+1} - \phi_n}{l_n} \right) \varepsilon_n \;\; . \tag{6.155}$$

One can determine the exact form of the change needed in $\phi_n$ by requiring that the local variation of the scalar field action be zero. Solving the resulting quadratic equation for $\Delta\phi_n$ one obtains a rather unwieldy expression, given to lowest order by

$$\begin{aligned} \Delta\phi_n = & \frac{\varepsilon_n}{2} \left[ \frac{\phi_n - \phi_{n-1}}{l_{n-1}} + \frac{\phi_{n+1} - \phi_n}{l_n} \right] \\ & + \frac{\varepsilon_n^2}{8} \left[ - \frac{\phi_n - \phi_{n-1}}{l_{n-1}^2} + \frac{\phi_{n+1} - \phi_n}{l_n^2} + \frac{\phi_{n+1} - 2\phi_n + \phi_{n-1}}{l_{n-1} l_n} \right] \\ & + O(\varepsilon_n^3) \;\; , \end{aligned} \tag{6.156}$$

and which is indeed of the expected form (as well as symmetric in the vertices $n-1$ and $n+1$). For fields which are reasonably smooth, this correction is suppressed if $|\phi_{n+1} - \phi_n|/l_n \ll 1$. On the other hand it should be clear that the measure $d\phi_n$ is no longer manifestly invariant, due to the rather involved transformation property of the scalar field.

The full functional integral for $N$ sites then reads

$$\mathscr{Z}_N \;=\; \prod_{n=1}^{N} \int_0^\infty dl_n^2 \, l_n^\sigma \int_{-\infty}^{\infty} d\phi_n \exp \left\{ -\lambda_0 \sum_{n=1}^{N} l_n - \tfrac{1}{2} \sum_{n=1}^{N} \frac{1}{l_n} \, (\phi_{n+1} - \phi_n)^2 \right\} \;\; . \tag{6.157}$$

In the absence of the scalar field one just has the $\mathscr{Z}_N(\lambda_0)$ of Eq. (6.98). The trivial translational mode in $\phi$ can be eliminated for example by setting $\sum_{n=1}^{N} \phi_n = 0$.

It is possible to further constrain the measure over the edge lengths by examining some local averages. Under a rescaling of the edge lengths $l_n \to \alpha\, l_n$ one can derive the following identity for $\mathscr{Z}_N$

$$\mathscr{Z}_N(\lambda_0, z) \;=\; \lambda_0^{-(5/2+\sigma)N} z^{-N/2} \mathscr{Z}_N(1,1) \;\; , \tag{6.158}$$

where we have replaced the coefficient $1/2$ of the scalar kinetic term by $z/2$. It follows then that

$$\langle l \rangle \;\equiv \frac{1}{N} \langle \sum_{n=1}^{N} l_n \rangle \;=\; (\frac{5}{2} + \sigma)\, \lambda_0^{-1} \tag{6.159}$$

and

$$\frac{1}{N} \langle \sum_{n=1}^{N} \frac{1}{l_n} \, (\phi_{n+1} - \phi_n)^2 \rangle \;=\; 1 \;\; . \tag{6.160}$$

Without loss of generality we can fix the average edge length to be equal to one, $\langle l \rangle = 1$, which then requires $\lambda_0 = \frac{5}{2} + \sigma$. In order for the model to be meaningful, the measure parameter is constrained by $\sigma > -5/2$, i.e. the measure over the edges cannot be too singular.

## 6.14 Lattice Fermions, Tetrads and Spin Rotations

On a simplicial manifold spinor fields $\psi_s$ and $\bar{\psi}_s$ are most naturally placed at the center of each d-simplex $s$. In the following we will restrict our discussion for simplicity to the four-dimensional case, and largely follow the original discussion given in (Drummond, 1986). As in the continuum (see for example Veltman, 1975), the construction of a suitable lattice action requires the introduction of the Lorentz group and its associated tetrad fields $e_\mu^a(s)$ within each simplex labeled by $s$.

Within each simplex one can choose a representation of the Dirac gamma matrices, denoted here by $\gamma^\mu(s)$, such that in the local coordinate basis

$$\{\gamma^\mu(s), \gamma^\nu(s)\} = 2 g^{\mu\nu}(s) \ . \tag{6.161}$$

These in turn are related to the ordinary Dirac gamma matrices $\gamma^a$, which obey

$$\left\{\gamma^a, \gamma^b\right\} = 2 \eta^{ab} \tag{6.162}$$

by

$$\gamma^\mu(s) = e_a^\mu(s) \gamma^a \ , \tag{6.163}$$

so that within each simplex the tetrads $e_\mu^a(s)$ satisfy the usual relation

$$e_a^\mu(s)\, e_b^\nu(s)\, \eta^{ab} = g^{\mu\nu}(s) \ . \tag{6.164}$$

In general the tetrads are not fixed uniquely within a simplex, being invariant under the local Lorentz transformations discussed earlier in Sect. 6.4.

In the continuum the action for a massless spinor field is given by

$$I = \int dx \sqrt{g}\, \bar{\psi}(x) \gamma^\mu D_\mu \psi(x) \ , \tag{6.165}$$

where $D_\mu = \partial_\mu + \frac{1}{2}\omega_{\mu ab}\sigma^{ab}$ is the spinorial covariant derivative containing the spin connection $\omega_{\mu ab}$. It will be convenient to first consider only two neighboring simplices $s_1$ and $s_2$, covered by a common coordinate system $x^\mu$. When the two tetrads in $s_1$ and $s_2$ are made to coincide, one can then use a common set of gamma matrices $\gamma^\mu$ within both simplices. Then in the absence of torsion the covariant derivative $D_\mu$ in Eq. (6.165) reduces to just an ordinary derivative. The fermion field $\psi(x)$ within the two simplices can then be suitably interpolated, by writing for example

$$\psi(x) \;=\; \theta(n\cdot x)\,\psi(s_1) + [1-\theta(n\cdot x)]\,\psi(s_2) \;\;, \tag{6.166}$$

where $n_\mu$ is the common normal to the face $f(s_1,s_2)$ shared by the simplices $s_1$ and $s_2$, and chosen to point into $s_1$. Inserting the expression for $\psi(x)$ from Eq. (6.166) into Eq. (6.165) and applying the divergence theorem (or equivalently using the fact that the derivative of a step function only has support at the origin) one obtains

$$I \;=\; \tfrac{1}{2} V^{(d-1)}(f)\,(\bar\psi_1+\bar\psi_2)\,\gamma^\mu\,n_\mu\,(\psi_1-\psi_2) \;\;, \tag{6.167}$$

where $V^{(d-1)}(f)$ represents the volume of the $(d-1)$-dimensional common interface $f$, a tetrahedron in four dimensions. But the contributions from the diagonal terms containing $\bar\psi_1\psi_1$ and $\bar\psi_2\psi_2$ vanish when summed over the faces of an $n$-simplex, by virtue of the useful identity

$$\sum_{p=1}^{n+1} V(f^{(p)})\,n_\mu^{(p)} \;=\; 0 \;\;, \tag{6.168}$$

where $V(f^{(p)})$ are the volumes of the $p$ faces of a given simplex, and $n_\mu^{(p)}$ the inward pointing unit normals to those faces.

So far the above partial expression for the lattice spinor action was obtained by assuming that the tetrads $e_a^{\mu}(s_1)$ and $e_a^{\mu}(s_2)$ in the two simplices coincide. If they do not, then they will be related by a matrix $\mathbf{R}(s_2,s_1)$ such that

$$e_a^{\mu}(s_2) \;=\; R^{\mu}{}_{\nu}(s_2,s_1)\,e_a^{\nu}(s_1) \;\;, \tag{6.169}$$

and whose spinorial representation $\mathbf{S}$ was given previously for example in Eq. (6.24). Such a matrix $\mathbf{S}(s_2,s_1)$ is now needed to additionally parallel transport the spinor $\psi$ from a simplex $s_1$ to the neighboring simplex $s_2$.

The invariant lattice action for a massless spinor takes therefore the form

$$I \;=\; \tfrac{1}{2} \sum_{\text{faces f(ss}')} V[f(s,s')]\,\bar\psi_s\,\mathbf{S}[\mathbf{R}(s,s')]\,\gamma^\mu(s')\,n_\mu(s,s')\,\psi_{s'} \;\;, \tag{6.170}$$

where the sum extends over all interfaces $f(s,s')$ connecting one simplex $s$ to a neighboring simplex $s'$. As shown in (Drummond, 1986) it can be further extended to include a dynamical torsion field.

The above spinorial action can be considered analogous to the lattice Fermion action proposed originally in (Wilson, 1973) for non-Abelian gauge theories. It is possible that it still suffers from the Fermion doubling problem, although the situation is less clear for a dynamical lattice (Lehto, Nielsen and Ninomiya, 1987; Christ and Lee, 1982).

## 6.15 Gauge Fields

In the continuum a locally gauge invariant action coupling an $SU(N)$ gauge field to gravity is

$$I_{\text{gauge}} = -\frac{1}{4g^2} \int d^4x \sqrt{g}\, g^{\mu\lambda}\, g^{\nu\sigma}\, F^a_{\mu\nu} F^a_{\lambda\sigma} \ , \tag{6.171}$$

with $F^a_{\mu\nu} = \nabla_\mu A^a_\nu - \nabla_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu$ and $a = 1 \dots N^2 - 1$.

On the lattice one can follow a procedure analogous to Wilson's construction on a hypercubic lattice, with the main difference that the lattice is now simplicial. Given a link $ij$ on the lattice one assigns group element $U_{ij}$, with each $U$ an $N \times N$ unitary matrix with determinant equal to one, and such that $U_{ji} = U^{-1}_{ij}$. Then with each triangle (plaquettes) $\Delta$ labeled by the three vertices $ijk$ one associates a product of three $U$ matrices,

$$U_\Delta \equiv U_{ijk} = U_{ij} U_{jk} U_{ki} \ . \tag{6.172}$$

The discrete action is then given by (Christ Friedberg and Lee, 1982)

$$I_{\text{gauge}} = -\frac{1}{g^2} \sum_\Delta V_\Delta \frac{c}{A^2_\Delta} \operatorname{Re}\left[\operatorname{tr}(1 - U_\Delta)\right] \ , \tag{6.173}$$

with 1 the unit matrix, $V_\Delta$ the 4-volume associated with the plaquettes $\Delta$, $A_\Delta$ the area of the triangle (plaquettes) $\Delta$, and $c$ a numerical constant of order one. If one denotes by $\tau_\Delta = cV_\Delta / A_\Delta$ the $d-2$-volume of the dual to the plaquette $\Delta$, then the quantity

$$\frac{\tau_\Delta}{A_\Delta} = c \frac{V_\Delta}{A^2_\Delta} \ , \tag{6.174}$$

is simply the ratio of this dual volume to the plaquettes area. The edge lengths $l_{ij}$ and therefore the metric enter the lattice gauge field action through these volumes and areas.

One important property of the gauge lattice action of Eq. (6.173) is its local invariance under gauge rotations $g_i$ defined at the lattice vertices, and for which $U_{ij}$ on the link $ij$ transforms as

$$U_{ij} \to g_i U_{ij} g^{-1}_j \ . \tag{6.175}$$

These leave the product

$$\operatorname{tr}\left[(g_i U_{ij} g^{-1}_j)(g_j U_{jk} g^{-1}_k)(g_k U_{ki} g^{-1}_i)\right] = \operatorname{tr}\left[U_{ij} U_{jk} U_{ki}\right] \ , \tag{6.176}$$

and therefore the action invariant. One can further show that the discrete action of Eq. (6.173) goes over in the lattice continuum limit to the correct Yang-Mills action for manifolds that are smooth and close to flat.

## 6.16 Lattice Gravitino

Supergravity in four dimensions naturally contains a spin-$3/2$ gravitino, the supersymmetric partner of the graviton. In the case of $\mathcal{N} = 1$ supergravity these are the only two degrees of freedom present. The action contains, beside the Einstein-Hilbert action for the gravitational degrees of freedom, the Rarita-Schwinger action for the gravitino, as well as a number of additional terms (and fields) required to make the action manifestly supersymmetric off-shell. The idea in formulating supergravity on a lattice is to try to construct a lattice action with enough symmetry (local gauge and supersymmetry) so that when one eventually takes a lattice continuum limit one gets back all the desired symmetries of the original continuum theory.

Consider a spin-$3/2$ Majorana fermion in four dimensions, which correspond to self-conjugate Dirac spinors $\psi_\mu$, where the Lorentz index $\mu = 1 \dots 4$. In flat space the action for such a field is given by the Rarita-Schwinger term

$$\mathscr{L}_{RS} = - \tfrac{1}{2} \, \varepsilon^{\alpha\beta\gamma\delta} \, \psi_\alpha^T \, C \, \gamma_5 \, \gamma_\beta \, \partial_\gamma \, \psi_\delta \ . \tag{6.177}$$

Locally the action is invariant under the gauge transformation

$$\psi_\mu(x) \ \to \ \psi_\mu(x) + \partial_\mu \, \varepsilon(x) \ , \tag{6.178}$$

where $\varepsilon(x)$ is an arbitrary local Majorana spinor.

The construction of a suitable lattice action for the spin-$3/2$ particle proceeds in a way that is rather similar to what one does in the spin-$1/2$ case. On a simplicial manifold the Rarita-Schwinger spinor fields $\psi_\mu(s)$ and $\bar\psi_\mu(s)$ are most naturally placed at the center of each $d$-simplex $s$. Like the spin-$1/2$ case, the construction of a suitable lattice action requires the introduction of the Lorentz group and its associated vierbein fields $e_\mu^a(s)$ within each simplex labeled by $s$.

Within each simplex one can choose a representation of the Dirac gamma matrices, denoted here by $\gamma^\mu(s)$, such that in the local coordinate basis $\{\gamma^\mu(s), \gamma^\nu(s)\} = 2 g^{\mu\nu}(s)$, with $\{\gamma^a, \gamma^b\} = 2 \eta^{ab}$ and $\gamma^\mu(s) = e_a^\mu(s) \, \gamma^a$. Then within each simplex the vierbeins $e_\mu^a(s)$ satisfy the usual relation $e_a^\mu(s) \, e_b^\nu(s) \, \eta^{ab} = g^{\mu\nu}(s)$. In general the vierbeins are not fixed uniquely within a simplex, being invariant under the local Lorentz transformations discussed earlier in Sect. 6.4.

Now in the presence of gravity the continuum action for a massless spin-$3/2$ field is given by

$$I_{3/2} = -\tfrac{1}{2} \int dx \sqrt{g} \, \varepsilon^{\mu\nu\lambda\sigma} \, \bar\psi_\mu(x) \, \gamma_5 \, \gamma_\nu \, D_\lambda \, \psi_\sigma(x) \ , \tag{6.179}$$

with the Rarita-Schwinger field subject to the Majorana constraint $\psi_\mu = C \bar\psi_\mu(x)^T$. Here the covariant derivative is defined as

$$D_\nu \psi_\rho = \partial_\nu \psi_\rho - \Gamma_{\nu\rho}^\sigma \, \psi_\sigma + \tfrac{1}{2} \, \omega_{\nu ab} \, \sigma^{ab} \, \psi_\rho \ , \tag{6.180}$$

and involves *both* the standard affine connection $\Gamma^{\sigma}_{\nu\rho}$, as well as the vierbein connection

$$\omega_{\nu\, ab} = \tfrac{1}{2} \left[ e_a{}^{\mu} (\partial_\nu e_{b\mu} - \partial_\mu e_{b\nu}) + e_a{}^{\rho} e_b{}^{\sigma} (\partial_\sigma e_{c\rho}) e^c{}_\nu \right] , \\ - (a \leftrightarrow b) \tag{6.181}$$

with Dirac spin matrices $\sigma_{ab} = \frac{1}{2i} [\gamma_a, \gamma_b]$, and $\varepsilon^{\mu\nu\rho\sigma}$ the usual Levi-Civita tensor, such that $\varepsilon_{\mu\nu\rho\sigma} = -g\, \varepsilon^{\mu\nu\rho\sigma}$.

Next one considers just two neighboring simplices $s_1$ and $s_2$, covered by a common coordinate system $x^\mu$. When the two vierbeins in $s_1$ and $s_2$ are made to coincide, one can then use a common set of gamma matrices $\gamma^\mu$ within both simplices. Then (in the absence of torsion) the covariant derivative $D_\mu$ in Eq. (6.179) reduces to just an ordinary derivative. The fermion field $\psi_\mu(x)$ within the two simplices can then be suitably interpolated, and one obtains a lattice action expression very similar to the spinor case. One can then relax the condition that the vierbeins $e^\mu_a(s_1)$ and $e^\mu_a(s_2)$ in the two simplices coincide. If they do not, then they will be related by a matrix $\mathbf{R}(s_2, s_1)$ such that

$$e^\mu_a(s_2) \;=\; R^\mu{}_\nu(s_2, s_1)\, e^\nu_a(s_1) \;, \tag{6.182}$$

and whose spinorial representation $\mathbf{S}$ was given previously in Eq. (6.24). But the new ingredient in the spin-3/2 case is that, besides requiring a spin rotation matrix $\mathbf{S}(s_2, s_1)$, now one also needs the matrix $R^\nu_\mu(s, s')$ describing the corresponding parallel transport of the Lorentz *vector* $\psi_\mu(s)$ from a simplex $s_1$ to the neighboring simplex $s_2$.

An invariant lattice action for a massless spin-3/2 particle takes therefore the form

$$I \;=\; -\tfrac{1}{2} \sum_{\text{faces f(ss}')} V[f(s,s')]\, \varepsilon^{\mu\nu\lambda\sigma}\, \bar\psi_\mu(s)\, \mathbf{S}[\mathbf{R}(s,s')]\, \gamma_\nu(s')\, n_\lambda(s,s')\, R^\rho{}_\sigma(s,s')\, \psi_\rho(s') \tag{6.183}$$

with

$$\bar\psi_\mu(s)\, \mathbf{S}[\mathbf{R}(s,s')]\, \gamma_\nu(s')\, \psi_\rho(s') \;\equiv\; \bar\psi_{\mu\,\alpha}(s)\, S^\alpha{}_\beta[\mathbf{R}(s,s')]\, \gamma_\nu{}^\beta{}_\gamma(s')\, \psi^\gamma_\rho(s') \;, \tag{6.184}$$

and the sum $\sum_{\text{faces f(ss}')}$ extends over all interfaces $f(s,s')$ connecting one simplex $s$ to a neighboring simplex $s'$. When compared to the spin-1/2 case, the most important modification is the second rotation matrix $R^\nu{}_\mu(s,s')$, which describes the parallel transport of the fermionic vector $\psi_\mu$ from the site $s$ to the site $s'$, which is required in order to obtain locally a Lorentz scalar contribution to the action.

In supergravity one more term is needed in the action. In order to achieve local supersymmetry invariance, one needs an additional quartic fermion self-interaction term, given in Eq. (1.176), and which is of the form (Ferrara, Freedman and van Nieuwenhuizen, 1976)

$$\mathscr{L}_4 = \; -\frac{1}{32\kappa^2\sqrt{g}} \left( \varepsilon^{\tau\alpha\beta\nu} \varepsilon_\tau{}^{\gamma\delta\mu} + \varepsilon^{\tau\alpha\mu\nu} \varepsilon_\tau{}^{\gamma\delta\beta} - \varepsilon^{\tau\beta\mu\nu} \varepsilon_\tau{}^{\gamma\delta\alpha} \right) \\ \times (\bar\psi_\alpha \gamma_\mu \psi_\beta)(\bar\psi_\gamma \gamma_\nu \psi_\delta) \;, \tag{6.185}$$

here with $\kappa^2 \equiv 4\pi G$. Then one can show that the combined Lagrangian (containing the gravity part, the gravitino action and the four-fermion term) is invariant, up to terms of order $(\psi)^5$, under the simultaneous transformations

$$\begin{aligned} \delta e^a{}_\mu &= i\kappa\,\bar{\varepsilon}\,\gamma^a\,\psi_\mu \\ \delta g_{\mu\nu} &= i\kappa\,\bar{\varepsilon}\,[\gamma_\mu\,\psi_\nu + \gamma_\nu\,\psi_\mu] \\ \delta\psi_\mu &= \kappa^{-1} D_\mu\,\varepsilon + \tfrac{1}{4}\,i\kappa\,(2\,\bar{\psi}_\mu\gamma_a\psi_b + \bar{\psi}_a\gamma_\mu\,\psi_b)\,\sigma^{ab}\,\varepsilon \ , \end{aligned} \tag{6.186}$$

where $\varepsilon(x)$ in an arbitrary Majorana spinor. The lattice action will in general not preserve all of these symmetries, but one can hope that they will be restored in the quantum continuum limit. The difficulties one encounters in transcribing supersymmetry on a lattice are discussed, for example, in the recent review (Feo, 2003).

## 6.17 Alternate Discrete Formulations

The simplicial lattice formulation offers a natural way of representing gravitational degrees in a discrete framework by employing inherently geometric concepts such as areas, volumes and angles. It is possible though to formulate quantum gravity on a flat hypercubic lattice, in analogy to Wilson's discrete formulation for gauge theories, by putting the connection centerstage. In this new set of theories the natural variables are then lattice versions of the spin connection and the vierbein. Also, because the spin connection variables appear from the very beginning, it is much easier to incorporate fermions later. Some lattice models have been based on the pure Einstein theory (Smolin, 1979; Das, Kaku and Townsend, 1979; Mannion and Taylor, 1981; Caracciolo and Pelissetto, 1988), while others attempt to incorporate higher derivative terms (Tomboulis, 1984; Kondo, 1984).

Difficult arise when attempting to put quantum gravity on a flat hypercubic lattice a la Wilson, since it is not entirely clear what the gravity analogue of the Yang-Mills connection is. In continuum formulations invariant under the Poincaré or de Sitter group the action is invariant under a local extension of the Lorentz transformations, but not under local translations (Kibble, 1961). Local translations are replaced by diffeomorphisms which have a different nature. One set of lattice discretizations starts from the action of (MacDowell and Mansouri, 1977a,b; West, 1978) whose local invariance group is the de Sitter group $Spin(4)$, the covering group of $SO(4)$. In the lattice formulation of (Smolin, 1979; Das, Kaku and Townsend, 1979) the lattice variables are gauge potentials $e_{a\mu}(n)$ and $\omega_{\mu ab}(n)$ defined on lattice sites $n$, generating local $Spin(4)$ matrix transformations with the aid of the de Sitter generators $P_a$ and $M_{ab}$. The resulting lattice action reduces classically to the Einstein action with cosmological term in first order form in the limit of the lattice spacing $a \to 0$; to demonstrate the quantum equivalence one needs an additional zero torsion constraint. In the end the issue of lattice diffeomorphism invariance remains somewhat open, with the hope that such an invariance will be restored in the full quantum theory.

As an example, we will discuss here the approach of (Mannion and Taylor, 1981) which relies on a four-dimensional lattice discretization of the Einstein-Cartan theory with gauge group $SL(2,C)$, and does not initially require the presence of a cosmological constant, as would be the case if one had started out with the de Sitter group $Spin(4)$. On a lattice of spacing $a$ with vertices labelled by $n$ and directions by $\mu$ one relates the relative orientations of nearest-neighbor local $SL(2,C)$ frames by

$$U_\mu(n) = \Big[U_{-\mu}(n+\mu)\Big]^{-1} = \exp[\,i B_\mu(n)] \ , \tag{6.187}$$

with $B_\mu = \frac{1}{2} a B_\mu^{ab}(n) J_{ba}$, $J_{ba}$ being the set of six generators of $SL(2,C)$, the covering group of the Lorentz group $SO(3,1)$, usually taken to be

$$\sigma_{ab} = \frac{1}{2i}[\gamma_a, \gamma_b] \ , \tag{6.188}$$

with $\gamma_a$'s the Dirac gamma matrices. The local lattice curvature is then obtained in the usual way by computing the product of four parallel transport matrices around an elementary lattice square,

$$U_\mu(n)\, U_\nu(n+\mu)\, U_{-\mu}(n+\mu+\nu)\, U_{-\nu}(n+\nu) \ , \tag{6.189}$$

giving in the limit of small $a$ by the Baker-Hausdorff formula the value $\exp[iaR_{\mu\nu}(n)]$, where $R_{\mu\nu}$ is the Riemann tensor defined in terms of the spin connection $B_\mu$

$$R_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + i[B_\mu, B_\nu] \ . \tag{6.190}$$

If one were to write for the action the usual Wilson lattice gauge form

$$\sum_{n,\mu,\nu} \mathrm{tr}[\,U_\mu(n)\, U_\nu(n+\mu)\, U_{-\mu}(n+\mu+\nu)\, U_{-\nu}(n+\nu)] \ , \tag{6.191}$$

then one would obtain a curvature squared action proportional to $\sim \int R_{\mu\nu}^{\ \ ab} R^{\mu\nu}_{\ \ ab}$ instead of the Einstein-Hilbert one. One needs therefore to introduce lattice vierbeins $e_\mu^{\ b}(n)$ on the sites by defining the matrices

$$E_\mu(n) = a\, e_\mu^{\ a}\, \gamma_a \ . \tag{6.192}$$

Then a suitable lattice action is given by

$$I = \frac{i}{16\kappa^2} \sum_{n,\mu,\nu,\lambda,\sigma} \mathrm{tr}[\,\gamma_5\, U_\mu(n)\, U_\nu(n+\mu)\, U_{-\mu}(n+\mu+\nu)\, U_{-\nu}(n+\nu)\, E_\sigma(n)\, E_\lambda(n)] \ . \tag{6.193}$$

The latter is invariant under local $SL(2,C)$ transformations $\Lambda(n)$ defined on the lattice vertices

$$U_\mu \to \Lambda(n)\, U_\mu(n)\, \Lambda^{-1}(n+\mu) \ , \tag{6.194}$$

for which the curvature transforms as

$$\begin{aligned} & U_\mu(n)\, U_\nu(n+\mu)\, U_{-\mu}(n+\mu+\nu)\, U_{-\nu}(n+\nu) \\ & \quad \to \Lambda(n)\, U_\mu(n)\, U_\nu(n+\mu)\, U_{-\mu}(n+\mu+\nu)\, U_{-\nu}(n+\nu)\, \Lambda^{-1}(n) \ , \end{aligned} \tag{6.195}$$

and the vierbein matrices as

$$E_\mu(n) \to \Lambda(n)\, E_\mu(n)\, \Lambda^{-1}(n) \ . \tag{6.196}$$

Since $\Lambda(n)$ commutes with $\gamma_5$, the expression in Eq. (6.193) is invariant. The metric is then obtained as usual by

$$g_{\mu\nu}(n) = \tfrac{1}{4}\, \mathrm{tr}[E_\mu(n)\, E_\nu(n)] \ . \tag{6.197}$$

From the expression for the lattice curvature $R_{\mu\nu}^{ab}$ given above if follows immediately that the lattice action in the continuum limit becomes

$$I = \frac{a^4}{4\kappa^2} \sum_n \varepsilon^{\mu\nu\lambda\sigma}\, \varepsilon_{abcd}\, R_{\mu\nu}^{\ \ ab}(n)\, e_\lambda^{\ c}(n)\, e_\sigma^{\ d}(n) + O(a^6) \ , \tag{6.198}$$

which is the Einstein action in Cartan form

$$I = \frac{1}{4\kappa^2} \int d^4x\, \varepsilon^{\mu\nu\lambda\sigma}\, \varepsilon_{abcd}\, R_{\mu\nu}^{\ \ ab}\, e_\lambda^{\ c}\, e_\sigma^{\ d} \ , \tag{6.199}$$

with the parameter $\kappa$ identified with the Planck length. One can add more terms to the action; in this theory a cosmological term can be represented by

$$\lambda_0 \sum_n \varepsilon^{\mu\nu\lambda\sigma}\, \mathrm{tr}[\,\gamma_5\, E_\mu(n)\, E_\nu(n)\, E_\sigma(n)\, E_\lambda(n)] \ . \tag{6.200}$$

Both Eqs. (6.193) and (6.200) are locally $SL(2,C)$ invariant. The functional integral is then given by

$$Z = \int \prod_{n,\mu} dB_\mu(n) \prod_{n,\sigma} dE_\sigma(n)\, \exp\left\{-I(B,E)\right\} \ , \tag{6.201}$$

and from it one can then compute suitable quantum averages. Here $dB_\mu(n)$ is the Haar measure for $SL(2,C)$; it is less clear how to choose the integration measure over the $E_\sigma$'s, and how it should suitably constrained, which obscures the issue of diffeomorphism invariance in this theory.

In these theories it is possible to formulate curvature squared terms as well. In general for a hypercubic lattices the formulation of $R^2$-type terms in four dimensions involves constraints between the connections and the tetrads, which are a bit difficult to handle. Also there is no simple way of writing down topological invariants, which are either related to the Einstein action (in two dimensions), or are candidates for extra terms to be included in the action. A flat hypercubic lattice action has been written with higher derivative terms which appears to be reflection positive but has a very cumbersome form. These difficulties need not be present on a simplicial lattice (except that it is not known how to write an exact expression for the Hirzebruch signature in lattice terms).

There is another way of discretizing gravity, still using largely geometric concepts as is done in the Regge theory. In the dynamical triangulation approach (David, 1985) one fixes the edge lengths to unity, and varies the incidence matrix. As a result the volume of each simplex is fixed at

$$V_d = \frac{1}{d!}\sqrt{\frac{d+1}{2^d}} \ , \tag{6.202}$$

and all dihedral angles are given by the constant value

$$\cos\theta_d = \frac{1}{d} \ , \tag{6.203}$$

so that for example in four dimensions one has $\theta_d = \arccos(1/4) \approx 75.5^o$. Local curvatures are then determined by how many simplices $n_s(h)$ meet on a given hinge,

$$\delta(h) = 2\pi - n_s(h)\,\theta_d \ . \tag{6.204}$$

The action contribution from a single hinge is therefore from Eq. (6.38) $\delta(h)A(h) = \frac{1}{4}\sqrt{3}[2\pi - n_s(h)\,\theta_d]$ with $n_s$ a positive integer. In this model the local curvatures are inherently discrete, and there is no equivalent lattice notion of continuous diffeomorphisms, or for that matter of continuous local deformations corresponding, for example, to shear waves. Indeed it seems rather problematic in this approach to make contact with the continuum theory, as the model does not contain a metric, at least not in an explicit way. This fact has some consequences for the functional measure, since there is really no useful criterion which could be used to restrict it to the form suggested by invariance arguments, as detailed earlier in the discussion of the continuum functional integral for gravity. The hope is that for lattices made of some large number of simplices one would recover some sort of discrete version of diffeomorphism invariance. The aims of the dynamical triangulation approach have recently been reviewed in (Ambjørn, Jurkiewicz and Loll, 2005), and we refer the reader to further references therein. A recent discussion of attempts at simulating the Lorentzian case, which in principle leads to complex weights in the functional integral which are known to be difficult to handle correctly in numerical simulations (since the latter generally rely on positive probabilities) can be found in (Loll, 2007).

Another lattice approach closely related to the Regge theory described in this review is based on the so-called spin foam models, which have their origin in an observation found in (Ponzano and Regge, 1968) relating the geometry of simplicial lattices to the asymptotics of Racah angular momentum addition coefficients. The original Regge-Ponzano concepts were later developed into a spin model for gravity

(Hasslacher and Perry, 1981) based on quantum spin variables attached to lattice links. In these models representations of $SU(2)$ label edges. One natural underlying framework for such theories is the canonical $3+1$ approach to quantum gravity, wherein quantum spin variables are naturally related to $SU(2)$ spin connections. Extensions to four dimensions have been attempted, and we refer the reader to the recent review of spin foam models in (Perez, 2003).

## 6.18 Lattice Invariance versus Continuum Invariance

In simplicial lattice gravity, and for that matter, in any theory of lattice gravity, one could wonder what the effects might be due to the restrictions on the metric arising from the generalized triangle inequalities of Eq. (6.74). Effectively these imply a soft cutoff at large edge lengths, and possibly, for scale invariant measures, a second cutoff at small edge lengths, in addition to the cosmological constant term of Eq. (6.43), which exponentially suppresses large volumes.[2]

On the lattice one would ideally like to preserve all of the symmetries of the continuum theory, including some version of diffeomorphism invariance. Since a truly diffeomorphism invariant cutoff does not seem to exists, there might be difficulties in implementing such a program. The hope therefore is that the lattice theory has enough symmetry built into it from the start to fully recover the symmetries of the original theory in some suitable lattice continuum limit. One example of such a mechanism is the demonstrated restoration of rotational invariance in many examples of ordinary lattice field theories in the vicinity of ultraviolet fixed points. What is meant by having enough symmetry built into the lattice theory is the following: that the unwanted terms arising from the lattice regularization can in some sense be considered small because they lead to vanishing contributions in the limit of small momenta and large distances.

For gauge theories the proof that small violations of gauge invariance do not affect the long distance properties of the theory, which is therefore still described by a locally gauge invariant effective action, goes as follows (Foerster, Nielsen and Ninomiya, 1980; Parisi, 1992). One first assumes that under local gauge transformations of the gauge fields $A_\mu(x)$

$$(A_\mu)_\Omega \; = \; \Omega^{-1} A_\mu \, \Omega + \Omega^{-1} \, \partial_\mu \, \Omega \;\; , \tag{6.205}$$

the action can be decomposed as a gauge invariant part $S_0(A)$ , plus a small non-invariant contribution $\delta S(A)$

$$S(A) = S_0(A) + \delta S(A) = \int dx \, [\mathscr{L}_0[A(x)] + \delta \mathscr{L}[A(x)]] \;\; . \tag{6.206}$$

[2] In the continuum such delicate field cutoff do not arise, since in a perturbative calculation the integration domain for the metric perturbation $h_{\mu\nu}$ is generally extended from minus infinity to plus infinity, leading in the usual way to straightforward Gaussian integrations, without regards to field constraints such as $\det g > 0$.

Furthermore it will be assumed that the functional measure $[dA]$ is also locally gauge invariant, so that one has

$$[dA_\Omega] = [dA] \qquad S_0(A_\Omega) = S_0(A) \ . \tag{6.207}$$

Here and in the following a lattice regularization will be implicit, in order to make various functional manipulations well defined. In addition, the fact that the gauge theory is compact will play a crucial role, as the proof generally does not go through if the gauge variables are not compact. The compactness of the gauge group implies for the measure over the gauge parameters

$$\int [d\Omega] = \text{finite} \ , \tag{6.208}$$

and furthermore by invariance $[d\Omega] = [d(\Sigma\Omega)]$ with $\Sigma$ an arbitrary group element. Thus both $[dA]$ and $[d\Omega]$ will be assumed to be the Haar measure over the group. Under a gauge transformation for the fields one has

$$S(A_{\Omega^{-1}}) = S_0(A) + \delta S(A_{\Omega^{-1}}) \ , \tag{6.209}$$

since the first term is assumed to be invariant. As a consequence only the term $\delta S(A)$ causes gauge breaking in $Z$.

Using perturbation theory in $\delta S(A)$ one can compute the vacuum expectation value of the gauge invariant quantity $F(A)$

$$\langle F(A) \rangle = \int [dA]\, e^{-S(A)}\, F(A) \,/\, \int [dA]\, e^{-S(A)} \ . \tag{6.210}$$

For small $\delta S(A)$ it is given by the expansion

$$\begin{aligned} \langle F(A) \rangle = \langle F(A) \rangle_0 - \int dx\, \langle \delta\mathscr{L}(x) F(A) \rangle_0 \\ + \tfrac{1}{2} \int dx dx' \, \langle \delta\mathscr{L}(x)\, \delta\mathscr{L}(x') F(A) \rangle_0^c + \dots \end{aligned} \tag{6.211}$$

Gauge invariance of the lowest order action $S_0(A)$ and measure $[dA]$, used in the average $\langle \dots \rangle_0$, implies that the first correction on the r.h.s. vanishes. The second order correction only gives a contribution when $x = x'$, which is given by

$$\tfrac{1}{2} \int dx \langle\, [\delta\mathscr{L}(x)]^2 F(A)\, \rangle_0^c = \tfrac{1}{2} \int dx \langle\, \delta\mathscr{I}(x) F(A)\, \rangle_0^c \ , \tag{6.212}$$

where $\delta\mathscr{I}(x)$ is some gauge invariant operator. One concludes that a small gauge breaking term leads to a correction which can be described by the insertion of a suitable gauge-invariant operator. Such effects could then be re-absorbed into a suitable redefinition of the bare coupling constants of the theory.

To actually show this, it will be useful to introduce an artificial integration $\int [d\Omega]$ over the gauge group parameters, by writing for the Euclidean partition function

$$\begin{aligned} Z &= \int [dA]\, e^{-S(A)} \\ &= \text{const.} \int [d\Omega]\, [dA]\, e^{-S(A)} \\ &= \text{const.} \int [d\Omega]\, [dA_\Omega]\, e^{-S(A)} \\ &= \text{const.} \int [d\Omega]\, [dA]\, e^{-S(A_{\Omega^{-1}})} \ . \end{aligned} \tag{6.213}$$

To take into account the effect of the $\Omega$ variables one first integrates over them, and then deduces from this procedure an effective action for the $A$ fields,

$$S_{\text{eff}}(A) \; = \; S_0(A) + \delta S_{\text{inv}}(A) \ . \tag{6.214}$$

The new effective action is made out of the original invariant contribution $S_0(A)$, plus a correction from $\delta S_{\text{inv}}(A)$, which is obtained from

$$\exp\{-\delta S_{\text{inv}}(A)\} \; = \; \int [d\Omega] \; \exp\{-S(A_{\Omega^{-1}})\} \ . \tag{6.215}$$

In general there is no guarantee that the additional contribution $\delta S_{\text{inv}}(A)$ will be local, i.e. of the form

$$\delta S_{\text{inv}}(A) \; = \; \int dx\, \delta \mathscr{L}_{\text{inv}}[A(x)] \ . \tag{6.216}$$

But the $\delta S_{\text{inv}}(A)$ term will indeed be local provided certain conditions are met, i.e. that the correlations in the $\Omega$ variables will be sufficiently short ranged. Then at distances large compared to the correlation length $\xi_\Omega$ of the $\Omega$ variables it is possible to expand $\delta \mathscr{L}_{\text{inv}}[A(x)]$ in terms of locally gauge invariant terms of the type $\text{tr}(F^{\mu\nu})^2$, $\text{tr}(\nabla_\sigma F^{\mu\nu})^2$, etc. At sufficiently large distances one expects only terms with the lowest dimensions to be important, the leading one being $\text{tr}(F^{\mu\nu})^2$, whose effect will be just to renormalize the bare gauge coupling.

Clearly the argument for the decoupling of the $\Omega$ variables only works if there are long range correlations in the unperturbed gauge variables $A_\mu$ (so that the two correlation lengths can be compared to each other, and thus the notion of "short range" makes sense) which implies $g \gg 1$ in an asymptotically free gauge theory, or in general that one is close to an ultraviolet fixed point of the gauge theory.

In summary, the expectation is that if the non-invariant correction is small there are no significant changes compared to the invariant theory, which is a rather remarkable result. On the other hand if the non-invariant correction is large then a phase transition appears and one moves into a qualitatively new phase. The argument is nice because of its simplicity, but still leaves one major issue open: namely what is meant exactly by a small non-invariant perturbation, an aspect that will presumably have to be addressed individually case by case.

How do these considerations apply to gravity? There are two major issues that stand in the way of taking over the result for gauge theories. the first one is the fact that the gauge group of continuum gravity, the diffeomorphism group, is not compact. On the lattice the situation is less clear, due to the triangle inequality constraints of Eq. (6.74), and the fact that in some lattice formulations one can take some lattice variables to be compact [see for example the lattice theory described in Eqs. (6.193) and (6.200)].

In the case of lattice gravity one would again decompose the lattice action into an invariant part $S_0(g)$, plus a small non-invariant contribution $\delta S(g)$

$$S(g) \;=\; S_0(g) + \delta S(g) \;=\; \int dx \, \left[ \mathscr{L}_0[g_{\mu\nu}(x)] + \delta \mathscr{L}[g_{\mu\nu}(x)] \right] \;, \tag{6.217}$$

and one would need to assume again that the functional measure $[dg]$ is also invariant. The volume of the diffemorphism group will of course cancel out when one computes vacuum averages, as in

$$\langle F(g) \rangle \;=\; \int [dg] \, e^{-S(g)} \, F(g) \; / \; \int [dg] \, e^{-S(g)} \;\; . \tag{6.218}$$

Proceeding in analogy to the gauge theory case, one finds, after integrating over the variables $\Omega$ of Eq. (1.11), that the effective action is made out of the original invariant contribution $S_0(g)$, plus the correction from $\delta S_{\rm inv}(g)$,

$$\exp\{-\delta S_{\rm inv}(g)\} \;=\; \int [d\Omega] \; \exp\{-S(g_{\Omega^{-1}})\} \;\; . \tag{6.219}$$

The non-compact nature of the group might cause problems in the above integral over gauge parameters $\Omega$. It is therefore more appropriate to look instead at the physically more relevant difference in action between two metrics $g$ and $g'$

$$\exp\left\{-\delta S_{\rm inv}(g) + \delta S_{\rm inv}(g')\right\} \;=\; \frac{\int [d\Omega] \;\; \exp\{-S(g_{\Omega^{-1}})\}}{\int [d\Omega'] \; \exp\left\{-S({g'}_{\Omega'^{-1}})\right\}} \;\; . \tag{6.220}$$

As in the gauge case, there is in general no guarantee that the additional contribution $\delta S_{\rm inv}(g)$ will be local, i.e. of the form

$$\delta S_{\rm inv}(g) \;=\; \int dx \; \delta \mathscr{L}_{\rm inv}[g_{\mu\nu}(x)] \;, \tag{6.221}$$

but $\delta S_{\rm inv}(g)$ will indeed be local provided the correlations in the $\Omega$ variables will be sufficiently short ranged.

Then at distances large compared to the correlation length $\xi_\Omega$ of the $\Omega$ variables it should be possible to expand $\delta \mathscr{L}_{\rm inv}[g_{\mu\nu}(x)]$ in terms of invariant terms of the type $\sqrt{g}$, $\sqrt{g}R$, $\sqrt{g}R_{\mu\nu}^2$, etc. At sufficiently large distances only terms with the lowest dimensions are expected to be important, the leading one being $\sqrt{g}R$, whose effect will be to renormalize the bare Newton coupling $G$. Finally, as in the gauge theory case, the argument for the decoupling of the $\Omega$ variables only works if there are long range correlations in the unperturbed gauge variables $g_{\mu\nu}$, which implies that one is close to a gravitational ultraviolet fixed point.

# Chapter 7
# Analytical Lattice Expansion Methods

## 7.1 Motivation

The following sections will discuss a number of instances where the lattice theory of quantum gravity can be investigated analytically, subject to some necessary simplifying assumptions.

The first problem deals with the lattice weak field expansion about a flat background. It will be shown that in this case the relevant modes are the lattice analogs of transverse-traceless deformations.

The second problem involves the strong coupling (large $G$) expansion, where the weight factor in the path integral is expanded in powers of $1/G$. The domain of validity of this expansion can be regarded as somewhat complementary to the weak field limit.

The third case to be discussed is what happens in lattice gravity in the limit of large dimensions $d$, which formally is similar in some ways to the large-$N$ expansion discussed previously in this review. In this limit one can derive exact estimates for the phase transition point and for the scaling dimensions.

## 7.2 Lattice Weak Field Expansion and Transverse-Traceless Modes

One of the simplest possible problems that can be treated in quantum Regge calculus is the analysis of small fluctuations about a fixed flat Euclidean simplicial background (Roček and Williams, 1981; 1984). In this case one finds that the lattice graviton propagator in a De Donder-like gauge is precisely analogous to the continuum expression.

To compute an expansion of the lattice Regge action

$$I_R \propto \sum_{\text{hinges}} \delta(l)\, A(l) \ , \tag{7.1}$$

to quadratic order in the lattice weak fields one needs first and second variations with respect to the edge lengths. In four dimensions the first variation of the lattice Regge action is given by

$$\delta I_R \propto \sum_{\text{hinges}} \delta \cdot \left( \sum_{\text{edges}} \frac{\partial A}{\partial l}\, \delta l \right) \ , \tag{7.2}$$

since Regge has shown that the term involving the variation of the deficit angle $\delta$ vanishes (here the variation symbol should obviously not be confused with the deficit angle). Furthermore in flat space all the deficit angles vanish, so that the second variation is given simply by

$$\delta^2 I_R \propto \sum_{\text{hinges}} \left( \sum_{\text{edges}} \frac{\partial \delta}{\partial l}\, \delta l \right) \cdot \left( \sum_{\text{edges}} \frac{\partial A}{\partial l}\, \delta l \right) \ . \tag{7.3}$$

Next a specific lattice structure needs to be chosen as a background geometry. A natural choice is to use a flat hypercubic lattice, made rigid by introducing face diagonals, body diagonals and hyperbody diagonals, which results into a subdivision of each hypercube into $d!$ (here 4!=24) simplices. This subdivision is shown in Fig. 7.1.

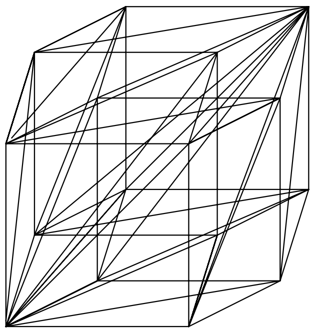


**Fig. 7.1** A four-dimensional hypercube divided up into four-simplices.

By a simple translation, the whole lattice can then be constructed from this one elemental hypercube. Consequently there will be $2^d - 1 = 15$ lattice fields per point, corresponding to all the edge lengths emanating in the positive lattice directions from any one vertex. Note that the number of degrees per lattice point is slightly larger than what one would have in the continuum, where the metric $g_{\mu\nu}(x)$ has $d(d+1)/2 = 10$ degrees of freedom per spacetime point $x$ in four dimensions (perturbatively, the physical degrees of freedom in the continuum are much less: $\frac{1}{2}d(d+1) - 1 - d - (d-1) = \frac{1}{2}d(d-3)$, for a traceless symmetric tensor, and after imposing gauge conditions). Thus in four dimensions each lattice hypercube will contain 4 body principals, 6 face diagonals, 4 body diagonals and one hyperbody diagonal. Within a given hypercube it is quite convenient to label the coordinates of the vertices using a binary notation, so that the four body principals with coordinates $(1,0,0,0)\ldots(0,0,0,1)$ will be labeled by integers 1,2,4,8, and similarly for the other vertices (thus for example the vertex $(0,1,1,0)$, corresponding to a face diagonal along the second and third Cartesian direction, will be labeled by the integer 6).

For a given lattice of fixed connectivity, the edge lengths are then allowed to fluctuate around an equilibrium value $l_i^0$

$$l_i = l_i^0\,(1+\varepsilon_i)\ . \tag{7.4}$$

In the case of the hypercubic lattice subdivided into simplices, the unperturbed edge lengths $l_i^0$ take on the values $1,\sqrt{2},\sqrt{3},2$, depending on edge type. The second variation of the action then reduces to a quadratic form in the 15-component small fluctuation vector $\varepsilon_n$

$$\delta^2 I_R\ \propto\ \sum_{mn} \varepsilon_m^T\, M_{mn}\, \varepsilon_n\ . \tag{7.5}$$

Here $M$ is the small fluctuation matrix, whose inverse determines the free lattice graviton propagator, and the indices $m$ and $n$ label the sites on the lattice. But just as in the continuum, $M$ has zero eigenvalues and cannot therefore be inverted until one supplies an appropriate gauge condition. Specifically, one finds that the matrix $M$ in four dimensions has four zero modes corresponding to periodic translations of the lattice, and a fifth zero mode corresponding to periodic fluctuations in the hyperbody diagonal. After block-diagonalization it is found that 4 modes completely decouple and are constrained to vanish, and thus the remaining degrees of freedom are 10, as in the continuum, where the metric has 10 independent components. The wrong sign for the conformal mode, which is present in the continuum, is also reproduced by the lattice propagator.

Due to the locality of the original lattice action, the matrix $M$ can be considered local as well, since it only couples edge fluctuations on neighboring lattice sites. In Fourier space one can write for each of the fifteen displacements $\varepsilon_n^{i+j+k+l}$, defined at the vertex of the hypercube with labels $(i,j,k,l)$,

$$\varepsilon_n^{i+j+k+l}\ =\ (\omega_1)^i(\omega_2)^j(\omega_4)^k(\omega_8)^l\ \varepsilon_n^0\ , \tag{7.6}$$

with $\omega_1 = e^{ik_1}$, $\omega_2 = e^{ik_2}$, $\omega_4 = e^{ik_3}$ and $\omega_8 = e^{ik_4}$ (it will be convenient in the following to use binary notation for $\omega$ and $\varepsilon$, but the regular notation for $k_i$). Here and in the following we have set the lattice spacing $a$ equal to one.

In this basis the matrix $M$ reduces to a block-diagonal form, with entries given by the $15\times 15$ dimensional matrices

$$M_\omega\ =\ \begin{pmatrix} A_{10} & B & 0\\ B^\dagger & 18 I_4 & 0\\ 0 & 0 & 0 \end{pmatrix}, \tag{7.7}$$

where $A_{10}$ is a $10 \times 10$ dimensional matrix, $B$ a $10 \times 4$ dimensional matrix and $I_4$ is the $4 \times 4$ dimensional identity matrix. Explicitly the above cited authors find

$$\begin{aligned}
(M_\omega)_{1,1} &= 6 \\
(M_\omega)_{1,2} &= \omega_1(\omega_4 + \omega_8) + \bar{\omega}_2(\bar{\omega}_4 + \bar{\omega}_8) \\
(M_\omega)_{1,3} &= 2 + 2\bar{\omega}_2 \\
(M_\omega)_{1,6} &= 2\omega_1 + 2\bar{\omega}_2\bar{\omega}_4 \\
(M_\omega)_{1,7} &= \bar{\omega}_2 + \bar{\omega}_4 \\
(M_\omega)_{1,14} &= 0 \\
(M_\omega)_{3,3} &= 4 \\
(M_\omega)_{3,5} &= \omega_2 + \bar{\omega}_4 \\
(M_\omega)_{3,12} &= 0 \\
(M_\omega)_{3,7} &= 1 + \bar{\omega}_4 \\
(M_\omega)_{3,13} &= 0 \ ,
\end{aligned} \tag{7.8}$$

where the remaining non-vanishing matrix elements can be obtained either by permuting appropriate indices, or by complex conjugation.

Besides one obvious zero eigenvalue, corresponding to a periodic fluctuation in $\varepsilon_{15}$, the matrix $M_\omega$ exhibits four additional zero modes corresponding to the four-parameter group of translations in flat space. An explicit form for these eigenmodes is

$$\begin{aligned}
\varepsilon_i &= (1 - \omega_i)\, x_i \\
\varepsilon_{i+j} &= \tfrac{1}{2}(1 - \omega_i\omega_j)\,(x_i + x_j) \\
\varepsilon_{i+j+k} &= \tfrac{1}{3}(1 - \omega_i\omega_j\omega_k)\,(x_i + x_j + x_k) \\
\varepsilon_{i+j+k+l} &= \tfrac{1}{4}(1 - \omega_i\omega_j\omega_k\omega_l)\,(x_i + x_j + x_k + x_l) \ ,
\end{aligned} \tag{7.9}$$

with $i, j, k, l = 1, 2, 4, 8$ and $i \neq j \neq k \neq l$.

The next step consists in transforming the lattice action $M_\omega$ into a form more suitable for comparison with the continuum action. To this end a set of transformations are performed sequentially, the first of which involves the matrix

$$S = \begin{pmatrix} I_{10} & 0 & 0 \\ -\frac{1}{18}B^\dagger & I_4 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{7.10}$$

which rotates $M_\omega$ into

$$M'_\omega = S^\dagger M_\omega S = \begin{pmatrix} A_{10} - \frac{1}{18}BB^\dagger & 0 & 0 \\ 0 & 18 I_4 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{7.11}$$

thus completely decoupling the (body diagonal) fluctuations $\varepsilon_7, \varepsilon_{11}, \varepsilon_{13}, \varepsilon_{14}$. These in turn can now be integrated out, as they appear in the action with no $\omega$ (i.e. derivative) term. As a result the number of dynamical degrees of freedom has been reduced from 15 to 10, the same number as in the continuum.

The remaining dynamics is thus encoded in the $10 \times 10$ dimensional matrix $L_\omega = A_{10} - \frac{1}{18} B B^\dagger$. By a second rotation, here affected by the matrix $T$, it can finally be brought into the form

$$\tilde{L}_\omega \;=\; T^\dagger L_\omega T \;=\; \left[8 - (\Sigma + \bar{\Sigma})\right] \begin{pmatrix} \frac{1}{2}\beta & 0 \\ 0 & I_6 \end{pmatrix} - C^\dagger C \;, \tag{7.12}$$

with the matrix $\beta$ given by

$$\beta \;=\; \frac{1}{2} \begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{pmatrix} \;. \tag{7.13}$$

The other matrix $C$ appearing in the second term is given by

$$C = \begin{pmatrix} f_1 & 0 & 0 & 0 & \tilde{f}_2 & \tilde{f}_4 & 0 & \tilde{f}_8 & 0 & 0 \\ 0 & f_2 & 0 & 0 & \tilde{f}_1 & 0 & \tilde{f}_4 & 0 & \tilde{f}_8 & 0 \\ 0 & 0 & f_4 & 0 & 0 & \tilde{f}_1 & \tilde{f}_2 & 0 & 0 & \tilde{f}_8 \\ 0 & 0 & 0 & f_8 & 0 & 0 & 0 & \tilde{f}_1 & \tilde{f}_2 & \tilde{f}_4 \end{pmatrix} \;, \tag{7.14}$$

with $f_i \equiv \omega_i - 1$ and $\tilde{f}_i \equiv 1 - \bar{\omega}_i$. Furthermore $\Sigma = \sum_i \omega_i$, and for small momenta one finds

$$8 - (\Sigma + \bar{\Sigma}) \;=\; 8 - \sum_{i=1}^{4} (e^{ik_i} + e^{-ik_i}) \;\sim\; k^2 + O(k^4) \;, \tag{7.15}$$

which shows that the surviving terms in the lattice action are indeed quadratic in $k$. The rotation matrix $T$ involved in the last transformation is given by

$$T \;=\; \begin{pmatrix} \Omega_4 \beta & 0 \\ 0 & I_6 \end{pmatrix} \begin{pmatrix} I_4 & 0 \\ \Omega_6 \gamma & I_6 \end{pmatrix} \;, \tag{7.16}$$

with $\Omega_4 = \mathrm{diag}(\omega_1, \omega_2, \omega_4, \omega_8)$ and $\Omega_6 = \mathrm{diag}(\omega_1\omega_2, \omega_1\omega_4, \omega_2\omega_4, \omega_1\omega_8, \omega_2\omega_8, \omega_4\omega_8)$, and

$$\gamma \;=\; -\frac{1}{2} \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \;. \tag{7.17}$$

At this point one is finally ready for a comparison with the continuum result, namely with the Lagrangian for pure gravity in the weak field limit as given in Eq. (1.7)

$$\mathscr{L}_{sym} = -\tfrac{1}{2}\partial_\lambda h_{\lambda\mu}\,\partial_\mu h_{\nu\nu} + \tfrac{1}{2}\partial_\lambda h_{\lambda\mu}\,\partial_\nu h_{\nu\mu} - \tfrac{1}{4}\partial_\lambda h_{\mu\nu}\,\partial_\lambda h_{\mu\nu} + \tfrac{1}{4}\partial_\lambda h_{\mu\mu}\,\partial_\lambda h_{\nu\nu} \;. \tag{7.18}$$

The latter can be conveniently split into two parts, as was done already in Eq. (1.67), as follows

$$\mathscr{L}_{sym} = -\tfrac{1}{2}\partial_\lambda h_{\alpha\beta} V_{\alpha\beta\mu\nu}\,\partial_\lambda h_{\mu\nu} + \tfrac{1}{2}C^2 \tag{7.19}$$

with

$$V_{\alpha\beta\mu\nu} = \tfrac{1}{2}\eta_{\alpha\mu}\eta_{\beta\nu} - \tfrac{1}{4}\eta_{\alpha\beta}\eta_{\mu\nu} \;, \tag{7.20}$$

or as a matrix,

$$V = \begin{pmatrix} \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} , \tag{7.21}$$

with metric components $11, 22, 33, 44, 12, 13, 14, 23, 24, 34$ more conveniently labeled sequentially by integers $1 \dots 10$, and the gauge fixing term $C_\mu$ given by the term in Eq. (1.68)

$$C_\mu = \partial_\nu h_{\mu\nu} - \tfrac{1}{2}\partial_\mu h_{\nu\nu} \;. \tag{7.22}$$

The above expression is still not quite the same as the lattice weak field action, but a simple transformation to trace reversed variables $\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\delta_{\mu\nu}h_{\lambda\lambda}$ leads to

$$\mathscr{L}_{sym} = \tfrac{1}{2}k_\lambda \bar{h}_i V_{ij} k_\lambda \bar{h}_j - \tfrac{1}{2}\bar{h}_i (C^\dagger C)_{ij}\bar{h}_j \;, \tag{7.23}$$

with the matrix $V$ given by

$$V_{ij} = \begin{pmatrix} \frac{1}{2}\beta & 0 \\ 0 & I_6 \end{pmatrix} , \tag{7.24}$$

with $k = i\partial$. Now $\beta$ is the same as the matrix in Eq. (7.13), and $C$ is nothing but the small $k$ limit of the matrix by the same name in Eq. (7.14), for which one needs to set $\omega_i - 1 \simeq i k_i$. The resulting continuum expression is then recognized to be identical to the lattice weak field results of Eq. (7.12).

This concludes the outline of the proof of equivalence of the lattice weak field expansion of the Regge action to the corresponding continuum expression. To summarize, there are several ingredients to this proof, the first of which is a relatively straightforward weak field expansion of both actions, and the second of which is the correct identification of the lattice degrees of freedom $\varepsilon_i(n)$ with their continuum counterparts $h_{\mu\nu}(x)$, which involves a sequence of non-trivial $\omega$-dependent transformations, expressed by the matrices $S$ and $T$. One more important aspect of the process is the disappearance of redundant lattice variables (five in the case of the hypercubic lattice), whose dynamics turns out to be trivial, in the sense that the associated degrees of freedom are non-propagating.

It is easy to see that the sequence of transformations expressed by the matrices $S$ of Eq. (7.10) and $T$ of Eq. (7.16), and therefore ultimately relating the lattice fluctuations $\varepsilon_i(n)$ to their continuum counterparts $h_{\mu\nu}(x)$, just reproduces the expected relationship between lattice and continuum fields. On the one hand one has $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, where $\eta_{\mu\nu}$ is the flat metric. At the same time one has from Eq. (6.3) for each simplex within a given hypercube

$$g_{ij} = \tfrac{1}{2}(l_{0i}^2 + l_{0j}^2 - l_{ij}^2) \ . \tag{7.25}$$

By inserting $l_i = l_i^0\,(1+\varepsilon_i)$, with $l_i^0 = 1, \sqrt{2}, \sqrt{3}, 2$ for the body principal ($i = 1,2,4,8$), face diagonal ($i = 3,5,6,9,10,12$), body diagonal ($i = 7,11,13,14$) and hyperbody diagonal ($i = 15$), respectively, one gets for example $(1+\varepsilon_1)^2 = 1 + h_{11}$, $(1+\varepsilon_3)^2 = 1 + \frac{1}{2}\,(h_{11} + h_{22}) + h_{12}$ etc., which in turn can then be solved for the $\varepsilon$'s in terms of the $h_{\mu\nu}$'s. One would then obtain

$$\begin{aligned}
\varepsilon_1 &= -1 + [1 + h_{11}]^{1/2} \\
\varepsilon_3 &= -1 + [1 + \tfrac{1}{2}(h_{11} + h_{22}) + h_{12})]^{1/2} \\
\varepsilon_7 &= -1 + [1 + \tfrac{1}{3}(h_{11} + h_{22} + h_{33}) \\
&\qquad\qquad + \tfrac{2}{3}(h_{12} + h_{23} + h_{13})]^{1/2} \\
\varepsilon_{15} &= -1 + [1 + \tfrac{1}{4}(h_{11} + h_{22} + h_{33} + h_{44}) \\
&\qquad\qquad + \tfrac{3}{4}(h_{12} + h_{13} + h_{14} + h_{23} + h_{24} + h_{34})]^{1/2} \ ,
\end{aligned} \tag{7.26}$$

and so on for the other edges, by suitably permuting indices. These relations can then be expanded out for weak $h$, giving for example

$$\begin{aligned}
\varepsilon_1 &= \tfrac{1}{2}\,h_{11} + O(h^2) \\
\varepsilon_3 &= \tfrac{1}{2}\,h_{12} + \tfrac{1}{4}\,(h_{11} + h_{22}) + O(h^2) \\
\varepsilon_7 &= \tfrac{1}{6}\,(h_{12} + h_{13} + h_{23}) + \tfrac{1}{6}\,(h_{23} + h_{13} + h_{12}) \\
&\quad + \tfrac{1}{6}\,(h_{11} + h_{22} + h_{33}) + O(h^2) \ ,
\end{aligned} \tag{7.27}$$

and so on. The above correspondence between the $\varepsilon$'s and the $h_{\mu\nu}$ are the underlying reason for the existence of the rotation matrices $S$ and $T$ of Eqs. (7.10) and (7.16), with one further important amendment: on the hypercubic lattice four edges within a given simplex are assigned to one vertex, while the remaining six edges are assigned to neighboring vertices, and require therefore a translation back to the base vertex of the hypercube, using the result of Eq. (7.6). This explains the additional factors of $\omega$ appearing in the rotation matrices $S$ and $T$. More importantly, one would expect such a combined rotation to be independent of what particular term in the lattice action one is considering, implying that it can be used to relate other lattice gravity contributions, such as the cosmological term and higher derivative terms, to their continuum counterparts.

The lattice action has a local gauge invariance, whose explicit form in the weak field limit was given in Eq. (7.9). This local invariance has $d$ parameters in $d$ dimensions and describes lattice diffeomorphisms. In the quantum theory, such local gauge invariance implies the existence of Ward identities for $n$-point functions. The choice of gauge in Eq. (7.22) is of course not the only possible one. Another possible choice is the so-called vacuum gauge for which in the continuum $h_{ik,k} = 0$, $h_{00} = h_{0i} = 0$. Expressed in terms of the lattice small fluctuation variables such a condition reads in momentum space

$$\begin{aligned}
e_8 &= 0 \\
e_9 &= \tfrac{1}{2}\,\omega_8\, e_1 \\
e_{10} &= \tfrac{1}{2}\,\omega_8\, e_2 \\
e_{12} &= \tfrac{1}{2}\,\omega_8\, e_4 \\
e_{11} &= \tfrac{1}{3}\,(1+\omega_8)e_3 - \tfrac{1}{6}\,(1-\omega_8)(\omega_2 e_1 + \omega_1 e_2) \\
e_{13} &= \tfrac{1}{3}\,(1+\omega_8)e_5 - \tfrac{1}{6}\,(1-\omega_8)(\omega_4 e_1 + \omega_1 e_4) \\
e_{14} &= \tfrac{1}{3}\,(1+\omega_8)e_6 - \tfrac{1}{6}\,(1-\omega_8)(\omega_2 e_4 + \omega_4 e_2)\ .
\end{aligned} \tag{7.28}$$

One can then evaluate the lattice action in such a gauge and again compare to the continuum expression. First one expands again the $e_i$'s in terms of the $h_{ij}$'s, as given in Eq. (7.26), and then expand out the $\omega$'s in powers of $k$. If one then sets $k_4 = 0$ one finds that the resulting contribution can be re-written as the sum of two parts (Hamber and Williams, 2005a,b), the first part being the transverse-traceless contribution

$$\tfrac{1}{4}\mathbf{k}^2\,\mathrm{Tr}\,[\,{}^3h[P\,{}^3hP - \tfrac{1}{2}P\mathrm{Tr}\,(P\,{}^3h)]] = \tfrac{1}{4}\mathbf{k}^2\,\bar{h}^{TT}_{ij}(\mathbf{k})\,h^{TT}_{ij}(\mathbf{k}) \tag{7.29}$$

$$\bar{h}^{TT}_{ij}\,h^{TT}_{ij} \equiv \mathrm{Tr}\,[\,{}^3h[P\,{}^3hP - \tfrac{1}{2}P\mathrm{Tr}\,(P\,{}^3h)]]\ , \tag{7.30}$$

with $P_{ij} = \delta_{ij} - k_i k_j/\mathbf{k}^2$ acting on the three-metric ${}^3h_{ij}$, and the second part arising due to the trace component of the metric

$$-\tfrac{1}{4}\mathbf{k}^2\,\mathrm{Tr}\,[P\mathrm{Tr}\,(P\,{}^3h)P\mathrm{Tr}\,(P\,{}^3h)] \;=\; \mathbf{k}^2\,\bar{h}^{T}_{ij}(\mathbf{k})\,h^{T}_{ij}(\mathbf{k})\ , \tag{7.31}$$

with $h^T = \frac{1}{2}P\mathrm{Tr}\,(P\,{}^3h)$. In the vacuum gauge $h_{ik,k} = 0$, $h_{ii} = 0$, $h_{0i} = 0$ one can further solve for the metric components $h_{12}$, $h_{13}$, $h_{23}$ and $h_{33}$ in terms of the two remaining degrees of freedom, $h_{11}$ and $h_{22}$,

$$h_{12} = -\frac{1}{2k_1k_2}(h_{11}k_1^2 + h_{22}k_2^2 + h_{11}k_3^2 + h_{22}k_3^2)$$

$$
\begin{aligned}
h_{13} &= -\frac{1}{2k_1k_3}(h_{11}k_1^2 - h_{22}k_2^2 - h_{11}k_3^2 - h_{22}k_3^2) \\
h_{23} &= -\frac{1}{2k_2k_3}(-h_{11}k_1^2 + h_{22}k_2^2 - h_{11}k_3^2 - h_{22}k_3^2) \\
h_{33} &= -h_{11} - h_{22} \ ,
\end{aligned}
\tag{7.32}
$$

and show that the second (trace) part vanishes.

The above manipulations underscore the fact that the lattice action, in the weak field limit and for small momenta, only propagates transverse-traceless modes, as for linearized gravity in the continuum. They can be used to derive an expression for the lattice analog of the result given in (Kuchař, 1970) and (Hartle, 1984) for the vacuum wave functional of linearized gravity, which gives therefore a suitable starting point for a lattice candidate for the same functional.

A cosmological constant term can be analyzed in the lattice weak field expansion along similar lines. According to Eqs. (6.41) or (6.42) it is given on the lattice by the total space-time volume, so that the action contribution is given by

$$
I_V = \lambda_0 \sum_{\text{edges h}} V_h \ , \tag{7.33}
$$

where $V_h$ is defined to be the volume associated with an edge $h$. The latter is obtained by subdividing the volume of each four-simplex into contributions associated with each hinge (here via a barycentric subdivision), and then adding up the contributions from each four-simplex touched by the given hinge. Expanding out in the small edge fluctuations one has

$$
I_V \sim \sum_n (\varepsilon_1^{(n)} + \varepsilon_2^{(n)} + \varepsilon_4^{(n)} + \varepsilon_8^{(n)}) + \tfrac{1}{2} \sum_{mn,ij} \varepsilon_i^{(m)\,T} \, M_{i,j}^{(m,n)} \, \varepsilon_j^{(n)} \ . \tag{7.34}
$$

One needs to be careful since the expansion of $\varepsilon_i$ in terms of $h_{\mu\nu}$ contains terms quadratic in $h_{\mu\nu}$, so that there are additional diagonal contributions to the small fluctuation matrix $L_\omega$,

$$
\varepsilon_1 + \varepsilon_2 + \varepsilon_4 + \varepsilon_8 = \tfrac{1}{2}(h_{11} + h_{22} + h_{33} + h_{44}) - \tfrac{1}{8}(h_{11}^2 + h_{22}^2 + h_{33}^2 + h_{44}^2) + \cdots \ . \tag{7.35}
$$

These additional contributions are required for the volume term to reduce to the continuum form of Eq. (1.55) for small momenta and to quadratic order in the weak field expansion.

Next the same set of rotations needs to be performed as for the Einstein term, in order to go from the lattice variables $\varepsilon_i$ to the continuum variables $\bar{h}_{\mu\nu}$. After the combined $S_\omega$- and $T_\omega$-matrix rotations of Eqs. (7.10) and (7.16) one obtains for the small fluctuation matrix $L_\omega$ arising from the gauge-fixed lattice Einstein-Regge term [see Eq. (7.12)]

$$
L_\omega = -\tfrac{1}{2} k^2 V, \tag{7.36}
$$

with the matrix $V$ given by Eq. (7.21). Since the lattice cosmological term can also be expressed in terms of the matrix $V$,

$$\sqrt{g} = 1 + \tfrac{1}{2} h_{\mu\mu} - \tfrac{1}{2} h_{\alpha\beta} V^{\alpha\beta\mu\nu} h_{\mu\nu} + O(h^3) \ , \tag{7.37}$$

one finds, as in the continuum, for the combined Einstein and cosmological constant terms

$$\lambda_0 \, (1 + \tfrac{1}{2} \, h_{\mu\mu}) + \tfrac{1}{2} \cdot \frac{k}{2} \, h_{\alpha\beta} V^{\alpha\beta\mu\nu} \, (\partial^2 + \frac{2\lambda_0}{k}) \, h_{\mu\nu} + O(h^3) \ , \tag{7.38}$$

corresponding in this gauge to the exchange of a particle of "mass" $\mu^2 = -2\lambda_0/k$, in agreement with the continuum weak field result of Eq. (1.79). As for the Regge-Einstein term, there are higher order lattice corrections to the cosmological constant term of $O(k)$ (which are completely absent in the continuum, since no derivatives are present there). These should be irrelevant in the lattice continuum limit.

## 7.3 Lattice Diffeomorphism Invariance

The appearance of exact zero modes in the weak field expansion of the lattice gravitational action is not specific to an expansion about flat space. One can consider the same procedure for variations about spaces which are classical solutions for the gravitational action with a cosmological constant term as in Eq. (6.43), such as the regular or irregular tessellations of the $d$-sphere. In principle it is possible to look at a general $d$-dimensional case, but here, for illustrative purposes, only two dimensions will be considered, in which case on is looking, in the simplest case, at the regular tessellations of the two-sphere. In the following the discussion will focus therefore at first on edge length fluctuations about the regular tetrahedron (with 6 edges), octahedron (12 edges), and icosahedron (30 edges). One could consider irregular tessellations as well, but this will not be pursued here, although one believes the results to have general validity for lattices sufficiently dense with points (Hartle, 1985; Hamber and Williams, 1997).

In two dimensions the lattice action for pure gravity is

$$I(l^2) \ = \ \lambda \sum_h V_h \, - \, 2k \sum_h \delta_h \, + \, 4a \sum_h \frac{\delta_h^2}{A_h} \ , \tag{7.39}$$

with the two-dimensional volume element equated here to the dual area surrounding a vertex $V_h = A_h$, and the local curvature given by $R_h = 2 \ \delta_h / A_h$. In the limit of small fluctuations around a smooth background, the above lattice action describes the continuum action

$$I[g] \ = \ \int d^2x \sqrt{g} \left[ \lambda \, - \, kR + aR^2 \right] \ . \tag{7.40}$$

For a manifold of fixed topology the term proportional to $k$ can be dropped, since $\sum_h \delta_h = 2\pi\chi$, where $\chi$ is the Euler characteristic. The classical solutions have constant curvature with $R = \pm\sqrt{\lambda/a}$ (there being no real solutions for $\lambda < 0$). The curvature-squared leads to some non-trivial interactions in two dimensions, although the resulting theory is not unitary. This is not important here, as we only plan to address for now the issue of lattice diffeomorphism invariance.

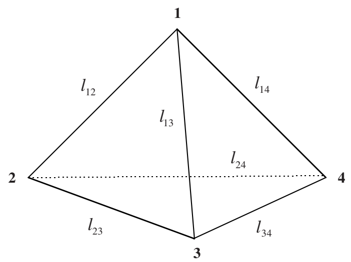



**Fig. 7.2** Tetrahedral tessellation of the two-sphere, with arbitrary edge length assignments.

After expanding about the equilateral configuration, the action at the stationary point reduces to

$$I \;=\; \lambda\; 8\pi\sqrt{a/\lambda} + a\; 8\pi/\sqrt{a/\lambda} \;=\; 16\pi\sqrt{a\,\lambda} \quad , \tag{7.41}$$

independently of the tessellation considered. Vanishing of the linear terms in the small fluctuation expansion gives for the average edge length

$$l_0 \;=\; \left[c\,\pi^2\,(4a/\lambda)\right]^{1/4} \quad , \tag{7.42}$$

with $c = 16/3, 4/3, 16/75$ for the tetrahedron, octahedron and icosahedron, respectively. For fluctuations about the classical solution for a tetrahedral tessellation of $S^2$ [see Fig. (7.2)] the small edge length fluctuation matrix gives rise to the following coefficients

$$\begin{aligned}
\varepsilon_{12}^2 &\to 16\sqrt{a\lambda}\;(54 - 6\sqrt{3}\pi + 5\pi^2)/81\pi \\
\varepsilon_{12}\,\varepsilon_{13} &\to 16\sqrt{a\lambda}\;\pi/9 \\
\varepsilon_{12}\,\varepsilon_{15} &\to 64\sqrt{a\lambda}\;(-27 + 3\sqrt{3}\pi + 2\pi^2)/81\pi \;\; ,
\end{aligned} \tag{7.43}$$

with the remaining coefficients being determined by symmetry. The small fluctuation matrix is therefore given by

$$\frac{8\pi\sqrt{a\lambda}}{9}\begin{pmatrix} \mu & 1 & 1 & 1 & 1 & 2-\mu \\ 1 & \mu & 1 & 2-\mu & 1 & 1 \\ 1 & 1 & \mu & 1 & 2-\mu & 1 \\ 1 & 2-\mu & 1 & \mu & 1 & 1 \\ 1 & 1 & 2-\mu & 1 & \mu & 1 \\ 2-\mu & 1 & 1 & 1 & 1 & \mu \end{pmatrix}, \tag{7.44}$$

where $\mu = 2(5\pi^2 - 6\sqrt{3}\pi + 54)/9\pi^2 \approx 1.5919$. (the $\lambda/a$ dependence has disappeared since the couplings $a$ and $\lambda$ only appear in the dimensionless combination $\sqrt{a\lambda}$). The eigenvalues of the above matrix (apart from the constants in front of it) are 0 (with multiplicity 2), $2(\mu-1)$ (with multiplicity 3) and 6 (with multiplicity 1). The zero modes correspond to flat directions, for which deformations of the edge lengths leave the lattice geometry unchanged. Their explicit form in the weak field limit was given in Eq. (7.9).

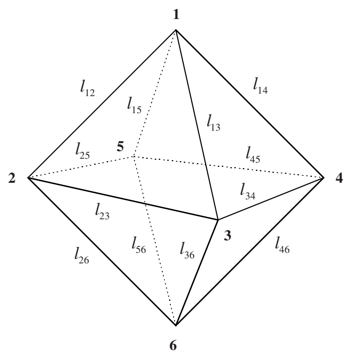



**Fig. 7.3** Octahedral tessellation of the two-sphere, with arbitrary edge length assignments.

For the octahedron [see Fig. (7.3)] one obtains instead the following coefficients of the small fluctuation matrix

$$\begin{aligned} \varepsilon_{12}^2 &\to 2\sqrt{a\lambda}\ (216 - 12\sqrt{3}\pi + 5\pi^2)/27\pi \\ \varepsilon_{12}\,\varepsilon_{13} &\to 8\sqrt{a\lambda}\ (-27 - 3\sqrt{3}\pi + 2\pi^2)/27\pi \\ \varepsilon_{12}\,\varepsilon_{14} &\to 4\sqrt{a\lambda}\ (54 + \pi^2)/9\pi \\ \varepsilon_{12}\,\varepsilon_{34} &\to 8\sqrt{a\lambda}\ (-54 + 3\sqrt{3}\pi + \pi^2)/27\pi \\ \varepsilon_{12}\,\varepsilon_{46} &\to 4\sqrt{a\lambda}\ (108 + 12\sqrt{3}\pi + \pi^2)/27\pi\ , \end{aligned} \tag{7.45}$$

again with the remaining coefficients being determined by symmetry. Up to a common factor of $2\sqrt{a\lambda}/27\pi$, the eigenvalues of the $12 \times 12$ small fluctuation matrix are given by $36\pi^2$ (with multiplicity 1), 972 (with multiplicity 2), and $8(3\sqrt{3}-\pi)^2$ (with multiplicity 3), and zero (with multiplicity 6).

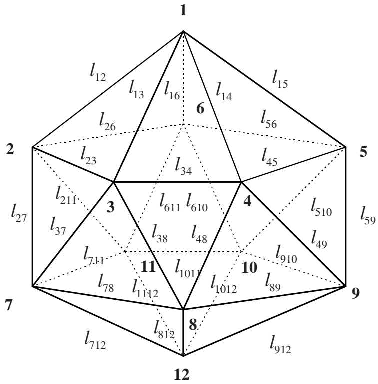



**Fig. 7.4** Icosahedral tessellation of the two-sphere, with arbitrary edge length assignments.

Finally, for the icosahedron [shown in Fig. (7.4)] one computes the following coefficients of the small fluctuation matrix

$$
\begin{aligned}
\varepsilon_{12}^2 &\to 16\sqrt{a\lambda}\,(270 - 6\sqrt{3}\pi + \pi^2)/135\pi \\
\varepsilon_{12}\,\varepsilon_{13} &\to 16\sqrt{a\lambda}\,(-675 - 30\sqrt{3}\pi + 8\pi^2)/675\pi \\
\varepsilon_{12}\,\varepsilon_{14} &\to 16\sqrt{a\lambda}\,(270 - 6\sqrt{3} + \pi^2)/135\pi \\
\varepsilon_{12}\,\varepsilon_{34} &\to 32\sqrt{a\lambda}\,(-675 + 15\sqrt{3}\pi + 2\pi^2)/675\pi \\
\varepsilon_{12}\,\varepsilon_{45} &\to 16\sqrt{a\lambda}\,(-675 + 15\sqrt{3}\pi + 2\pi^2)/675\pi \\
\varepsilon_{12}\,\varepsilon_{38} &\to 16\sqrt{a\lambda}\,(-675 + 15\sqrt{3}\pi + 2\pi^2)/675\pi \\
\varepsilon_{12}\,\varepsilon_{48} &\to 16\sqrt{a\lambda}\,(675 + 30\sqrt{3}\pi + \pi^2)/675\pi \ ,
\end{aligned}
\tag{7.46}
$$

with the remaining coefficients being determined by symmetry. Up to a common factor of $8\sqrt{a\lambda}/675\pi$, the eigenvalues of the $30 \times 30$ small edge length fluctuation matrix are given by 12340.173 (with multiplicity 3), 7238.984 (with multiplicity 5), $888.264 = 90\pi^2$ (with multiplicity 1), 20.887 (with multiplicity 3), and zero (with multiplicity 18).

The presence of the zero modes is interpreted as a lattice manifestation of the diffeomorphism invariance of the gravitational action. One can summarize the previous results so far as

$$
\begin{aligned}
\text{Tetrahedron } (N_0 = 4):&\ \ 2 \text{ zero modes} \\
\text{Octahedron}(N_0 = 6):&\ \ 6 \text{ zero modes} \\
\text{Icosahedron}(N_0 = 12):&\ \ 18 \text{ zero modes} \ .
\end{aligned}
\tag{7.47}
$$

If the number of zero modes for each triangulation of the sphere is denoted by $N_{z.m.}$, then the results can be re-expressed as

$$N_{z.m.} = 2N_0 - 6 , \tag{7.48}$$

which agrees with the expectation that in the continuum limit, $N_0 \to \infty$, $N_{z.m.}/N_0$ should approach the constant value $d$ in $d$ space-time dimensions, which is the number of local parameters for a diffeomorphism. On the lattice the diffeomorphisms correspond to local deformations of the edge lengths about a vertex, which leave the local geometry physically unchanged, the latter being described by the values of local lattice operators corresponding to local volumes, and curvatures. The lesson is that the correct count of zero modes will in general only be recovered asymptotically for sufficiently large triangulations, where $N_0$ is roughly much larger than the number of neighbors to a point in $d$ dimensions. A similar pattern is expected in higher dimensions, although in general one would expect such results to hold only for deformations of flat space which are not too large. In particular one should always keep in mind the presence of the triangle inequalities, which do not allow deformations of the edges past a certain configuration space boundary.

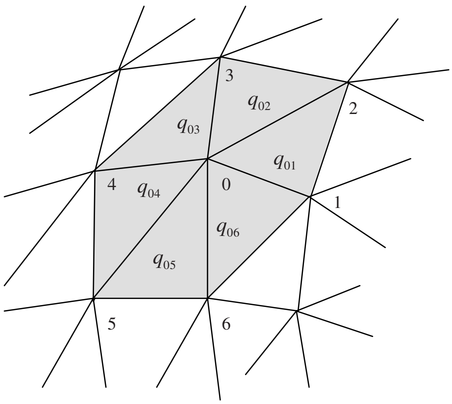



**Fig. 7.5** Notation for an arbitrary simplicial lattice, where the edge lengths meeting at the vertex 0 have been deformed away from a regular lattice by a small amount $q_i$ (minimally deformed equilateral lattice).

The previous discussion dealt with the expansion of the gravitational action about a regular lattice: a regular tessellation of the sphere, a manifold of constant curvature. One might wonder whether the results depend on the lattice having a particular symmetry, but this can be shown not to be the case. To complete our discussion, we turn therefore to the slightly more complex task of exhibiting explicitly the local lattice invariance for an arbitrary background simplicial complex. The idea here is to look at lattices that are deformations of a regular lattice, and small edge fluctuations around them. To this end we write for the edge length deformations

$$l_i^2 \;=\; l_{0i}^2 + q_i + \delta l_i^2 \ , \tag{7.49}$$

where $q_i$ describes an arbitrary but small deviation from a regular lattice, and $\delta l_i^2$ is a gauge fluctuation, whose form needs to be determined. We shall keep terms $O(q^2)$ and $O(q\,\delta l^2)$, but neglect terms $O(\delta l^4)$.

The squared volumes $V_n^2(\sigma)$ of n-dimensional simplices $\sigma$ are given by homogeneous polynomials of order $(l^2)^n$. In particular for the area of a triangle $A_\Delta$ with arbitrary edges $l_1, l_2, l_3$ one has

$$\delta A_\Delta^2 \;=\; \tfrac{1}{8}(-l_1^2+l_2^2+l_3^2)\,\delta l_1^2 + \tfrac{1}{8}(l_1^2-l_2^2+l_3^2)\,\delta l_2^2 + \tfrac{1}{8}(l_1^2+l_2^2-l_3^2)\,\delta l_3^2 \ , \tag{7.50}$$

and similarly for the other quantities which are needed in order to construct the action. For our notation in two dimensions we refer to Fig. (7.5). The subsequent Figs. 7.6 and 7.7 illustrate the difference between a *gauge* deformation of the surface, which leaves the area and curvature at the point labeled by 0 invariant, and a *physical* deformation which corresponds to a re-assignment of edge lengths meeting at the vertex 0 such that it alters the area and curvature at 0. In the following we will characterize unambiguously what we mean by the two different operations.

Consider therefore an expansion about a deformed equilateral lattice, for which $l_{0i}=1$ to start with. A motivation for this choice is provided by the fact that in the numerical studies of two-dimensional gravity the averages of the squared edge lengths in the three principal directions turn out to be equal, $\langle l_1^2\rangle = \langle l_2^2\rangle = \langle l_3^2\rangle$. The baricentric area associated with vertex 0 is then given by

$$\begin{aligned} A = A_0(q) + \frac{1}{2\cdot 3^{5/2}} \Big[ & \delta l_{01}^2 \ (3+q_{06}-4q_{01}+q_{02}+q_{16}+q_{12}) \\ + & \delta l_{02}^2 \ (3+q_{01}-4q_{02}+q_{03}+q_{12}+q_{23}) \\ + & \delta l_{03}^2 \ (3+q_{02}-4q_{03}+q_{04}+q_{23}+q_{34}) \\ + & \delta l_{04}^2 \ (3+q_{03}-4q_{04}+q_{05}+q_{34}+q_{45}) \\ + & \delta l_{05}^2 \ (3+q_{04}-4q_{05}+q_{06}+q_{45}+q_{56}) \\ + & \delta l_{06}^2 \ (3+q_{05}-4q_{06}+q_{01}+q_{56}+q_{16}) \Big] \\ + & \qquad O(\delta l^4) \ . \end{aligned} \tag{7.51}$$

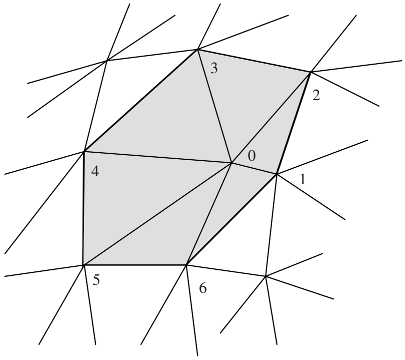
The normalization here is such that $A_0 = \frac{\sqrt{3}}{2}$ for $q_i = 0$. Equivalently one can write, in more compact notation, at the vertex 0

$$A = A_0(q) + \tfrac{1}{3}\mathbf{v}_A(q)\cdot \boldsymbol{\delta}\mathbf{l^2} + O(\delta l^4) \ , \tag{7.52}$$

with $\boldsymbol{\delta}\mathbf{l^2} = (\delta l_{01}^2, \ldots, \delta l_{06}^2)$. After adding the contributions from the neighboring vertices one obtains

$$\sum_{P_0 \ldots P_6} A \;=\; \sum_{P_0 \ldots P_6} A_0(q) + \mathbf{v}_A(q)\cdot \boldsymbol{\delta}\mathbf{l^2} + O(\delta l^4) \ . \tag{7.53}$$

**Fig. 7.6** Local gauge deformations of the lattice act on the edge lengths meeting at the vertex 0, and are performed in such a way that the area and curvature at the vertex 0 are left unchanged.




Therefore the area associated with the vertex 0 will remain unchanged provided the variations in the squared edge lengths meeting at 0 satisfy the constraint

$$\mathbf{v}_A(q) \cdot \delta \mathbf{l^2} \; = \; 0 \; . \tag{7.54}$$

This is nothing but the curved space equivalent of the well known flat equilateral lattice condition for deformations to be of the pure gauge form, given in Eq. (7.9), form

$$\sum_{i=1}^{6} \delta l_i^2(n) \; = \; 0 \; . \tag{7.55}$$

Furthermore, if one considers a dual lattice subdivision, a similar result can be shown to hold.

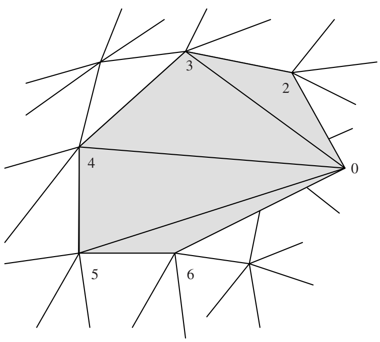



**Fig. 7.7** Physical deformations change the area and curvature at the vertex 0, thus changing the lattice geometry.

A similar calculation can be done for the curvature associated with vertex 0. One has for the deficit angle at 0

$$\begin{aligned}\delta = \delta_0(q) + \frac{1}{3^{3/2}} \Big[ \; & \delta l_{01}^2 \; (3 - 2q_{06} - q_{01} - 2q_{02} + q_{16} + q_{12}) \\ + \; & \delta l_{02}^2 \; (3 - 2q_{01} - q_{02} - 2q_{03} + q_{12} + q_{23}) \\ + \; & \cdots \; \Big] + O(\delta l^4) \; , \end{aligned} \tag{7.56}$$

and therefore for the variation of the sum of the deficit angles surrounding 0

$$\Delta \left( \sum_h \delta_h \right) = \sum_{P_0 \dots P_6} \Delta \delta \tag{7.57}$$

and

$$\sum_{P_0 \dots P_6} \delta = \sum_{P_0 \dots P_6} \delta_0(q) + \mathbf{v}_R(q) \cdot \delta \mathbf{l^2} + O(\delta l^4) \tag{7.58}$$

with in this case, as expected, $\mathbf{v}_R(q) \equiv 0$.

Finally for the curvature squared associated with vertex 0 one computes

$$\begin{aligned}\frac{\delta^2}{A} = \frac{\delta_0^2}{A_0}(q) + \frac{4}{3^{3/2}} \Big[ & \delta l_{01}^2 \big( \; q_{01} + q_{02} + q_{03} + q_{04} + q_{05} + q_{06} \\ & \quad - q_{12} - q_{23} - q_{34} - q_{45} - q_{56} - q_{16} \big) \\ & + \delta l_{02}^2 \big( \; q_{01} + q_{02} + q_{03} + q_{04} + q_{05} + q_{06} \\ & \quad - q_{12} - q_{23} - q_{34} - q_{45} - q_{56} - q_{16} \big) + \cdots \Big] + O(\delta l^4) \; . \end{aligned} \tag{7.59}$$

Adding up all seven contributions one gets

$$\Delta \left( \sum_h \delta_h^2 / A_h \right) = \sum_{P_0 \dots P_6} \Delta (\delta^2 / A) \; , \tag{7.60}$$

and therefore

$$\sum_{P_0 \dots P_6} \delta^2 / A = \sum_{P_0 \dots P_6} \left( \delta^2 / A \right)_0 + \mathbf{v}_{R^2}(q) \cdot \delta \mathbf{l^2} + O(\delta l^4) \; . \tag{7.61}$$

In this case the curvature squared associated with the vertex 0 will remain unchanged, provided the variations in the squared edge lengths meeting at 0 satisfy the constraint

$$\mathbf{v}_{R^2}(q) \cdot \delta \mathbf{l^2} = 0 \; , \tag{7.62}$$

which provides a second constraint on the edge length variations $\delta \mathbf{l^2}$ at the vertex 0. Again this constraint is the generalization to curved space of the condition of Eq. (7.55) for flat space.

A similar calculation can be performed for square lattices with $l_{01} = l_{02} = 1$ and $l_{03} = \sqrt{2}$, and one obtains analogous results. In general one can show explicitly how gauge variations of the edge lengths at each vertex can be defined by requiring that the lattice action contributions be locally invariant (Hamber and Williams, 1999). Here we have looked at small deformations, but larger deformations can be treated along the same lines, provided one is careful not to violate the triangle inequalities, which impose a sharp non-perturbative cutoff in orbit space. Finally the same approach can be extended to higher dimensions, leading to similar (but rather more complicated, when written out explicitly) results. The main conclusions do not change, namely that there is a local $d$-parameter invariance of the lattice action for small deformations of the edges, analogous to the local diffeomorphism invariance of the continuum theory.

## 7.4 Strong Coupling Expansion

In this section the strong coupling [large $G$ or small $k = 1/(8\pi G)$] expansion of the lattice gravitational functional integral will be discussed. The resulting series is expected to be useful up to some $k = k_c$, where $k_c$ is the lattice critical point, at which the partition function develops a singularity.

There will be two main aspects to the following discussion. The first aspect will be the development of a systematic expansion for the partition function and the correlation functions in powers of $k$, and a number of rather general considerations that follow from it. The second main aspect will be a detailed analysis and interpretation of the individual terms which appear order by order in the strong coupling expansion. This second part will lead to a later discussion of what happens for large $d$.

One starts from the lattice regularized path integral with action Eq. (6.43) and measure Eq. (6.76). In the following we will focus at first on the four-dimensional case. Then the four-dimensional Euclidean lattice action contains the usual cosmological constant and Regge scalar curvature terms of Eq. (6.90)

$$I_{latt} \;=\; \lambda \sum_h V_h(l^2) - k \sum_h \delta_h(l^2) A_h(l^2) \;, \tag{7.63}$$

with $k = 1/(8\pi G)$, and possibly additional higher derivative terms as well. The action only couples edges which belong either to the same simplex or to a set of neighboring simplices, and can therefore be considered as *local*, just like the continuum action. It leads to a lattice partition function defined in Eq. (6.91)

$$Z_{latt} \;=\; \int [d\, l^2] \; e^{-\lambda_0 \sum_h V_h + k \sum_h \delta_h A_h} \;, \tag{7.64}$$

where, as customary, the lattice ultraviolet cutoff is set equal to one (i.e. all length scales are measured in units of the lattice cutoff). For definiteness the measure will be of the form

$$\int [d\,l^2] \;=\; \int_0^\infty \prod_s \, [V_d(s)]^\sigma \, \prod_{ij} dl_{ij}^2 \, \Theta[l_{ij}^2] \;\; . \tag{7.65}$$

The lattice partition function $Z_{latt}$ should be compared to the continuum Euclidean Feynman path integral of Eq. (2.34),

$$Z_{cont} \;=\; \int [d\,g_{\mu\nu}] \; e^{-\lambda \int dx \sqrt{g} + \frac{1}{16\pi G} \int dx \sqrt{g} R} \;\; . \tag{7.66}$$

When doing an expansion in the kinetic term proportional to $k$, it will be convenient to include the $\lambda$-term in the measure. We will set therefore in this Section as in Eq. (6.93)

$$d\mu(l^2) \;\equiv\; [d\,l^2] \, e^{-\lambda_0 \sum_h V_h} \;\; . \tag{7.67}$$

It should be clear that this last expression represents a fairly non-trivial quantity, both in view of the relative complexity of the expression for the volume of a simplex, Eq. (6.5), and because of the generalized triangle inequality constraints already implicit in $[d\,l^2]$. But, like the continuum functional measure, it is certainly *local*, to the extent that each edge length appears only in the expression for the volume of those simplices which explicitly contain it. Also, we note that in general the integral $\int d\mu$ can only be evaluated numerically; nevertheless this can be done, at least in principle, to arbitrary precision. Furthermore, $\lambda_0$ sets the overall scale and can therefore be set equal to one without any loss of generality.

Thus the effective strong coupling measure of Eq. (7.67) has the properties that (a) it is local in the lattice metric of Eq. (6.3), to the same extent that the continuum measure is ultra-local, (b) it restricts all edge lengths to be positive, and (c) it imposes a soft cutoff on large simplices due to the $\lambda_0$-term and the generalized triangle inequalities. Apart from these constraints, it does *not* significantly restrict the fluctuations in the lattice metric field at short distances. It will be the effect of the curvature term to restrict such fluctuation, by coupling the metric field between simplices, in the same way as the derivatives appearing in the continuum Einstein term couple the metric between infinitesimally close space-time points.

As a next step, $Z_{latt}$ is expanded in powers of $k$,

$$Z_{latt}(k) \;=\; \int d\mu(l^2) \; e^{k \sum_h \delta_h A_h} \;=\; \sum_{n=0}^{\infty} \frac{1}{n!} k^n \int d\mu(l^2) \left( \sum_h \delta_h A_h \right)^n \;\; . \tag{7.68}$$

It is easy to show that $Z(k) = \sum_{n=0}^{\infty} a_n \, k^n$ is analytic at $k = 0$, so this expansion should be well defined up to the nearest singularity in the complex $k$ plane. A quantitative estimate for the expected location of such a singularity in the large-$d$ limit will be given later in Sect. 7.6. Beyond this singularity $Z(k)$ can sometimes be extended, for example, via Padé or differential approximants.[1] The above expansion is of course analogous to the high temperature expansion in statistical mechanics systems, where the on-site terms are treated exactly and the kinetic or hopping term is treated as a perturbation. Singularities in the free energy or its derivatives can usually be pinned down with the knowledge of a large enough number of terms in the relevant expansion (Domb and Green, 1976).

Next consider a fixed, arbitrary hinge on the lattice, and call the corresponding curvature term in the action $\delta A$. Such a contribution will be denoted in the following, as is customary in lattice gauge theories, a *plaquette* contribution. For the average curvature on that hinge one has

$$< \delta A > = \frac{\sum_{n=0}^{\infty} \frac{1}{n!} k^n \int d\mu(l^2)\, \delta A \left( \sum_h \delta_h A_h \right)^n}{\sum_{n=0}^{\infty} \frac{1}{n!} k^n \int d\mu(l^2) \left( \sum_h \delta_h A_h \right)^n} \quad . \tag{7.69}$$

After expanding out in $k$ the resulting expression, one obtains for the cumulants

$$< \delta A > = \sum_{n=0}^{\infty} c_n k^n \tag{7.70}$$

with

$$c_0 = \frac{\int d\mu(l^2)\, \delta A}{\int d\mu(l^2)} \quad , \tag{7.71}$$

whereas to first order in $k$ one has

$$c_1 = \frac{\int d\mu(l^2)\, \delta A \left( \sum_h \delta_h A_h \right)}{\int d\mu(l^2)} - \frac{\int d\mu(l^2)\, \delta A \cdot \int d\mu(l^2) \sum_h \delta_h A_h}{\left( \int d\mu(l^2) \right)^2} \quad . \tag{7.72}$$

This last expression clearly represents a measure of the fluctuation in $\delta A$, namely $[\langle(\sum_h \delta_h A_h)^2\rangle - \langle\sum_h \delta_h A_h\rangle^2]/N_h$, using the homogeneity properties of the lattice $\delta A \to \sum_h \delta_h A_h / N_h$. Here $N_h$ is the number of hinges in the lattice. Equivalently, it can be written in an even more compact way as $N_h[\langle(\delta A)^2\rangle - \langle\delta A\rangle^2]$.

To second order in $k$ one has

$$c_2 = N_h^2 \, [\langle(\delta A)^3\rangle - 3\langle\delta A\rangle\langle(\delta A)^2\rangle + 2\langle\delta A\rangle^3]/2 \quad . \tag{7.73}$$

[1] A first order transition cannot affect the singularity structure of $Z(k)$ as viewed from the strong coupling phase, as the free energy is $C_\infty$ at a first order transition.

At the next order one has

$$c_3 = N_h^3 \, [\langle (\delta A)^4 \rangle - 4\langle \delta A \rangle \langle (\delta A)^3 \rangle - 3\langle (\delta A)^2 \rangle^2 + 12\langle (\delta A)^2 \rangle \langle \delta A \rangle^2 - 6\langle \delta A \rangle^4]/6 \ , \tag{7.74}$$

and so on. Note that the expressions in square parentheses become rapidly quite small, $O(1/N_h^n)$ with increasing order $n$, as a result of large cancellations that must arise eventually between individual terms inside the square parentheses. In principle, a careful and systematic numerical evaluation of the above integrals (which is quite feasible in practice) would allow the determination of the expansion coefficients in $k$ for the average curvature $< \delta A >$ to rather high order.

As an example, consider a non-analyticity in the average scalar curvature

$$\mathscr{R}(k) \ = \ \frac{< \int dx \sqrt{g(x)} \, R(x) >}{< \int dx \sqrt{g(x)} >} \approx \frac{< \sum_h \delta_h A_h >}{< \sum_h V_h >} \ , \tag{7.75}$$

assumed for concreteness to be of the form of an algebraic singularity at $k_c$, namely

$$\mathscr{R}(k) \underset{k \to k_c}{\sim} A_{\mathscr{R}} \, (k_c - k)^{\delta} \ , \tag{7.76}$$

with $\delta$ some exponent. It will lead to a behavior, for the general term in the series in $k$, of the type

$$(-1)^n A_{\mathscr{R}} \, \frac{(\delta - n + 1)(\delta - n + 2) \ldots \delta}{n! \, k_c^{n-\delta}} \, k^n \ . \tag{7.77}$$

Given enough terms in the series, the singularity structure can then be investigated using a variety of increasingly sophisticated series analysis methods.

It can be advantageous to isolate in the above expressions the *local* fluctuation term, from those terms that involve *correlations* between different hinges. To see this, one needs to go back, for example, to the first order expression in Eq. (7.72) and isolate in the sum $\sum_h$ the contribution which contains the selected hinge with value $\delta A$, namely

$$\sum_h \delta_h A_h \ = \ \delta A + {\sum_h}' \, \delta_h A_h \ , \tag{7.78}$$

where the primed sum indicates that the term containing $\delta A$ is *not* included. The result is

$$\begin{aligned} c_1 \ = \ & \frac{\int d\mu(l^2) \, (\delta A)^2}{\int d\mu(l^2)} - \frac{\left( \int d\mu(l^2) \, \delta A \right)^2}{\left( \int d\mu(l^2) \right)^2} \\ & + \frac{\int d\mu(l^2) \, \delta A \sum_h{}' \, \delta_h A_h}{\int d\mu(l^2)} - \frac{\left( \int d\mu(l^2) \, \delta A \right) \left( \int d\mu(l^2) \sum_h{}' \, \delta_h A_h \right)}{\left( \int d\mu(l^2) \right)^2} \ . \end{aligned} \tag{7.79}$$

One then observes the following: the first two terms describe the *local* fluctuation of $\delta A$ on a given hinge; the third and fourth terms describe *correlations* between $\delta A$ terms on *different* hinges. But because the action is *local*, the only non-vanishing contribution to the last two terms comes from edges and hinges which are in the immediate vicinity of the hinge in question. For hinges located further apart (indicated below by "*not nn*") one has that their fluctuations remain uncorrelated, leading to a vanishing variance

$$\frac{\int d\mu(l^2)\,\delta A \sum_{h\,\mathrm{not}\,nn}{}'\,\delta_h A_h}{\int d\mu(l^2)} - \frac{\left(\int d\mu(l^2)\,\delta A\right)\left(\int d\mu(l^2) \sum_{h\,\mathrm{not}\,nn}{}'\,\delta_h A_h\right)}{\left(\int d\mu(l^2)\right)^2} = 0 \ , \tag{7.80}$$

since for uncorrelated random variables $X_n$'s, $< X_n X_m > - < X_n >< X_m >= 0$. Therefore the only non-vanishing contributions in the last two terms in Eq. (7.79) come from hinges which are *close* to each other.

The above discussion makes it clear that a key quantity is the *correlation* between different plaquettes,

$$< (\delta A)_h\,(\delta A)_{h'} > = \frac{\int d\mu(l^2)\,(\delta A)_h\,(\delta A)_{h'}\,e^{k\sum_h \delta_h A_h}}{\int d\mu(l^2)\,e^{k\sum_h \delta_h A_h}} \ , \tag{7.81}$$

or, better, its *connected* part (denoted here by $< \ldots >_C$)

$$< (\delta A)_h\,(\delta A)_{h'} >_C \ \equiv \ < (\delta A)_h\,(\delta A)_{h'} > \ - \ < (\delta A)_h >< (\delta A)_{h'} > \ , \tag{7.82}$$

which subtracts out the trivial part of the correlation. Here again the exponentials in the numerator and denominator can be expanded out in powers of $k$, as in Eq. (7.69). The lowest order term in $k$ will involve the correlation

$$\int d\mu(l^2)\,(\delta A)_h\,(\delta A)_{h'} \ . \tag{7.83}$$

But unless the two hinges are close to each other, they will fluctuate in an uncorrelated manner, with $< (\delta A)_h\,(\delta A)_{h'} > - < (\delta A)_h >< (\delta A)_{h'} >= 0$. In order to achieve a non-trivial correlation, the path between the two hinges $h$ and $h'$ needs to be tiled by at least as many terms from the product $(\sum_h \delta_h A_h)^n$ in

$$\int d\mu(l^2)\,(\delta A)_h\,(\delta A)_{h'} \left(\sum_h \delta_h A_h\right)^n \ , \tag{7.84}$$

as are needed to cover the distance $l$ between the two hinges. One then has

$$< (\delta A)_h\,(\delta A)_{h'} >_C \sim \ k^l \ \sim \ e^{-l/\xi} \ , \tag{7.85}$$

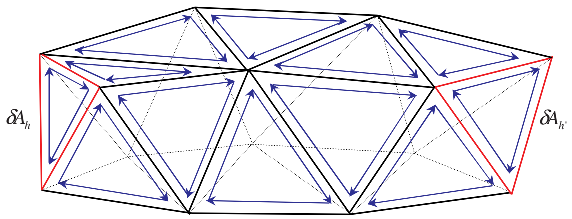



**Fig. 7.8** Correlations between action contributions on hinge $h$ and hinge $h'$ arise to lowest order in the strong coupling expansions from diagrams describing a narrow tube connecting the two hinges. Here vertices represent points in the dual lattice, with the tube-like closed surface tiled with parallel transport polygons. For each link of the dual lattice, the $SO(4)$ parallel transport matrices $\mathbf{R}$ of Sect. 6.4 are represented by an arrow.

with the correlation length $\xi = 1/|\log k| \to 0$ to lowest order as $k \to 0$ (here we have used the usual definition of the correlation length $\xi$, namely that a generic correlation function is expected to decay as $\exp(-\text{distance}/\xi)$ for large separations).[2] This last result is quite general, and holds for example irrespective of the boundary conditions (unless of course $\xi \sim L$, where $L$ is the linear size of the system, in which case a path can be found which wraps around the lattice).

But further thought reveals that the above result is in fact not completely correct, due to the fact that in order to achieve a non-vanishing correlation one needs, at least to lowest order, to connect the two hinges by a narrow tube (Hamber and Williams, 2006). The previous result should then read correctly as

$$< (\delta A)_h \, (\delta A)_{h'} >_C \sim \; (k^{n_d})^l \quad , \tag{7.86}$$

where $n_d \, l$ represents the minimal number of dual lattice polygons needed to form a closed surface connecting the hinges $h$ and $h'$, with $l$ the actual distance (in lattice units) between the two hinges. Fig. 7.8 provides an illustration of the situation.

With some additional effort many additional terms can be computed in the strong coupling expansion. In practice the method is generally not really competitive with direct numerical evaluation of the path integral via Monte Carlo methods. But it does provide a new way of looking at the functional integral, and provide the basis for new approaches, such as the large $d$ limit to be discussed in the second half of the next section.

---

[2] This statement, taken literally, oversimplifies the situation a bit, as depending on the spin (or tensor structure) of the operator appearing in the correlation function, the large distance decay of the corresponding correlator is determined by the lightest excitation in that specific channel. But in the gravitational context one is mostly concerned with correlators involving spin two (transverse-traceless) objects, evaluated at fixed geodesic distance.

## 7.5 Gravitational Wilson Loop

An important question for any theory of quantum gravity is what gravitational observables should look like, i.e. which expectation values of operators (or ratios thereof) have meaning and physical interpretation in the context of a manifestly covariant formulation, in particular in a situation where metric fluctuations are not necessarily bounded. Such averages naturally include expectation values of the (integrated) scalar curvature and other related quantities (involving for example curvature-squared terms), as well as correlations of operators at fixed geodesic distance. Another set of physical observables corresponds to the gravitational analog of the Wilson loop (Modanese, 1995), which provides information about the parallel transport of vectors, and therefore on the effective curvature, around large, near-planar loops. In contrast to gauge theories, the Wilson loop in quantum gravity does not provide useful information on the static potential, which is obtained instead for the correlation between particle world-lines (Modanese, 1995; Hamber and Williams, 1995) Here we will concentrate on defining and exploring physical properties of the gravitational Wilson loop at strong coupling (Hamber and Williams, 2007).

Before embarking on the gravitational case, it might be useful to recall more generally the well-known fact (see for example Peskin and Schröder, 1995) that many low energy physical properties in gauge theories cannot be computed reliably in weak coupling perturbation theory. Thus, for example, in non-Abelian $SU(N)$ gauge theories a confining potential for static sources placed in the fundamental representation is found at sufficiently strong coupling, by examining the behavior of the Wilson loop (Wilson, 1974), defined for a large closed planar loop $C$ as

$$W(C) = < \mathrm{tr}\, \mathscr{P} \exp\left\{ ig \oint_C A_\mu(x) dx^\mu \right\} > \ , \tag{7.87}$$

with $A_\mu \equiv t_a A^a_\mu$ and the $t_a$'s the group generators of $SU(N)$ in the fundamental representation. Specifically, in the pure gauge theory at strong coupling the leading contribution to the Wilson loop can be shown to follow an area law for sufficiently large loops

$$W(C) \underset{A\to\infty}{\sim} \exp(-A(C)/\xi^2) \ , \tag{7.88}$$

where $A(C)$ is the minimal area spanned by the planar loop $C$ (Balian Drouffe and Itzykson, 1975). The quantity $\xi$ is the gauge field correlation length, defined for example from the exponential decay of the Euclidean correlation function of two infinitesimal loops separated by a distance $|x|$,

$$G_\square(x) = < \mathrm{tr}\, \mathscr{P} \exp\left\{ ig \oint_{C'_\varepsilon} A_\mu(x') dx'^\mu \right\}(x) \ \mathrm{tr}\, \mathscr{P} \exp\left\{ ig \oint_{C''_\varepsilon} A_\mu(x'') dx''^\mu \right\}(0) >_c \ . \tag{7.89}$$

Here the $C_\varepsilon$'s are two infinitesimal loops centered around $x$ and 0 respectively, suitably defined on the lattice as elementary square loops, and for which one has at sufficiently large separations

$$G_{\Box}(x) \underset{|x|\to\infty}{\sim} \exp(-|x|/\xi) \ . \tag{7.90}$$

The inverse of the correlation length $\xi$ is known to correspond to the lowest mass excitation in the gauge theory, the scalar glueball.

The gauge theory definition can be adapted to the lattice gravitational case. It will turn out that it is most easily achieved by using a slight variant of Regge calculus, in which the action coincides with the usual Regge action in the near-flat limit. Here we will use extensively the notion of lattice parallel transport discussed in Sect. 6.4, and how areas are defined on the dual lattice. From those properties the behavior of large Wilson loops can be derived, and much of what is done is in close parallel with the analogous procedure in lattice gauge theories at strong coupling. One finds that the results for the lattice gravitational loop imply that for sufficiently strong coupling (large bare Newton's constant $G$) the behavior of the loop at large distances is consistent with a positive vacuum curvature, and therefore with (Euclidean) De Sitter space.

The construction of the lattice action and measure, leading to a discretized form for the gravitational functional integral, are discussed in Sects. (6.4), (6.5) and (6.9). In the following it will be convenient to include at strong coupling the cosmological constant term in the measure, since this contribution is ultralocal and contains no derivatives of the metric, giving rise to an effective strong coupling measure $d\mu(l^2)$,

$$d\mu(l^2) \ \equiv \ [d\,l^2]\, e^{-\lambda_0 \sum_h V_h} \ . \tag{7.91}$$

In view of previous discussions, this last expression represents a fairly non-trivial quantity, both in view of the relative complexity of the expression for the volume of a simplex, and because of the generalized triangle inequality constraints already implicit in the definition of $[d\,l^2]$.

The main assumption used here regarding the effective strong coupling measure $d\mu(l^2)$ will be the existence of a stable ground state with a well-defined average lattice spacing, as implied by direct numerical evaluations of the lattice integrals in four dimensions, at least for sufficiently strong coupling. In the following the lattice measure $[d\,l^2]$ will be a suitable discretization of the continuum functional measure, and therefore of the form

$$\int [d\,l^2] \ = \ \int_0^\infty \prod_s \, [V_d(s)]^\sigma \, \prod_{ij} dl_{ij}^2 \, \Theta[l_{ij}^2] \ , \tag{7.92}$$

with $\sigma$ again a real parameter, and $\Theta$ a function of the squared edge lengths ensuring that the generalized triangle inequalities discussed in Sect. (6.9) are satisfied.

At strong coupling the measure and cosmological constant terms form the dominant part of the functional integral, since the Einstein part of the action is vanishingly small in this limit. Yet, and in contrast to strongly coupled lattice Yang-Mills theories, the functional integral is still non-trivial to compute analytically in this limit, mainly due to these triangle inequality constraints. Therefore, in order to be able to derive some analytical estimates for correlation functions in the strong coupling limit, one needs still to develop some set of approximation methods. In principle the reliability of the approximations can later be tested by numerical means, for example by integrating directly over edges using the explicit lattice measure given above.

One approach that appears natural in the gravity context follows along the lines of what is normally done in gauge theories, namely an integration over compact group variables, using the invariant measure over the gauge group. It is of this method that we wish to take advantage here, as we believe that it is well suited for gravity as well. In order to apply such a technique to gravity one needs (i) to formulate the lattice theory in such a way that group variables are separated and therefore appear explicitly; (ii) integrate over the group variables using an invariant measure; and (iii) approximate the relevant correlation functions in such a way that the group integration can be performed exactly, using for example mean field methods for the parts that appear less tractable. In such a program one is aided by the fact that in the strong coupling limit one is expanding about a well defined ground state, and that the measure and the interactions are *local*, coupling only lattice variable (edges or rotations) which are a few lattice spacings apart. The downside of such methods is that one is no longer evaluating the functional integral for quantum gravity exactly, even in the strong coupling limit; the upside is that one obtains a clear analytical estimate, which later can be in principle systematically tested by numerical methods (which are in principle exact).

In the gravity case the analogs of the gauge variables of Yang-Mills theories are given by the connection, so it is natural therefore to look for a first order formulation of Regge gravity (Caselle D'Adda and Magnea, 1989), discussed in Sect. (4.2). The main feature of this approach is that one treats the metric $g_{\mu\nu}$ and the affine connection $\Gamma^{\lambda}_{\mu\nu}$ as independent variables. There one can safely consider functionally integrating separately over the affine connection and the metric, treated as independent variables, with the correct relationship between metric and connection arising then as a consequence of the dynamics. In the lattice theory we will follow a similar spirit, separating out explicitly in the lattice action the degrees of freedom corresponding to local rotations (the analogs of the $\Gamma$'s in the continuum), which we will find to be most conveniently described by orthogonal matrices $\mathbf{R}$.

The next step is a use of the properties of local rotation matrices in the context of the lattice theory, and how these relate to the lattice gravitational action. It was shown in Sect. (6.4) that with each neighboring pair of simplices $s, s+1$ one can associate a Lorentz transformation $R^{\mu}{}_{\nu}(s, s+1)$, which describes how a given vector $V^{\mu}$ transforms between the local coordinate systems in these two simplices, and that the above transformation is directly related to the continuum path-ordered ($P$) exponential of the integral of the local affine connection $\Gamma^{\lambda}_{\mu\nu}(x)$ via

$$R^{\mu}{}_{\nu} = \left[ P \, e^{\int_{\textbf{between simplices}}^{\textbf{path}} \Gamma_{\lambda} dx^{\lambda}} \right]^{\mu}{}_{\nu} , \tag{7.93}$$

with the connection having support only on the common interface between the two simplices. Also, for a closed elementary path $C_h$ encircling a hinge $h$ and passing

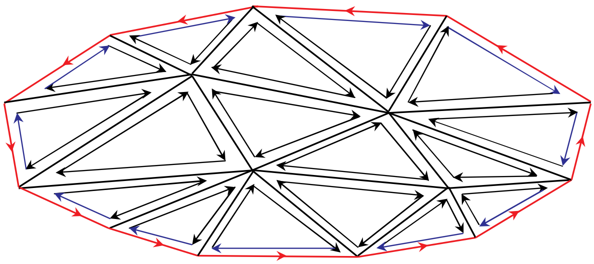


**Fig. 7.9** Gravitational analog of the Wilson loop. A vector is parallel-transported along the larger outer loop. The enclosed minimal surface is tiled with parallel transport polygons, here chosen to be triangles for illustrative purposes. For each link of the dual lattice, the elementary parallel transport matrices $\mathbf{R}(s,s')$ are represented by arrows. In spite of the fact that the (Lorentz) matrices $\mathbf{R}$ can fluctuate strongly in accordance with the local geometry, two contiguous, oppositely oriented arrows always give $\mathbf{R}\mathbf{R}^{-1} = 1$.

through each of the simplices that meet at that hinge one has for the total rotation matrix $\mathbf{R} \equiv \prod_s R_{s,s+1}$ associated with the given hinge

$$\Big[\prod_s R_{s,s+1}\Big]^{\mu}_{\ \nu} = \Big[e^{\delta(h)U(h)}\Big]^{\mu}_{\ \nu} \ , \tag{7.94}$$

as in Eq. (6.32). This matrix describes the parallel transport of a vector round the loop.

More generally one might want to consider a near-planar, but non-infinitesimal, closed loop $C$, as shown in Fig. (7.9). Along this closed loop the overall rotation matrix will still be given by

$$R^{\mu}_{\ \nu}(C) = \Big[\prod_{s\subset C} R_{s,s+1}\Big]^{\mu}_{\ \nu} \ . \tag{7.95}$$

In analogy with the infinitesimal loop case, one would like to state that for the overall rotation matrix one has

$$R^{\mu}_{\ \nu}(C) \approx \Big[e^{\delta(C)U(C))}\Big]^{\mu}_{\ \nu} \ , \tag{7.96}$$

where $U_{\mu\nu}(C)$ is now an area bivector perpendicular to the loop, which will work only if the loop is close to planar so that $U_{\mu\nu}$ can be taken to be approximately constant along the path $C$. By a near-planar loop around the point $P$, we mean one that is constructed by drawing outgoing geodesics, on a plane through $P$.

If that is true, then one can define an appropriate coordinate scalar by contracting the above rotation matrix $\mathbf{R}(C)$ with the some appropriate bivector, namely

$$W(C) = \omega_{\alpha\beta}(C)\, R^{\alpha\beta}(C) \ , \tag{7.97}$$

where the bivector, $\omega_{\alpha\beta}(C)$, is intended as being representative of the overall geometric features of the loop (for example, it can be taken as an average of the hinge bivector $\omega_{\alpha\beta}(h)$ along the loop).

In the quantum theory one is interested in the average of the above loop operator $W(C)$, as in Eq. (7.87). The previous construction is indeed quite analogous to the Wilson loop definition in ordinary lattice gauge theories, where it is defined via the trace of path ordered products of $SU(N)$ color rotation matrices. In gravity though the Wilson loop does not give any information about the static potential; instead it provides some insight into the large-scale curvature of the manifold, just as the infinitesimal loop contribution entering the lattice action of Eqs. (6.90) and (6.39) provides, through its averages, information on the very short distance, local curvature.

One the other hand for any continuum manifold one can define locally the parallel transport of a vector around a near-planar loop $C$. Indeed parallel transporting a vector around a closed loop represents a suitable operational way of detecting curvature locally. If the curvature of the manifold is small, one can treat the larger loop the same way as the small one; then the expression of Eq. (7.96) for the rotation matrix $\mathbf{R}(C)$ associated with a near-planar loop can be re-written in terms of a surface integral of the large-scale Riemann tensor, projected along the surface area element bivector $A^{\alpha\beta}(C)$ associated with the loop,

$$R^{\mu}{}_{\nu}(C) \approx \left[ e^{\frac{1}{2}\int_S R^{\cdot}{}_{\cdot\alpha\beta} A^{\alpha\beta}(C)} \right]^{\mu}{}_{\nu} . \tag{7.98}$$

Thus a direct calculation of the Wilson loop provides a way of determining the *effective* curvature at large distance scales, even in the case where short distance fluctuations in the metric may be significant. Conversely, the rotation matrix appearing in the elementary Wilson loop of Eqs. (6.29), (6.32) and (7.94) only provides information about the parallel transport of vectors around *infinitesimal* loops, with size comparable to the ultraviolet cutoff.

Let us now look in detail at how to construct a Wilson loop in quantum gravity. Since this involves finding the expectation value of a product of rotation matrices round a loop, the natural procedure is to treat these rotation matrices as variables and to integrate over their product, weighted by the exponential of minus the Regge action. The expression for this action has been given in terms of functions of the edge lengths, but an alternative (Fröhlich, 1981; Caselle D' Adda and Magnea, 1989) is to find an expression for it in terms of the rotation matrices. For the dual loop around each hinge, the product of the rotation matrices gives the exponential of the deficit angle, $\delta$, times the rotation generator, $U$, [see Eq. (6.32)] and one needs to find a way of extracting the deficit angle from this product of matrices, at the same time as constructing a scalar function to be averaged. The obvious way of doing this is to contract the product of the $R$-matrices with the rotation generator, $U$, and then take the trace. This is equivalent to the action obtained by contracting the elementary rotation matrix $\mathbf{R}(C)$ of Eq. (6.32), with the hinge bivector of Eq. (6.30), as done in Eq. (6.39),

$$I_{\rm com}(l^2) \;=\; -\,\frac{k}{2} \sum_{\rm hinges\;h} A_h \, U_{\alpha\beta}(h) \, R^{\alpha\beta}(h) \;\; . \tag{7.99}$$

The above construction can be regarded as analogous to Wilson's lattice gauge theory, for which the action also involves traces of products of $SU(N)$ color rotation matrices. This contraction produces the sine of the deficit angle times the area of the triangular hinge and so for small deficit angles it is equivalent to the Regge action.

But one notices that when evaluating Wilson loops, the final result often involves the trace of the bivector $U$, which is zero. It will be useful therefore to contract with a linear combination of $U$ and the unit matrix, and use as a pure gravity action

$$I_h \;=\; \frac{k}{4} \, A_h \, Tr[(U_h \,+\, \varepsilon \, I_4) \, (\mathbf{R}_h \,-\, \mathbf{R}_h^{-1})] \;\; , \tag{7.100}$$

where $I_4$ is the unit matrix in four dimensions, and $\varepsilon$ an arbitrary real parameter. We have subtracted the inverse of the rotation matrix for the hinge for reasons that will become apparent when one evaluates Wilson loops. In the end one is interested in the limit of small but non-zero $\varepsilon$. One might worry that the $\varepsilon$ term might modify the weak field limit, but this is not the case. Since $\mathbf{R}$ equals the exponential of $\delta$ times $U$, it may be expanded in a power series in $\delta$, which is then contracted with the $(U+\varepsilon I_4)$ and the trace taken. Using $Tr(U^{2n+1})=0$ and $Tr(U^{2n}) \;=\; 2\,(-1)^n$ one finds

$$I_h \;=\; -\,k\,A_h \; \sin(\delta_h) \;\; , \tag{7.101}$$

independently of the value of the parameter $\varepsilon$. Thus $\varepsilon$ is in fact an arbitrary parameter, which can be conveniently taken to be non-zero, as we shall see.

There is now a slight amount of freedom in how we define the Wilson loop, for a path $C$ in the dual lattice of a simplicial space. The main choices seem to be either

$$(i) \;\;\; W(C) \;=<\; Tr(R_1 \, R_2 \, ... \, R_n) \;>\; ; \tag{7.102}$$

or

$$(ii) \;\;\; W(C) \;=<\; Tr[(U_C \,+\, \varepsilon \, I_4) \, R_1 \, R_2 \, ... \, ... \, R_n] \;>\;\; . \tag{7.103}$$

Here the $R_i$ are the rotation matrices along the path; in (ii), there is a factor of $(U_C+\varepsilon I_4)$, containing some "average" direction bivector, $U_C$, for the loop, which, after all, is assumed to be almost planar. The position of the $U_C$ term in the product of $R_i$'s is not arbitrary; to give a unique answer, it needs to be placed before an $R$ which begins one of the plaquette contributions to the action.

One would like to take as independent fluctuating variables the rotation matrices $R_i$ and the loop bivectors $U_i$, in a first order formalism similar in spirit to that used in (Caselle, D'Adda and Magnea, 1989). This last statement clearly requires some clarification, as both the rotation matrices and the loop bivectors depend on the choice of the original edge lengths, as well as on the orientation of the local coordinate system, and cannot therefore in general be considered as independent variables (as should have already been clear from the detailed discussion of the properties of rotation matrices given in the previous section). On the lattice strong edge length fluctuations get reflected in large fluctuations in the local geometry, which in turn imply large correlated fluctuations in both the deficit angles and in the orientations of the elementary loop. It would therefore seem at first that one would have to integrate over both sets of coupled variables simultaneously, with some non-trivial measure derived from the original lattice measure over edge lengths, which in turn would make the problem of computing the Wilson loop close to intractable, even in the strong coupling limit. In particular one has to take notice of the fact that the lattice deficit angles and the loop bivectors are related to the metric and connection, as they appear in a first order formulation, in a rather non-trivial way.

But there are two important aspect that come into play when evaluating the expectation value of the gravitational Wilson loop for strongly coupled gravity, the first one being that the overall geometric features of the large near-planar loop provide a natural orientation, specified for example by a global loop bivector $U_C$. As will become clear from explicit calculations given below, in the strong coupling limit the tiling of the large Wilson loop surface by elementary parallel transport loops, which in general have random orientations, *requires* that their normals be preferentially oriented perpendicular to the plane of the loop, since otherwise a non-minimal surface must result, which leads to a necessarily higher order contribution in the strong coupling limit.

In the case of a hinge surrounded by the large loop with bivector $U_C$, one is therefore allowed to write for the bivector operator $U_h$ associated with that hinge, labelled by $h$,

$$U_h \;=\; U_C + \delta U_h \;\;, \tag{7.104}$$

where $\delta U_h$ is the quantum fluctuation associated with hinge bivector at $h$. But assuming the fluctuation in $\delta U_h$ to be zero is an unnecessarily strong requirement, and in the following it will be sufficient to take $<\delta U_h>= 0$ and $<(\delta U_h)^2>\neq 0$, which can be regarded as a mean-field type treatment for the loop bivectors. It will be important therefore in the following to keep in mind this distinction between the fluctutating hinge bivector $U_h$, and its quantum average.

The second important aspect of the calculation is that at strong coupling the edge lengths, and therefore the local geometry, fluctuate in a way that is uncorrelated over distances greater than a few lattice spacing. Thus, mainly due to the ultralocal nature of the gravitational lattice measure at strong coupling, the fluctuations in the $U's$ can be taken as essentially uncorrelated as well, again over distances greater than a few lattice spacings, which further simplifies the problem considerably.

One would expect that for a geometry fluctuating strongly at short distances (corresponding therefore to the small $k$ limit) the infinitesimal parallel transport matrices $R(s,s')$ should be distributed close to randomly, with a measure close to the uniform Haar measure, and with little correlation between neighboring hinges. In such instance one would have for the local quantum averages of the infinitesimal lattice parallel transports $<R>= 0$, but $<R\,R^{-1}>\neq 0$, which would require, for a non-vanishing lowest order contribution to the Wilson loop, that the loop at least be tiled by elementary loops with action contributions from Eqs. (6.90) or (7.99), thus forming a minimal surface spanning the loop. Then, in close analogy to the Yang-Mills case of Eq. (7.88), the leading contribution to the gravitational Wilson loop would be expected to follow an area law,

$$W(C) \sim \text{const.}\, k^{A(C)} \sim \exp(-A(C)/\xi^2) \ , \tag{7.105}$$

where $A(C)$ is the minimal physical area spanned by the near-planar loop $C$, and $\xi$ the gravitational correlation length, equal to $\xi = 1/\sqrt{|\ln k|}$ for small $k$. For a close-to-circular loop of perimeter $P$ one would use $A(C) \approx P^2/4\pi$.

We choose now to focus on the Euclidean case in four dimensions, where the rotation matrices will be elements of $SO(4)$. In evaluating the averages over the rotation matrices in the expectation values in the Wilson loops, the integrations we have to perform will be of the form

$$\int \left( \prod_{i=1}^{n} d\mu_H(R_i) \right) \; Tr\,[...(U_j + \varepsilon I_4)...R_k...] \\ \times \; \exp\left( -\frac{k}{4} \sum_{\text{hinges}\,\text{h}} A_h\, Tr[(U_h + \varepsilon I_4)\,(\mathbf{R}_h - \mathbf{R}_h^{-1})] \right) / \mathscr{N} , \tag{7.106}$$

where the normalization factor is given by

$$\mathscr{N} = \int \left( \prod_{i=1}^{n} d\mu_H(R_i) \right) \exp\left( -\frac{k}{4} \sum_{\text{hinges}\,\text{h}} A_h\, Tr[(U_h + \varepsilon I_4)\,(\mathbf{R}_h - \mathbf{R}_h^{-1})] \right) \ . \tag{7.107}$$

This factor will be omitted from subsequent expressions, for notational simplicity.

For smooth enough geometries, with small curvatures, the rotation matrices can be chosen to be close to the identity. Small fluctuations in the geometry will then imply small deviations in the $R$'s from the identity matrix. However, for strong coupling ($k \to 0$) the usual lattice measure $\int d\mu(l^2)$ does not significantly restrict fluctuations in the lattice metric field. As a result we will assume that these fields can be regarded, at least in this regime, as basically unconstrained random variables, only subject to the relatively mild constraints implicit in the measure $d\mu(l^2)$. Thus as $k \to 0$, the geometry is generally far from smooth, since there is no coupling term to enforce long range order (the coefficient of the lattice Einstein term goes to zero), and one has as a consequence large local fluctuations in the geometry. The matrices **R** will therefore fluctuate with the local geometry, and average out to zero, or a value close to zero, which suggests the use of the Haar measure over the group variables, as in ordinary $SU(N)$ lattice gauge theories. The uniform (Haar) measure over the group $SO(n)$ is given by

$$d\mu_H(R) = \left( \prod_{i=1}^{n} \Gamma(i/2)/2^n \, \pi^{n(n+1)/2} \right) \prod_{i=1}^{n-1} \prod_{j=1}^{i} \sin^{j-1}\theta^i_{\ j}\, d\theta^i_{\ j} \ , \tag{7.108}$$

with $0 \leq \theta_k^1 < 2\pi$, $0 \leq \theta_k^j < \pi$ (see for example Vilenkin and Klimyk, 1993). In practice one does not need to explicitly integrate over $SO(4)$ angles. Instead one uses the following properties of the normalized Haar measure,

$$\int d_H R \; = \; 1 \; ; \tag{7.109}$$

$$\int d_H R \;\; Tr(A\, R) \; Tr(R^{-1}\, B) \; = \; \frac{1}{4} \, Tr(A\, B) \;\; , \tag{7.110}$$

for arbitrary $4 \times 4$ matrices $A$ and $B$, which also implies

$$\int d_H R \;\; R_{ij} \, R_{kl}^{-1} \; = \; \frac{1}{4} \, \delta_{il} \, \delta_{jk} \;\; . \tag{7.111}$$

As stated previously, we will regard the individual hinge bivectors $U_h$ as aligned on average with the Wilson loop bivector $U_C$.

One is now ready to evaluate $< W >$ for some simple loops. Later we will discuss the general behavior for arbitrary loops, ending with a consideration of the asymptotic behavior for large loops. In the following only the second definition $(ii)$ of the Wilson loop $< W >$ in Eq. (7.103) will be considered; when one works out the details for the other choice one finds very similar results, and in the end the main conclusions are unchanged.

Consider a single hinge of area $A$, at which four 4-simplices meet see Fig. (7.10). The loop $C$ will consist of four segments between the Voronoi centers of the simplices. Let the rotation matrices on these segments be $R_1, R_2, R_3, R_4$, and the rotation generator for the hinge $U$. Then one has

$$W(C) \; = < \; Tr[(U \; + \; \varepsilon \, I_4) \, R_1 \, R_2 \, R_3 \, R_4] \;\; > \;\; . \tag{7.112}$$

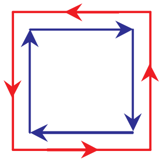


**Fig. 7.10** A parallel transport loop with four oriented links on the boundary. The parallel transport matrices **R** along the links, represented here by arrows, appear in pairs and are sequentially integrated over using the uniform measure.

Since the only non-vanishing contribution to the integration over the $R$'s will come from the product of an $R_i$ with the corresponding $R_i^{-1}$, then the lowest order contribution in $k$ will come from the term in the expansion of the exponential of minus the action which is linear in $\mathbf{R}^{-1}$. One obtains

$$\begin{aligned} \frac{k}{4} A \int d_H R_1 \, d_H R_2 \, d_H R_3 \, d_H R_4 \quad & Tr[(U \; + \; \varepsilon \, I_4) \, R_1 \, R_2 \, R_3 \, R_4] \\ & \times \, Tr[(U \; + \; \varepsilon \, I_4) \, R_4^{-1} \, R_3^{-1} \, R_2^{-1} \, R_1^{-1}] \;\; . \end{aligned} \tag{7.113}$$

Integration over the $R$'s results in

$$\frac{k}{4}\,\frac{1}{4}\,A\,Tr(U^2 + \varepsilon^2 I_4) = -\frac{k}{8}\,A\,(1 - 2\,\varepsilon^2)\,, \tag{7.114}$$

and the integration over the $U$'s (a sum over all possible orientations of the loop) is trivial.

Next consider a less trivial situation where the Wilson loop goes around a number of hinges and there is at least one internal hinge, i.e. a hinge where the elementary loop surrounding it is not part of the Wilson loop. For simplicity, we shall consider the case of one such loop. For the labeling of the rotation matrices and the hinges, the reader can annotate Fig. (7.11) in a way consistent with the expressions below.

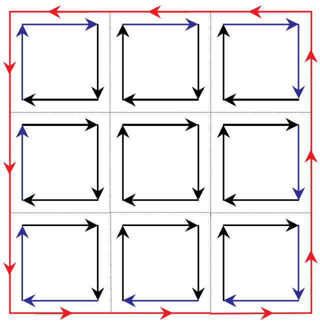


**Fig. 7.11** A larger parallel transport loop with twelve oriented links on the boundary. As before, the parallel transport matrices along the links appear in pairs and are sequentially integrated over using the uniform measure. The new ingredient in this configuration is an elementary loop at the center not touching the boundary.

In this case, the lowest order contribution comes from a ninth-order term in the expansion of the exponential of the action. One obtains the following result

$$\frac{k^9}{4^9}\,\frac{1}{4^{17}}\left(\prod_{i=1}^{9} A_i\right) Tr[(U_C+\varepsilon I_4)\,(U_1+\varepsilon I_4)] \left(\prod_{i=2}^{9} Tr(U_i\ +\varepsilon I_4)\right)$$
$$= \frac{k^9}{4^{18}}\left(\prod_{i=1}^{9} A_i\right) \varepsilon^8\,[Tr(U_C U_1)+4\varepsilon^2]\ . \tag{7.115}$$

In this last equation one sets $U_1 = U_C + \delta U_1$, with $< \delta U_1 >= 0$, after which the sum over the loop's orientation also becomes trivial. The above result also shows that it is better to take $\varepsilon > 0$, otherwise the answer vanishes to this order. But this is not a problem, as the correct lattice action is recovered irrespective of the value of $\varepsilon$, as shown earlier in Eq. (7.101).

Finally the value of a Wilson loop, when the loop is very large and surrounds $n$ hinges, can be seen to be of the general form

$$\frac{k^n}{4^{2n}} \left( \prod_{i=1}^{n} A_i \right) \varepsilon^{\alpha} \, [p + q \varepsilon^2]^{\beta} \quad , \tag{7.116}$$

where $\alpha + \beta = n$. If $\bar{A}$ is of the order of the geometric or arithmetic mean of the individual loops, this can be approximated by

$$\left( \frac{k\bar{A}}{16} \right)^n \varepsilon^{\alpha} \, [p + q \varepsilon^2]^{\beta} \quad . \tag{7.117}$$

This shows again that one should take $\varepsilon > 0$, otherwise the answer vanishes to this order, and one needs to go to higher order in the expansion in $k$. This is quite legitimate, as the correct lattice action is recovered irrespective of the value of $\varepsilon$, as in Eq. (7.101). Then using $n = A_C/\bar{A}$, one can write the area-dependent first factor as

$$\exp[\, (A_C/\bar{A}) \, \log(k\bar{A}/16) \,] = \exp(-A_C/\xi^2) \quad , \tag{7.118}$$

where $\xi \equiv [\bar{A}/|\log(k\bar{A}/16)|]^{1/2}$. Recall that this is in the case of strong coupling, when $k \to 0$. The above is the main result so far. The rapid decay of the quantum gravitational Wilson loop as a function of the area is seen here simply as a general and direct consequence of the assumed disorder in the uncorrelated fluctuations of the parallel transport matrices $\mathbf{R}(s,s')$ at strong coupling.

We note here the important point that the gravitational correlation length $\xi$ is defined independently of the expectation value of the Wilson loop. Indeed a key quantity in gauge theories as well as gravity is the *correlation* between different plaquettes, which in simplicial gravity is given by see Eq. (7.81)

$$< (\delta A)_h \, (\delta A)_{h'} > = \frac{\displaystyle\int d\mu(l^2) \, (\delta A)_h \, (\delta A)_{h'} \, e^{k \sum_h \delta_h A_h}}{\displaystyle\int d\mu(l^2) \, e^{k \sum_h \delta_h A_h}} \quad . \tag{7.119}$$

In order to achieve a non-vanishing correlation one needs, at least to lowest order, to connect the two hinges by a narrow tube, so that

$$< (\delta A)_h \, (\delta A)_{h'} >_C \sim \; (k^{n_t})^l \sim \; e^{-d(h,h')/\xi} \quad , \tag{7.120}$$

where the “distance” $n_t \, l$ represents the minimal number of dual lattice polygons needed to form a closed surface connecting the hinges $h$ and $h'$ (as an example, for a narrow tube made out of cubes connecting two squares one has $n_t$=4). In the above expression $d(h,h')$ represents the actual physical distance between the two hinges, and the correlation length is given in this limit ($k \to 0$) by $\xi \sim l_0/n_t |\log k|$. where $l_0$ is the average lattice spacing. Here we have used the usual definition of the correlation length $\xi$, namely that a generic correlation function is expected to decay as $\exp(-\text{distance}/\xi)$ for large separations. Fig. (7.8) provides an illustration of the situation.

The strong coupling area law behavior predicted for a large Wilson loop in Eq. (7.118) should be compared with the results for this in numerical simulations of lattice gravity. For small deficit angles (small curvature), the action used here [involving Eq. (7.100)] is sufficiently close to the usual Regge action of Eq. (6.90) that the standard simulations can be used for comparison. Universality arguments would suggest a similar behavior for the gravitational Wilson loop for a wide class of lattice actions, constructed so as to reproduce the Einstein-Hilbert continuum action in the continuum limit.

The final step is an interpretation of this last and main result in semi classical terms. As discussed at the beginning of this section, the rotation matrix appearing in the gravitational Wilson loop can be related classically to a well-defined physical process: a vector is parallel transported around a large loop, and at the end it is compared to its original orientation. The vector's rotation is then directly related to some sort of average curvature enclosed by the loop. The total rotation matrix $\mathbf{R}(C)$ is given in general by a path-ordered ($\mathscr{P}$) exponential of the integral of the affine connection $\Gamma^{\lambda}_{\mu\nu}$ via

$$R^{\alpha}{}_{\beta}(C) = \left[ \mathscr{P} \exp\left\{ \oint_{\textbf{path C}} \Gamma^{\cdot}_{\lambda\,\cdot} dx^{\lambda} \right\} \right]^{\alpha}_{\;\beta} . \tag{7.121}$$

In such a semi classical description of the parallel transport process of a vector around a very large loop, one can re-express the connection in terms of a suitable coarse-grained, or semi-classical, Riemann tensor, using Stokes' theorem

$$R^{\alpha}{}_{\beta}(C) \sim \left[ \exp\left\{ \tfrac{1}{2} \int_{S(C)} R^{\cdot}{}_{\cdot\mu\nu} A^{\mu\nu}_{C} \right\} \right]^{\alpha}_{\;\beta} , \tag{7.122}$$

where here $A^{\mu\nu}_{C}$ is the usual area bivector associated with the loop in question,

$$A^{\mu\nu}_{C} = \tfrac{1}{2} \oint dx^{\mu}\, x^{\nu} . \tag{7.123}$$

The use of semi-classical arguments in relating the above rotation matrix $\mathbf{R}(C)$ to the surface integral of the Riemann tensor assumes (as usual in the classical context) that the curvature is slowly varying on the scale of the very large loop. Since the rotation is small for weak curvatures, one can write

$$R^{\alpha}{}_{\beta}(C) \sim \left[ 1 + \tfrac{1}{2} \int_{S(C)} R^{\cdot}{}_{\cdot\mu\nu} A^{\mu\nu}_{C} + \ldots \right]^{\alpha}_{\;\beta} . \tag{7.124}$$

At this stage one is ready to compare the above expression to the quantum result of Eq. (7.118), and in particular one should relate the coefficients of the area terms, which leads to the identification of the magnitude of the large scale semi-classical curvature with the genuinely quantum quantity $1/\xi^2$. Since one expression [Eq. (7.124)] is a matrix and the other [Eq. (7.118)] is a scalar, we shall take the trace after first contracting the rotation matrix with $(U_C + \varepsilon I_4)$, as in our second definition of the Wilson loop, giving

$$W(C) \sim \mathrm{Tr}\left( (U_C + \varepsilon I_4) \exp\left\{ \tfrac{1}{2} \int_{S(C)} R^{\cdot}{}_{\cdot \mu\nu} A_C^{\mu\nu} \right\} \right) \ . \tag{7.125}$$

Next, as is standard in simplicial gravity, we consider the lattice analog of a background manifold with constant or near-constant large scale curvature,

$$\begin{aligned} R_{\mu\nu\lambda\sigma} &= \tfrac{1}{3}\lambda \,(g_{\mu\nu}\, g_{\lambda\sigma} - g_{\mu\lambda}\, g_{\nu\sigma}) \\ R_{\mu\nu\lambda\sigma} R^{\mu\nu\lambda\sigma} &= \tfrac{8}{3}\lambda^2 \ , \end{aligned} \tag{7.126}$$

so that here in our case we can set

$$R^{\alpha}{}_{\beta\mu\nu} = \bar{R}\, U^{\alpha}{}_{\beta}\, U_{\mu\nu} \ , \tag{7.127}$$

where $\bar{R}$ is some average curvature over the loop, and the $U$'s here will be taken to coincide with $U_C$. The trace of the product of $(U_C + \varepsilon I_4)$ with this expression gives

$$Tr(\bar{R}\, U_C^2\, A_C) = -2\, \bar{R}\, A_C \ , \tag{7.128}$$

where one has used $U_{\mu\nu} A_C^{\mu\nu} = 2A_C$ (the choice of direction for the bivectors will be such that the latter is true for all loops). This is to be compared with the linear term from the other exponential expression, $-A_C/\xi^2$. Thus the average curvature is computed to be of the order

$$\bar{R} \sim 1/\xi^2 \ , \tag{7.129}$$

at least in the small $k = 1/8\pi G$ limit. An equivalent way of phrasing the last result is that $1/\xi^2$ should be identified, up to a constant of proportionality, with the scaled cosmological constant $\lambda$, with the latter being regarded as a measure of the intrinsic curvature of the vacuum. We see therefore that a direct calculation of the Wilson loop for gravity provides an insight into whether the manifold is De Sitter or anti-De Sitter *at large distances*.

## 7.6 Discrete Gravity in the Large-*d* Limit

In the large-$d$ limit the geometric expressions for volume, areas and angles simplify considerably, and as will be shown below one can obtain a number of interesting results for lattice gravity. These can then be compared to earlier investigations of continuum Einstein gravity in the same limit (Strominger, 1981).

Here we will consider a general simplicial lattice in $d$ dimensions, made out of a collection of flat $d$-simplices glued together at their common faces so as to constitute a triangulation of a smooth continuum manifold, such as the $d$-torus or the surface of a sphere. Each simplex is endowed with $d+1$ vertices, and its geometry is completely specified by assigning the lengths of its $d(d+1)/2$ edges. We will label the vertices by $1,2,3,\ldots,d+1$ and denote the square edge lengths by $l_{12}^2 = l_{21}^2$, ... , $l_{1,d+1}^2$.

As discussed in Sect. 6.3, the volume of a $d$-simplex can be computed from the determinant of a $(d+2)\times(d+2)$ matrix,

$$V_d \;=\; \frac{(-1)^{\frac{d+1}{2}}}{d!\,2^{d/2}} \begin{vmatrix} 0 & 1 & 1 & \cdots \\ 1 & 0 & l_{12}^2 & \cdots \\ 1 & l_{21}^2 & 0 & \cdots \\ 1 & l_{31}^2 & l_{32}^2 & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 1 & l_{d+1,1}^2 & l_{d+1,2}^2 & \cdots \end{vmatrix}^{1/2} . \tag{7.130}$$

If one calls the above matrix $M_d$ then previous expression can the re-written as

$$V_d \;=\; \frac{(-1)^{\frac{d+1}{2}}}{d!\,2^{d/2}} \sqrt{\det M_d} \ . \tag{7.131}$$

In general the formulae for volumes and angles are quite complicated and therefore of limited use in large dimensions. The next step consists in expanding them out in terms of small edge length variations, by setting

$$l_{ij}^2 \;=\; l_{ij}^{(0)\,2} \;+\; \delta\, l_{ij}^2 \ . \tag{7.132}$$

From now on we will set $\delta\, l_{ij}^2 = \varepsilon_{ij}$. Unless stated otherwise, we will be considering the expansion about the equilateral case, and set $l_{ij}^{(0)} = 1$; later on this restriction will be relaxed. In the equilateral case one has for the volume of a simplex

$$V_d \;=\; \frac{1}{d!}\sqrt{\frac{d+1}{2^d}} \ . \tag{7.133}$$

From the well-known expansion for determinants

$$\begin{aligned} \det(1+M) \;&=\; e^{\operatorname{tr}\ln(1+M)} \\ &=\; 1 + \operatorname{tr}M + \frac{1}{2!}\left[(\operatorname{tr}M)^2 - \operatorname{tr}M^2\right] + \ldots \ . \end{aligned} \tag{7.134}$$

one finds after a little algebra

$$V_d \underset{d\to\infty}{\sim} \frac{\sqrt{d}}{d!\,2^{d/2}} \left\{1 - \tfrac{1}{2}\,\varepsilon_{12}^2 + \ldots + \tfrac{1}{d}\,(\varepsilon_{12} + \ldots + \varepsilon_{12}\,\varepsilon_{13} + \ldots) + O(d^{-2})\right\} \ . \tag{7.135}$$

Note that the terms linear in $\varepsilon$, which would have required a shift in the ground state value of $\varepsilon$ for non-vanishing cosmological constant $\lambda_0$, vanish to leading order in

$1/d$. The complete volume term $\lambda_0 \sum V_d$ appearing in the action can then be easily written down using the above expressions.

In $d$ dimensions the dihedral angle in a $d$-dimensional simplex of volume $V_d$, between faces of volume $V_{d-1}$ and $V'_{d-1}$, is obtained from Eq. (6.12)

$$\sin \theta_d = \frac{d}{d-1} \frac{V_d V_{d-2}}{V_{d-1} V'_{d-1}} \ . \tag{7.136}$$

In the equilateral case one has for the dihedral angle

$$\theta_d = \arcsin \frac{\sqrt{d^2-1}}{d} \underset{d\to\infty}{\sim} \frac{\pi}{2} - \frac{1}{d} - \frac{1}{6d^3} + \dots \tag{7.137}$$

which will require *four* simplices to meet on a hinge, to give a deficit angle of $2\pi - 4 \times \frac{\pi}{2} \approx 0$ in large dimensions. One notes that in large dimensions the simplices look locally (i.e. at a vertex) more like hypercubes. Several $d$-dimensional simplices will meet on a $(d-2)$-dimensional hinge, sharing a common face of dimension $d-1$ between adjacent simplices. Each simplex has $(d-2)(d-1)/2$ edges "on" the hinge, some more edges are then situated on the two "interfaces" between neighboring simplices meeting at the hinge, and finally one edge lies "opposite" to the hinge in question.

In the large $d$ limit one then obtains, to leading order for the dihedral angle at the hinge with vertices labelled by $1 \dots d-1$

$$\begin{aligned} \theta_d \sim_{d\to\infty} \ & \arcsin \frac{\sqrt{d^2-1}}{d} + \varepsilon_{d,d+1} + \varepsilon_{1,d}\, \varepsilon_{1,d+1} + \dots \\ & + \tfrac{1}{d} \big( -\varepsilon_{1,d} + \dots - \tfrac{1}{2} \varepsilon_{1,d}^2 + \dots - \tfrac{1}{2} \varepsilon_{d,d+1}^2 \\ & \quad - \varepsilon_{12}\, \varepsilon_{1,d+1} - \varepsilon_{1,d}\, \varepsilon_{3,d+1} - \varepsilon_{1,d}\, \varepsilon_{d,d+1} + \dots \big) \ . \\ & + O(d^{-2}) \ . \end{aligned} \tag{7.138}$$

From the expressions in Eq. (7.135) for the volume and Eq. (7.138) for the dihedral angle one can then evaluate the $d$-dimensional Euclidean lattice action, involving cosmological constant and scalar curvature terms as in Eq. (6.43)

$$I(l^2) = \lambda_0 \sum V_d - k \sum \delta_d V_{d-2} \ , \tag{7.139}$$

where $\delta_d$ is the $d$-dimensional deficit angle, $\delta_d = 2\pi - \sum_{\text{simplices}} \theta_d$. The lattice functional integral is then

$$Z(\lambda_0, k) = \int [d\, l^2] \exp\left(-I(l^2)\right) \ . \tag{7.140}$$

To evaluate the curvature term $-k \sum \delta_d V_{d-2}$ appearing in the gravitational lattice action one needs the hinge volume $V_{d-2}$, which is easily obtained from Eq. (7.135), by reducing $d \to d-2$.

We now specialize to the case where four simplices meet at a hinge. When expanded out in terms of the $\varepsilon$'s one obtains for the deficit angle

$$\begin{aligned}\delta_d &= 2\pi - 4\cdot\frac{\pi}{2} + \sum_{\text{simplices}} \tfrac{1}{d} - \varepsilon_{d,d+1} + \ldots - \varepsilon_{1,d}\,\varepsilon_{1,d+1} + \ldots \\ &\quad -\tfrac{1}{d}\left(-\varepsilon_{1,d} - \tfrac{1}{2}\varepsilon_{1,d}^2 - \tfrac{1}{2}\varepsilon_{d,d+1}^2 - \varepsilon_{12}\,\varepsilon_{1,d+1} - \varepsilon_{1,d}\,\varepsilon_{3,d+1} - \varepsilon_{1,d}\,\varepsilon_{d,d+1} + \ldots\right) \\ &\quad + O\!\left(\frac{1}{d^2}\right) .\end{aligned} \tag{7.141}$$

The action contribution involving the deficit angle is then, for a single hinge,

$$-k\,\delta_d\,V_{d-2} = (-k)\,\frac{2\,d^{3/2}\,(d-1)}{d!\,2^{d/2}}\left(-\varepsilon_{d,d+1} + \ldots - \varepsilon_{1,d}\,\varepsilon_{1,d+1} + \ldots\right) . \tag{7.142}$$

It involves two types of terms: one linear in the (single) edge opposite to the hinge, as well as a term involving a product of two distinct edges, connecting any hinge vertex to the two vertices opposite to the given hinge. Since there are four simplices meeting on one hinge, one will have 4 terms of the first type, and $4(d-1)$ terms of the second type.

To obtain the total action, a sum over all simplices, resp. hinges, has still to be performed. Dropping the irrelevant constant term and summing over edges one obtains for the total action $\lambda_0 \sum V_d - k \sum \delta_d\, V_{d-2}$ in the large $d$ limit

$$\lambda_0\left(-\tfrac{1}{2}\sum \varepsilon_{ij}^2\right) - 2kd^2\left(-\sum \varepsilon_{jk} - \sum \varepsilon_{ij}\,\varepsilon_{ik}\right) , \tag{7.143}$$

up to an overall multiplicative factor $\sqrt{d}/d!\,2^{d/2}$, which will play no essential role in the following.

The next step involves the choice of a specific lattice. Here we will evaluate the action for the cross polytope $\beta_{d+1}$. The cross polytope $\beta_n$ is the regular polytope in $n$ dimensions corresponding to the convex hull of the points formed by permuting the coordinates $(\pm 1, 0, 0, ..., 0)$, and has therefore $2n$ vertices. It is named so because its vertices are located equidistant from the origin, along the Cartesian axes in $n$-space. The cross polytope in $n$ dimensions is bounded by $2^n$ $(n-1)$-simplices, has $2n$ vertices and $2n(n-1)$ edges.

In three dimensions, it represents the convex hull of the octahedron, while in four dimensions the cross polytope is the 16-cell (Coxeter, 1948; Coxeter, 1974). In the general case it is dual to a hypercube in $n$ dimensions, with the "dual" of a regular polytope being another regular polytope having one vertex in the center of each cell of the polytope one started with. Fig. 7.12 shows as an example the polytope $\beta_8$.

When we consider the surface of the cross polytope in $d+1$ dimensions, we have an object of dimension $n-1=d$, which corresponds to a triangulated manifold with no boundary, homeomorphic to the sphere. From Eq. (7.141) the deficit angle is then given to leading order by

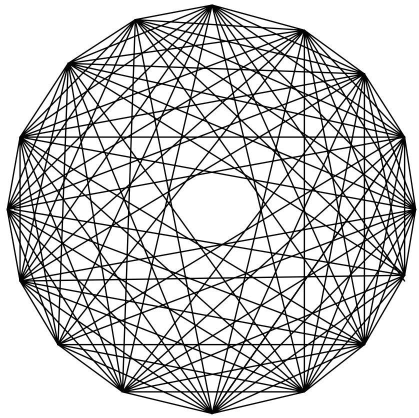


**Fig. 7.12** Cross polytope $\beta_n$ with $n = 8$ and $2n = 16$ vertices, whose surface can be used to define a simplicial manifold of dimension $d = n - 1 = 7$. For general $d$, the cross polytope $\beta_{d+1}$ will have $2(d+1)$ vertices, connected to each other by $2d(d+1)$ edges.

$$\delta_d = 0 + \frac{4}{d} - \left(\varepsilon_{d,d+1} + 3\,\text{terms} + \varepsilon_{1,d}\,\varepsilon_{1,d+1} + \ldots\right) + \ldots \tag{7.144}$$

and therefore close to flat in the large $d$ limit (due to our choice of an equilateral starting configuration). Indeed if the choice of triangulation is such that the deficit angle is not close to zero, then the discrete model leads to an average curvature whose magnitude is comparable to the lattice spacing or ultraviolet cutoff, which from a physical point of view does not seem very attractive: one obtains a space-time with curvature radius comparable to the Planck length.

When evaluated on such a manifold the lattice action becomes

$$\frac{\sqrt{d}\,2^{d/2}}{d!}\,2\left(\lambda_0 - k\,d^3\right)\left[1 - \frac{1}{8}\sum \varepsilon_{ij}^2 + \frac{1}{d}\left(\frac{1}{4}\sum \varepsilon_{ij} + \frac{1}{8}\sum \varepsilon_{ij}\,\varepsilon_{ik}\right) + O(1/d^2)\right] . \tag{7.145}$$

Dropping the $1/d$ correction the action is proportional to

$$-\tfrac{1}{2}\left(\lambda_0 - k\,d^3\right)\sum \varepsilon_{ij}^2 \ . \tag{7.146}$$

Since there are $2d(d+1)$ edges in the cross polytope, one finds therefore that, at the critical point $kd^3 = \lambda_0$, the quadratic form in $\varepsilon$, defined by the above action, develops $2d(d+1) \sim 2d^2$ zero eigenvalues.

This result is quite close to the $d^2/2$ zero eigenvalues expected in the continuum for large $d$, with the factor of four discrepancy presumably attributed to an underlying intrinsic ambiguity that arises when trying to identify lattice points with points in the continuum.

It is worth noting here that the competing curvature ($k$) and cosmological constant ($\lambda_0$) terms will have comparable magnitude when

$$k_c = \frac{\lambda_0\,l_0^2}{d^3} \ . \tag{7.147}$$

Here we have further allowed for the possibility that the average lattice spacing $l_0 = \langle l^2 \rangle^{1/2}$ is not equal to one (in other words, we have restored the appropriate overall scale for the average edge length, which is in fact largely determined by the value of $\lambda_0$).

The average lattice spacing $l_0$ can easily be estimated from the following argument. The volume of a general equilateral simplex is given by Eq. (7.133), multiplied by an additional factor of $l_0^d$. In the limit of small $k$ the average volume of a simplex is largely determined by the cosmological term, and can therefore be computed from

$$< V > = -\frac{\partial}{\partial \lambda_0} \log \int [dl^2] \, e^{-\lambda_0 V(l^2)} \quad , \tag{7.148}$$

with $V(l^2) = (\sqrt{d+1}/d!\,2^{d/2})\, l^d \equiv c_d l^d$. After doing the integral over $l^2$ with measure $dl^2$ and solving this last expression for $l_0^2$ one obtains

$$l_0^2 = \frac{1}{\lambda_0^{2/d}} \left[ \frac{2}{d} \, \frac{d!\,2^{d/2}}{\sqrt{d+1}} \right]^{2/d} \quad , \tag{7.149}$$

(which, for example, gives $l_0 = 2.153$ for $\lambda_0 = 1$ in four dimensions, in reasonable agreement with the actual value $l_0 \approx 2.43$ found near the transition point).

This then gives for $\lambda_0 = 1$ the estimate $k_c = \sqrt{3}/(16 \cdot 5^{1/4}) = 0.0724$ in $d = 4$, to be compared with $k_c = 0.0636(11)$ obtained in (Hamber, 2000) by direct numerical simulation in four dimensions. Even in $d = 3$ one finds again for $\lambda_0 = 1$, from Eqs. (7.147) and (7.149), $k_c = 2^{5/3}/27 = 0.118$, to be compared with $k_c = 0.112(5)$ obtained in (Hamber and Williams, 1993) by direct numerical simulation.

Using Eq. (7.149) inserted into Eq. (7.147) one obtains in the large $d$ limit for the dimensionless combination $k/\lambda_0^{(d-2)/d}$

$$\frac{k_c}{\lambda_0^{1-2/d}} = \frac{2^{1+2/d}}{d^3} \left[ \frac{\Gamma(d)}{\sqrt{d+1}} \right]^{2/d} \quad . \tag{7.150}$$

To summarize, an expansion in powers of $1/d$ can be developed, which relies on a combined use of the weak field expansion. It can be regarded therefore as a double expansion in $1/d$ and $\varepsilon$, valid wherever the fields are smooth enough and the geometry is close to flat, which presumably is the case in the vicinity of the lattice critical point at $k_c$.

A somewhat complementary $1/d$ expansion can be set up, which does not require weak fields, but relies instead on the strong coupling (small $k = 1/8\pi G$, or large $G$) limit. As such it will be a double expansion in $1/d$ and $k$. Its validity will be in a regime where the fields are not smooth, and in fact will involve lattice field configurations which are very far from smooth at short distances.

The general framework for the strong coupling expansion for pure quantum gravity was outlined in the previous section, and is quite analogous to what one does in gauge theories (Balian, Drouffe and Itzykson, 1975). One expands $Z_{latt}$ in powers of $k$ as in Eq. (7.68)

$$Z_{latt}(k) = \int d\mu(l^2)\; e^{k\sum_h \delta_h A_h} = \sum_{n=0}^{\infty} \frac{1}{n!} k^n \int d\mu(l^2) \left( \sum_h \delta_h A_h \right)^n . \quad (7.151)$$

Then one can show that dominant diagrams contributing to $Z_{latt}$ correspond to closed surfaces tiled with elementary transport loops. In the case of the hinge-hinge connected correlation function the leading contribution at strong coupling come from closed surfaces anchored on the two hinges, as in Eq. (7.86).

It will be advantageous to focus on general properties of the parallel transport matrices **R**, discussed previously in Sect. 6.4. For smooth enough geometries, with small curvatures, these rotation matrices can be chosen to be close to the identity. Small fluctuations in the geometry will then imply small deviations in the **R**'s from the identity matrix. But for strong coupling ($k \to 0$) the measure $\int d\mu(l^2)$ does not significantly restrict fluctuations in the lattice metric field. As a result we will assume that these fields can be regarded, in this regime, as basically unconstrained random variables, only subject to the relatively mild constraints implicit in the measure $d\mu$. The geometry is generally far from smooth since there is no coupling term to enforce long range order (the coefficient of the lattice Einstein term is zero), and one has as a consequence large local fluctuations in the geometry. The matrices **R** will therefore fluctuate with the local geometry, and average out to zero, or a value close to zero. In the sense that, for example, the $SO(4)$ rotation

$$\mathbf{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} , \quad (7.152)$$

averages out to zero when integrated over $\theta$. In general an element of $SO(n)$ is described by $n(n-1)/2$ independent parameters, which in the case at hand can be conveniently chosen as the six $SO(4)$ Euler angles. The uniform (Haar) measure over the group is then

$$d\mu_H(\mathbf{R}) = \frac{1}{32\pi^9} \int_0^{2\pi} d\theta_1 \int_0^{\pi} d\theta_2 \int_0^{\pi} d\theta_3 \int_0^{\pi} d\theta_4 \sin\theta_4 \int_0^{\pi} d\theta_5 \sin\theta_5 \int_0^{\pi} d\theta_6 \sin^2\theta_6 . \quad (7.153)$$

This is just a special case of the general $n$ result, which reads

$$d\mu_H(\mathbf{R}) = \left( \prod_{i=1}^{n} \Gamma(i/2)/2^n\, \pi^{n(n+1)/2} \right) \prod_{i=1}^{n-1} \prod_{j=1}^{i} \sin^{j-1}\theta^i_j\, d\theta^i_j , \quad (7.154)$$

with $0 \le \theta^1_k < 2\pi$, $0 \le \theta^j_k < \pi$.

These averaging properties of rotations are quite similar of course to what happens in $SU(N)$ Yang-Mills theories, or even more simply in (compact) QED, where the analogs of the $SO(d)$ rotation matrices **R** are phase factors $U_\mu(x) = e^{iaA_\mu(x)}$. There one has $\int \frac{dA_\mu}{2\pi} U_\mu(x) = 0$ and $\int \frac{dA_\mu}{2\pi} U_\mu(x) U^\dagger_\mu(x) = 1$. In addition, for two contiguous closed paths $C_1$ and $C_2$ sharing a common side one has

$$e^{i \oint_{C_1} \mathbf{A} \cdot \mathbf{dl}} e^{i \oint_{C_2} \mathbf{A} \cdot \mathbf{dl}} = e^{i \oint_{C} \mathbf{A} \cdot \mathbf{dl}} = e^{i \int_{S} \mathbf{B} \cdot \mathbf{n} dA} , \tag{7.155}$$

with $C$ the slightly larger path encircling the two loops. For a closed surface tiled with many contiguous infinitesimal closed loops the last expression evaluates to 1, due to the divergence theorem. In the lattice gravity case the discrete analog of this last result is considerably more involved, and ultimately represents the (exact) lattice analog of the contracted Bianchi identities. An example of a closed surface tiled with parallel transport polygons (here chosen for simplicity to be triangles) is shown in Fig. 7.13.

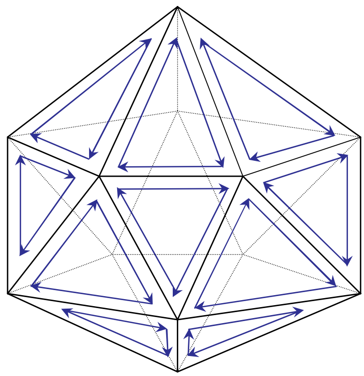


**Fig. 7.13** Elementary closed surface tiled with parallel transport polygons, here chosen to be triangles for illustrative purposes. For each link of the dual lattice, the elementary parallel transport matrices $\mathbf{R}(s,s')$ are represented by an arrow. In spite of the fact that the (Lorentz) matrices $\mathbf{R}$ fluctuate with the local geometry, two contiguous, oppositely oriented arrows always give $\mathbf{R}\mathbf{R}^{-1} = 1$.

As one approaches the critical point, $k \to k_c$, one is interested in random surfaces which are of very large extent. Let $n_p$ be the number of polygons in the surface, and set $n_p = T^2$ since after all one is describing a surface. The critical point then naturally corresponds to the appearance of surfaces of infinite extent,

$$n_p = T^2 \sim \frac{1}{k_c - k} \to \infty . \tag{7.156}$$

A legitimate parallel is to the simpler case of scalar field theories, where random walks of length $T$ describing particle paths become of infinite extent at the critical point, situated where the inverse of the (renormalized) mass $\xi = m^{-1}$, expressed in units of the ultraviolet cutoff, diverges.

In the present case of polygonal random surfaces, one can provide the following concise argument in support of the identification in Eq. (7.156). First approximate the discrete sums over $n$, as they appear for example in the strong coupling expansion for the average curvature, Eq. (7.69) or its correlation, Eq. (7.119), by continuous integrals over areas

$$\sum_{n=0}^{\infty} c_n \left(\frac{k}{k_c}\right)^n \rightarrow \int_0^{\infty} dA A^{\gamma-1} \left(\frac{k}{k_c}\right)^A = \Gamma(\gamma) \left(\log \frac{k_c}{k}\right)^{-\gamma} , \tag{7.157}$$

where $A \equiv T^2$ is the area of a given surface. The $A^{\gamma-1}$ term can be regarded as counting the multiplicity of the surface (its entropy, in statistical mechanics terms). The exponent $\gamma$ depends on the specific quantity one is looking at. For the average curvature one has from Eq. (7.76) $\gamma = -\delta$, while for its derivative, the curvature fluctuation (the curvature correlation function at zero momentum), one expects $\gamma = 1 - \delta$. The saddle point is located at

$$A = \frac{(\gamma - 1)}{\log \frac{k_c}{k}} \underset{k \to k_c}{\sim} \frac{(\gamma - 1)\, k_c}{k_c - k} . \tag{7.158}$$

From this discussion one then concludes that close to the critical point very large areas dominate, as claimed in Eq. (7.156).

Furthermore, one would expect that the universal geometric scaling properties of such a (closed) surface would not depend on its short distance details, such as whether it is constructed out of say triangles or more complex polygons. In general excluded volume effects at finite $d$ will provide constraints on the detailed geometry of the surface, but as $d \to \infty$ these constraints can presumably be neglected and one is dealing then with a more or less unconstrained random surface. This should be regarded as a direct consequence of the fact that as $d \to \infty$ there are infinitely many dimensions for the random surface to twist and fold into, giving a negligible contribution from unallowed (by interactions) directions. In the following we will assume that this is indeed the case, and that no special pathologies arise, such as the collapse of the random surface into narrow tube-like, lower dimensional geometric configurations. Then in the large $d$ limit the problem simplifies considerably.

Related examples for what is meant in this context are the simpler cases of random walks in infinite dimensions, random polymers and random surfaces in gauge theories (Drouffe, Parisi and Sourlas, 1979), which have been analysed in detail in the large-$d$ limit. There too the problem simplifies considerably in such a limit since excluded volume effects (self-intersections) can be neglected there as well. A summary of these results, with a short derivation, is given in the appendices of (Hamber and Williams, 2006).

Following (Gross, 1984) one can define the partition function for such an ensemble of unconstrained random surfaces, and one finds that the mean square size of the surface increases logarithmically with the intrinsic area of the surface. This last result is usually interpreted as the statement that an unconstrained random surface has infinite fractal (or Hausdorff) dimension. Although made of very many triangles (or polygons), the random surface remains quite compact in overall size, as viewed from the original embedding space. In a sense, an unconstrained random surface is a much more compact object than an unconstrained random walk, for which $< \mathbf{X}^2 > \sim T$. Identifying the size of the random surface with the gravitational correlation length $\xi$ then gives

$$\xi \sim \sqrt{\log T} \underset{k \to k_c}{\sim} |\log(k_c - k)|^{1/2} \ . \tag{7.159}$$

From the definition of the exponent $\nu$, namely $\xi \sim (k_c - k)^{-\nu}$, the above result then implies $\nu = 0$ (i.e. a weak logarithmic singularity) at $d = \infty$.

It is of interest to contrast the result $\nu \sim 0$ for gravity in large dimensions with what one finds for scalar (Wilson and Fisher, 1972; Wilson, 1973) and gauge (Drouffe, Parisi and Sourlas, 1979) fields, in the same limit $d = \infty$. So far, known results can be summarized as follows

$$\begin{aligned} &\text{scalar field} && \nu = \tfrac{1}{2} \\ &\text{lattice gauge field} && \nu = \tfrac{1}{4} \\ &\text{lattice gravity} && \nu = 0 \ . \end{aligned} \tag{7.160}$$

It should be regarded as encouraging that the new value obtained here, namely $\nu = 0$ for gravitation, appears to some extent to be consistent with the general trend observed for lower spin, at least at infinite dimension. What happens in finite dimensions? The situation becomes much more complicated since the self-intersection properties of the surface have to be taken into account. But a simple geometric argument then suggests in finite but large dimensions $\nu = 1/(d-1)$ (Hamber and Williams, 2004).

## 7.7 Mean Field Theory

In this section we will describe briefly a simple mean-field approach to quantum gravity, which contains some features observed in the numerical simulations. Write for the local average curvature $\mathscr{R}$,

$$\mathscr{R}(k) = \frac{< \int d^d x \sqrt{g} R >}{< \int d^d x \sqrt{g} >} \ , \tag{7.161}$$

an effective action (or effective potential) which entirely neglects any further effects of the metric degrees of freedom,

$$I_{eff}(\mathscr{R}) = (k - k_c) V |\mathscr{R}| + a V |\mathscr{R}|^\lambda \ , \tag{7.162}$$

with as usual $k \equiv 1/8\pi G$ and $a > 0$ some additional coupling; in the strong coupling phase of gravity $k < k_c$.

The above effective action is inspired by the analogy with the Landau theory for order-disorder transitions (Landau and Lifshitz, 1980), and the term proportional to $a$ is supposed to represent, in some crude and effective theory way, the effects of the interactions. Classically one has of course $k_c = 0$, but fluctuations give rise to a nonzero value for the critical coupling that separates the smooth ($k < k_c$) from the rough phase ($k > k_c$). The last term can be thought of parametrizing the lattice and continuum higher derivative terms, as well as the effects of radiative corrections, which include the measure contribution.

Numerical simulations show that in the smooth phase of gravity $\mathscr{R} < 0$, so one can set $\mathscr{R} = -|\mathscr{R}|$ in this phase; a physically acceptable phase with local average curvature $\mathscr{R} > 0$ (the rough phase) does not seem to exist. Then

$$\frac{\partial I_{eff}}{\partial \mathscr{R}} = (k_c - k) V - a \lambda V \, (-\mathscr{R})^{\lambda - 1} \; , \tag{7.163}$$

with stationary point at

$$\mathscr{R}_0(k) = -(a\lambda)^{-1/(\lambda-1)} \, (k_c - k)^{1/(\lambda-1)} \; . \tag{7.164}$$

This in a sense justifies the original form for $I_{eff}$, since it is known that the average curvature is non-analytic at $k_c$, [see for example Eq. (8.53)], and identifies the curvature critical exponent as $\delta = 1/(\lambda - 1)$. Furthermore it gives, from $\delta = d\nu - 1$,

$$\lambda^{-1} = 1 - \frac{1}{d\,\nu} \; , \tag{7.165}$$

where $\nu$ is the correlation length exponent. For the fluctuation in the local curvature one then obtains

$$\chi_{\mathscr{R}_0} \sim (k_c - k)^{-(\lambda-2)/(\lambda-1)} \; , \tag{7.166}$$

with exponent $\alpha = 2 - d\nu = (\lambda - 2)/(\lambda - 1)$, and in agreement with Eq. (7.165).

Classically one expects of course no higher order corrections, which would then correspond to either $a = 0$ or $\lambda = 1$. But in four dimensions numerical simulations give $\nu = 1/3$, which gives instead $\lambda = 4$. In general for large enough dimensions one expects $\nu = 1/(d-1)$ [see Sect. (7.6)], which would then imply $\lambda = d$. Furthermore, as long as $\mathscr{R} < 0$ the above solution is stable, since

$$\begin{aligned} \frac{\partial^2 I_{eff}}{\partial \mathscr{R}^2} &= aV \, \lambda \, (\lambda - 1) \, (-\mathscr{R})^{\lambda-2} \; , \\ &= a \, V \, \lambda \, (\lambda - 1) \, (a\lambda)^{-(\lambda-2)/(\lambda-1)} \, (k_c - k)^{(\lambda-2)/(\lambda-1)} \; , \end{aligned} \tag{7.167}$$

which incidentally requires $\lambda > 1$ for the second derivative of $I_{eff}$ to be finite at the origin $\mathscr{R} = 0$. Note that in this approach there is always only one minimum for $k < k_c$.

For $\mathscr{R} > 0$ the effective action is complex, as it should, since no stable ground state is found in the lattice theory for $\mathscr{R} > 0$. Two further predictions arise out of this model. The first one is that the amplitude of the average curvature should diverge when the parameter $a$ becomes sufficiently small, as

$$A_{\mathscr{R}} \sim a^{-1/(\lambda-1)} \; . \tag{7.168}$$

The second one is that the minimum should become increasingly shallow as $a \to 0$, which leads to larger fluctuations in the average curvature. But in the end one does not expect the mean field theory to be quantitatively accurate, just as it is not for scalar field theories in low dimensions. It only represents an effective theory for the average local curvature, which is represented here as a single scalar quantity.

# Chapter 8
# Numerical Studies

## 8.1 Nonperturbative Gravity

The exact evaluation of the lattice functional integral for quantum gravity by numerical methods allows one to investigate a regime which is generally inacessible by perturbation theory, where the coupling $G$ is strong and quantum fluctutations are expected to be large.

The hope in the end is to make contact with the analytic results obtained, for example, in the $2+\varepsilon$ expansion, and determine which scenarios are physically realized in the lattice regularized model, and then perhaps even in the real world.

Specifically, one can enumerate several major questions that one would like to get at least partially answered. The first one is: which scenarios suggested by perturbation theory are realized in the lattice theory? Perhaps a stable ground state for the quantum theory cannot be found, which would imply that the regulated theory is still inherently pathological. Furthermore, if a stable ground state exists for some range of bare parameters, does it require the inclusion of higher derivative couplings in an essential way, or is the minimal theory, with an Einstein and a cosmological term, sufficient? Does the presence of dynamical matter, say in the form of a massless scalar field, play an important role, or is the non-perturbative dynamics of gravity determined largely by the pure gravity sector (as in Yang-Mills theories)?

More generally, is there any indication that the non-trivial ultraviolet fixed point scenario is realized in the lattice theory in four dimensions? This would imply, as in the non-linear sigma model, the existence of at least two physically distinct phases and non-trivial exponents. Which quantity can be used as an order parameter to physically describe, in a *qualitative*, way the two phases? A clear physical characterization of the two phases would allow one, at least in principle, to decide which phase, if any, could be realized in nature. Ultimately this might or might not be possible based on purely qualitative aspects. As will discussed below, the lattice continuum limit is taken in the vicinity of the fixed point, so close to it is the physically most relevant regime. At the next level one would hope to be able to establish a *quantitative* connection with those continuum perturbative results which are not affected by uncontrollable errors, such as for example the $2+\varepsilon$ expansion of Sect. 3.5. Since the lattice cutoff and the method of dimensional regularization cut the theory off in the ultraviolet in rather different ways, one needs to compare universal quantities which are *cutoff-independent*. One example is the critical exponent $\nu$, as well as any other non-trivial scaling dimension that might arise. Within the $2+\varepsilon$ expansion only *one* such exponent appears, to *all* orders in the loop expansion, as $\nu^{-1} = -\beta'(G_c)$. Therefore one central issue in the lattice regularized theory is the value of the universal exponent $\nu$.

Knowledge of $\nu$ would allow one to be more specific about the running of the gravitational coupling. One purpose of the discussion in Sect. 3.3 was to convince the reader that the exponent $\nu$ determines the renormalization group running of $G(\mu^2)$ in the vicinity of the fixed point, as in Eq. (3.22) for the non-linear $\sigma$-model, and more appropriately in Eq. (3.117) for quantized gravity. From a practical point of view, on the lattice it is difficult to determine the running of $G(\mu^2)$ directly from correlation functions , since the effects from the running of $G$ are generally small. Instead one would like to make use of the analog of Eqs. (3.29), (3.59) and (3.60) for the non-linear $\sigma$-model, and, again, more appropriately of Eqs. (3.121) and possibly (3.127) for gravity to determine $\nu$, and from there the running of $G$. But the correlation length $\xi = m^{-1}$ is also difficult to compute, since it enters the curvature correlations at fixed geodesic distance, which are hard to compute for (genuinely geometric) reasons to be discussed later. Furthermore, these generally decay exponentially in the distance at strong $G$, and can therefore be difficult to compute due to the signal to noise problem of numerical simulations.

Fortunately the exponent $\nu$ can be determined instead, and with good accuracy, from singularities of the derivatives of the path integral $Z$, whose singular part is expected, on the basis of very general arguments, to behave in the vicinity of the fixed point as $F \equiv -\frac{1}{V}\ln Z \sim \xi^{-d}$ where $\xi$ is the gravitational correlation length. From Eq. (3.121) relating $\xi(G)$ to $G-G_c$ and $\nu$ one can then determine $\nu$, as well as the critical coupling $G_c$.

## 8.2 Observables, Phase Structure and Critical Exponents

The starting point is once again the lattice regularized path integral with action as in Eq. (6.43) and measure as in Eq. (6.76). Then the lattice action for pure four-dimensional Euclidean gravity contains a cosmological constant and Regge scalar curvature term as in Eq. (6.90)

$$I_{latt} \;=\; \lambda_0 \sum_h V_h(l^2) \,-\, k \sum_h \delta_h(l^2) A_h(l^2) \;\;, \tag{8.1}$$

with $k = 1/(8\pi G)$, and leads to the regularized lattice functional integral

$$Z_{latt} \;=\; \int [d\,l^2]\; e^{-\lambda_0 \sum_h V_h \,+\, k \sum_h \delta_h A_h} \;\;, \tag{8.2}$$

where, as customary, the lattice ultraviolet cutoff is set equal to one (i.e. all length scales are measured in units of the lattice cutoff). The lattice measure is given in Eq. (6.76) and will be therefore of the form

$$\int [d\,l^2] \;=\; \int_0^\infty \prod_s \,[V_d(s)]^\sigma \,\prod_{ij} dl_{ij}^2 \,\Theta[l_{ij}^2] \;\;, \tag{8.3}$$

with $\sigma$ a real parameter given below.

Ultimately the above lattice partition function $Z_{latt}$ is intended as a regularized form of the continuum Euclidean Feynman path integral of Eq. (2.34),

$$Z_{cont} \;=\; \int [d\,g_{\mu\nu}]\; e^{-\lambda_0 \int dx \sqrt{g} + \frac{1}{16\pi G} \int dx \sqrt{g} R} \;\;, \tag{8.4}$$

with functional measure over the $g_{\mu\nu}(x)$'s of the form

$$\int [d\,g_{\mu\nu}] \equiv \prod_x \,[g(x)]^{\sigma/2} \prod_{\mu \geq \nu} dg_{\mu\nu}(x) \;\;, \tag{8.5}$$

where $\sigma$ is a real parameter constrained by the requirement $\sigma \geq -(d+1)$. For $\sigma = \frac{1}{2}(d-4)(d+1)$ one obtains the De Witt measure of Eq. (2.18), while for $\sigma = -(d+1)$ one recovers the original Misner measure of Eq. (2.22). In the following we will mostly be interested in the four-dimensional case, for which $d = 4$ and therefore $\sigma = 0$ for the DeWitt measure.

It is possible to add higher derivative terms to the lattice action and investigate how the results are affected. The original motivation was that they would improve the convergence properties of functional integral for the lattice theory, but extensive numerical studies suggest that they don't seem to be necessary after all. In any case, with such terms included the lattice action for pure gravity acquires the two additional terms whose lattice expressions can be found in Eqs. (6.111) and (6.122),

$$\begin{aligned} I_{latt} = & \sum_h \Big[ \lambda_0 V_h - k\, \delta_h A_h - b A_h^2 \delta_h^2 / V_h \Big] \\ & + \tfrac{1}{3}\,(a+4b) \sum_s V_s \sum_{h,h' \subset s} \varepsilon_{h,h'} \left( \omega_h \left[ \frac{\delta}{A_C} \right]_h - \omega_{h'} \left[ \frac{\delta}{A_C} \right]_{h'} \right)^2 \;. \end{aligned} \tag{8.6}$$

The above action is intended as a lattice form for the continuum action

$$I = \int dx \sqrt{g} \left[ \lambda_0 - \tfrac{1}{2} k R - \tfrac{1}{4} b R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} + \tfrac{1}{2}\,(a+4b)\, C_{\mu\nu\rho\sigma} C^{\mu\nu\rho\sigma} \right] \;, \tag{8.7}$$

and is therefore of the form in Eqs. (1.137) and (6.123). Because of its relative complexity, in the following the Weyl term will not be considered any further, and $b$ will chosen so that $b = -\frac{1}{4}a$. Thus the only curvature term to be discussed here will be a Riemann squared contribution, with a (small) positive coefficient $+\frac{1}{4}a \to 0$.

## 8.3 Invariant Local Gravitational Averages

Among the simplest quantum mechanical averages is the one associated with the local curvature

$$\mathscr{R}(k) \sim \frac{< \int dx \sqrt{g} R(x) >}{< \int dx \sqrt{g} >} \quad . \tag{8.8}$$

The curvature associated with the quantity above is the one that would be detected when parallel-transporting vectors around infinitesimal loops, with size comparable to the average lattice spacing $l_0$. Closely related to it is the fluctuation in the local curvature

$$\chi_{\mathscr{R}}(k) \sim \frac{< (\int dx \sqrt{g} R)^2 > - < \int dx \sqrt{g} R >^2}{< \int dx \sqrt{g} >} \quad . \tag{8.9}$$

The latter is related to the connected curvature correlation at zero momentum

$$\chi_{\mathscr{R}} \sim \frac{\int dx \int dy < \sqrt{g(x)} R(x) \sqrt{g(y)} R(y) >_c}{< \int dx \sqrt{g(x)} >} \quad . \tag{8.10}$$

Both $\mathscr{R}(k)$ and $\chi_{\mathscr{R}}(k)$ are directly related to derivatives of $Z$ with respect to $k$,

$$\mathscr{R}(k) \sim \frac{1}{V} \frac{\partial}{\partial k} \ln Z \tag{8.11}$$

and

$$\chi_{\mathscr{R}}(k) \sim \frac{1}{V} \frac{\partial^2}{\partial k^2} \ln Z \quad . \tag{8.12}$$

Thus a divergence or non-analyticity in $Z$, as caused for example by a phase transition, is expected to show up in these local averages as well. Note that the above expectation values are manifestly invariant, since they are related to derivatives of $Z$.

On the lattice one prefers to define quantities in such a way that variations in the average lattice spacing $\sqrt{< l^2 >}$ are compensated by an appropriate factor determined from dimensional considerations. In the case of the average curvature one defines therefore the lattice quantity $\mathscr{R}$ as

$$\mathscr{R}(k) \equiv < l^2 > \frac{< 2 \sum_h \delta_h A_h >}{< \sum_h V_h >} \quad , \tag{8.13}$$

and similarly for the curvature fluctuation,

$$\chi_{\mathscr{R}}(k) \equiv \frac{< (\sum_h \delta_h A_h)^2 > - < \sum_h \delta_h A_h >^2}{< \sum_h V_h >} \quad . \tag{8.14}$$

Fluctuations in the local curvature probe graviton correlations, and are expected to be sensitive to the presence of a massless spin two particle. Note that both of the above expressions are dimensionless, and are thefore unaffected by an overall rescaling of the edge lengths. As in the continuum, they are proportional to first and second derivatives of $Z_{latt}$ with respect to $k$.

One can contrast the behavior of the preceding averages, related to the curvature, with the corresponding quantities involving the local volumes $V_h$ (the quantity $\sqrt{g}\,dx$ in the continuum). Consider the average volume per site

$$\langle V \rangle \equiv \frac{1}{N_0} < \sum_h V_h > \quad , \tag{8.15}$$

and its fluctuation, defined as

$$\chi_V(k) \equiv \frac{< (\sum_h V_h)^2 > - < \sum_h V_h >^2}{< \sum_h V_h >} \quad , \tag{8.16}$$

where $V_h$ is the volume associated with the hinge $h$. The last two quantities are again simply related to derivatives of $Z_{latt}$ with respect to the bare cosmological constant $\lambda_0$, as for example in

$$< V > \sim \frac{\partial}{\partial \lambda_0} \ln Z_{latt} \tag{8.17}$$

and

$$\chi_V(k) \sim \frac{\partial^2}{\partial \lambda_0^2} \ln Z_{latt} \quad . \tag{8.18}$$

Some useful relations and sum rules can be derived, which follow directly from the scaling properties of the discrete functional integral. Thus a simple scaling argument, based on neglecting the effects of curvature terms entirely (which, as will be seen below, vanish in the vicinity of the critical point), gives an estimate of the average volume per edge [for example from Eqs. (7.148) and (7.149)]

$$< V_l > \sim \frac{2\,(1+\sigma\,d)}{\lambda_0\,d} \underset{d=4,\,\sigma=0}{\sim} \frac{1}{2\lambda_0} \quad , \tag{8.19}$$

where $\sigma$ is the functional measure parameter in Eqs. (2.27) and (6.76). In four dimensions direct numerical simulations with $\sigma = 0$ (corresponding to the lattice DeWitt measure) agree quite well with the above formula.

Some exact lattice identities can be obtained from the scaling properties of the action and measure. The bare couplings $k$ and $\lambda_0$ in the gravitational action are dimensionful in four dimensions, but one can define the dimensionless ratio $k^2/\lambda_0$, and rescale the edge lengths so as to eliminate the overall length scale $\sqrt{k/\lambda_0}$. As a consequence the path integral for pure gravity,

$$Z_{latt}(\lambda_0, k, a, b) = \int [dl^2]\, e^{-I(l^2)} \quad , \tag{8.20}$$

obeys the scaling law

$$Z_{latt}(\lambda_0,k,a,b) \;=\; (\lambda_0)^{-N_1/2} Z_{latt}\left(1, \frac{k}{\sqrt{\lambda_0}}, a, b\right) \;\;, \tag{8.21}$$

where $N_1$ represents the number of edges in the lattice, and the $dl^2$ measure ($\sigma = 0$) has been selected. This implies in turn a sum rule for local averages, which for the $dl^2$ measure reads

$$2\lambda_0 < \sum_h V_h > \; - \, k < \sum_h \delta_h A_h > \; - \, N_1 \;=\; 0 \;\;, \tag{8.22}$$

and is easily derived from Eq. (8.21) and the definitions in Eqs. (8.11) and (8.17). $N_0$ represents the number of sites in the lattice, and the averages are defined per site (for the hypercubic lattice divided up into simplices as in Fig. 7.1, $N_1 = 15$). This last formula can be very useful in checking the accuracy of numerical evaluations of the path integral. A similar sum rule holds for the fluctuations

$$\begin{aligned} 4\,\lambda_0^2 & \left[ < (\sum_h V_h)^2 > - < \sum_h V_h >^2 \right] \\ & - k^2 \left[ < (\sum_h \delta_h A_h)^2 > - < \sum_h \delta_h A_h >^2 \right] \; - \; 2N_1 \;=\; 0 \;\;. \end{aligned} \tag{8.23}$$

In light of the above discussion one can therefore consider without loss of generality the case of unit bare cosmological coupling $\lambda_0 = 1$ (in units of the cutoff). Then all lengths are expressed in units of the fundamental microscopic length scale $\lambda_0^{-1/4}$.

## 8.4 Invariant Correlations at Fixed Geodesic Distance

Compared to ordinary field theories, new issues arise in quantum gravity due to the fact that the physical distance between any two points $x$ and $y$

$$d(x,y\,|\,g) \;=\; \min_{\xi} \int_{\tau(x)}^{\tau(y)} d\tau \, \sqrt{g_{\mu\nu}(\xi) \tfrac{d\xi^\mu}{d\tau} \tfrac{d\xi^\nu}{d\tau}} \;\;, \tag{8.24}$$

is a fluctuating function of the background metric $g_{\mu\nu}(x)$. In addition, the Lorentz group used to classify spin states is meaningful only as a local concept.

In the continuum the shortest distance between two events is determined by solving the equation of motion (equation of free fall, or geodesic equation)

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma^\mu_{\lambda\sigma} \frac{dx^\lambda}{d\tau} \frac{dx^\sigma}{d\tau} = 0 \;\;. \tag{8.25}$$

On the lattice the geodesic distance between two lattice vertices $x$ and $y$ requires the determination of the shortest lattice path connecting several lattice vertices, and having the two given vertices as endpoints. This can be done at least in principle by enumerating all paths connecting the two points, and then selecting the shortest one. Equivalently it can be computed from the scalar field propagator, as in Eq. (6.145).

Consequently physical correlations have to be defined at fixed geodesic distance $d$, as in the following correlation between scalar curvatures

$$< \int dx \int dy \sqrt{g} R(x) \sqrt{g} R(y)\, \delta(|x-y|-d) > \quad . \tag{8.26}$$

Generally these do not go to zero at large separation, so one needs to define the connected part, by subtracting out the value at $d = \infty$. These will be indicated in the following by the connected $<>_c$ average, and we will write the resulting connected curvature correlation function at fixed geodesic distance compactly as

$$G_R(d) \;\sim < \sqrt{g}\, R(x) \sqrt{g}\, R(y)\, \delta(|x-y|-d) >_c \quad . \tag{8.27}$$

One can define several more invariant correlation functions at fixed geodesic distance for other operators involving curvatures (Hamber, 1994). The gravitational correlation function just defined is similar to the one in non-Abelian gauge theories, Eq. (3.146).

In the lattice regulated theory one can define similar correlations, using for example the correspondence of Eqs. (6.38) or (6.110) for the scalar curvature

$$\sqrt{g} R(x) \;\to\; 2 \sum_{h \supset x} \delta_h A_h \quad . \tag{8.28}$$

If the deficit angles are averaged over a number of contiguous hinges $h$ sharing a common vertex $x$, one is naturally lead to the connected correlation function

$$G_R(d) \;\equiv < \sum_{h \supset x} \delta_h A_h \sum_{h' \supset y} \delta_{h'} A_{h'}\; \delta(|x-y|-d) >_c \,, \tag{8.29}$$

which probes correlations in the scalar curvatures. In practice the above lattice correlations have to be computed by a suitable binning procedure: one averages out all correlations in a geodesic distance interval $[d, d+\Delta d]$ with $\Delta d$ comparable to one lattice spacing $l_0$. See Fig. 8.1. Similarly one can construct the connected correlation functions for local volumes at fixed geodesic distance

$$G_V(d) \;\equiv < \sum_{h \supset x} V_h \sum_{h' \supset y} V_{h'}\; \delta(|x-y|-d) >_c \quad . \tag{8.30}$$

In general one expects for the curvature correlation either a power law decay, for distances sufficiently larger than the lattice spacing $l_0$,

$$< \sqrt{g}\, R(x) \sqrt{g}\, R(y)\, \delta(|x-y|-d) >_c \; \underset{d \gg l_0}{\sim} \; \frac{1}{d^{2n}} \quad , \tag{8.31}$$

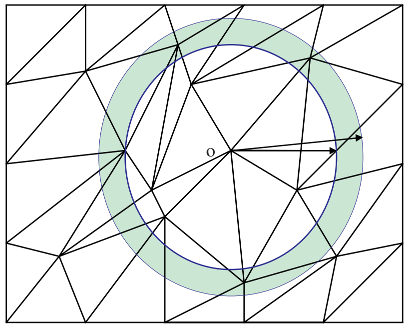



**Fig. 8.1** Geodesic distance and correlations. On each metric configuration correlation functions are computed for lattice vertices within the physical distance range $[d, d+\Delta d]$.

with $n$ some exponent characterizing the power law decay, or at very large distances an exponential decay, characterized by a correlation length $\xi$,

$$< \sqrt{g}\, R(x) \,\sqrt{g}\, R(y)\; \delta(|x-y|-d) >_c \underset{d \gg \xi}{\sim} e^{-d/\xi} \quad . \tag{8.32}$$

In fact the invariant correlation length $\xi$ is generally defined (in analogy with what one does for other theories) through the long-distance decay of the connected, invariant correlations at fixed geodesic distance $d$. In the pure power law decay case of Eq. (8.31) the correlation length $\xi$ is of course infinite. One can show from scaling considerations (see below) that the power $n$ in Eq. (8.31) is related to the critical exponent $\nu$ by $n = 4 - 1/\nu$.

In the presence of a finite correlation length $\xi$ one needs therefore to carefully distinguish between the "short distance" regime

$$l_0 \ll d \ll \xi \ , \tag{8.33}$$

where Eq. (8.31) is valid, and the "long distance" regime

$$\xi \ll d \ll L \ , \tag{8.34}$$

where Eq. (8.32) is appropriate. Here $l_0 = \sqrt{< l^2 >}$ is the average lattice spacing, and $L = V^{1/4}$ the linear size of the system.

Recently the issue of defining diffeomorphism invariant correlations in quantum gravity has been re-examined from a new perspective (Giddings, Marolf and Hartle, 2006).

## 8.5 Wilson Lines and Static Potential

In a gauge theory such as QED the static potential can be computed from the manifestly gauge invariant Wilson loop. To this end one considers the process where a particle-antiparticle pair are created at time zero, separated by a fixed distance $R$, and re-annihilated at a later time $T$. In QED the amplitude for such a process associated with the closed loop $\Gamma$ is given by the Wilson loop

$$W(\Gamma) \;=< \exp\Big\{ ie \oint_\Gamma A_\mu(x) dx^\mu \Big\} > \;, \tag{8.35}$$

which is a manifestly gauge invariant quantity. Performing the required Gaussian average using the (Euclidean) free photon propagator one obtains

$$< \exp\Big\{ ie \oint_\Gamma A_\mu dx^\mu \Big\} > = \; \exp\Big\{ -\tfrac{1}{2} e^2 \oint_\Gamma \oint_\Gamma dx^\mu dy^\nu \Delta_{\mu\nu}(x-y) \Big\} \;. \tag{8.36}$$

For a rectangular loop of sides $R$ and $T$ one has after a short calculation

$$< \exp\Big\{ ie \oint_\Gamma A_\mu dx^\mu \Big\} > \simeq \; \exp\Big\{ -\frac{e^2}{4\pi\varepsilon}(T+R) + \frac{e^2}{4\pi}\frac{T}{R} + \frac{e^2}{2\pi^2}\log(\frac{T}{\varepsilon}) + \cdots \Big\} \tag{8.37}$$

$$\underset{T \gg R}{\sim} \; \exp\Big[ -V(R)\, T \Big] \;, \tag{8.38}$$

where $\varepsilon \to 0$ is an ultraviolet cutoff. In the last line use has been made of the fact that for large imaginary times the exponent in the amplitude involves the energy for the process multiplied by the time $T$. Thus for $V(R)$ itself one obtains

$$V(R) = -\lim_{T\to\infty} \frac{1}{T} \log < \exp\Big\{ ie \oint_\Gamma A_\mu dx^\mu \Big\} > \sim \mathrm{cst.} - \frac{e^2}{4\pi R} \;, \tag{8.39}$$

which is the correct Coulomb potential for two oppositely charged particles.

To obtain the static potential it is not necessary to consider closed loops. Alternatively, in a periodic box of length $T$ one can introduce two long oppositely oriented parallel lines in the time direction, separated by a distance $R$ and closed by the periodicity of the lattice, and associated with oppositely charged particles,

$$< \exp\Big\{ ie \int_\Gamma A_\mu dx^\mu \Big\} \exp\Big\{ ie \int_{\Gamma'} A_\nu dy^\nu \Big\} > \tag{8.40}$$

In the large time limit one can then show that the result for the potential $V(R)$ is the same.

In the gravitational case there is no notion of "oppositely charged particles", so one cannot use the closed Wilson loop to extract the potential (Modanese, 1995). One is therefore forced to consider a process in which one introduces two separate world-lines for the two particles. It is well known that the equation for free fall can be obtained by extremizing the space-time distance travelled. Thus the quantity

$$\mu \int_{\tau(a)}^{\tau(b)} d\tau \sqrt{g_{\mu\nu}(x) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}} \ , \tag{8.41}$$

can be taken as the Euclidean action contribution associated with the heavy spinless particle of mass $\mu$.

Next consider two particles of mass $M_1$, $M_2$, propagating along parallel lines in the "time" direction and separated by a fixed distance $R$. Then the coordinates for the two particles can be chosen to be $x^\mu = (\tau, r, 0, 0)$ with $r$ either 0 or $R$. The amplitude for this process is a product of two factors, one for each heavy particle. Each is of the form

$$L(0;\, M_1) = \exp\left\{ -M_1 \int d\tau \sqrt{g_{\mu\nu}(x) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}} \right\} \ , \tag{8.42}$$

where the first argument indicates the spatial location of the Wilson line. For the two particles separated by a distance $R$ the amplitude is

$$\text{Amp.} \equiv W(0,R;\, M_1,M_2) = L(0;\, M_1)\, L(R;\, M_2) \ . \tag{8.43}$$

For weak fields one sets $g_{\mu\nu} = \delta_{\mu\nu} + h_{\mu\nu}$, with $h_{\mu\nu} \ll 1$, and therefore $g_{\mu\nu}(x) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}$ $= 1 + h_{00}(x)$. Then the amplitude reduces to

$$W(M_1, M_2) = \exp\left\{ -M_1 \int_0^T d\tau \sqrt{1 + h_{00}(\tau)} \right\} \exp\left\{ -M_2 \int_0^T d\tau' \sqrt{1 + h_{00}(\tau')} \right\} \ . \tag{8.44}$$

In perturbation theory the averaged amplitude can then be easily evaluated (Hamber and Williams, 1995)

$$< W(0,R;\, M_1,M_2) > = \exp\left\{ -T\ (M_1 + M_2 - G\, \frac{M_1 M_2}{R}) + \cdots \right\} \ , \tag{8.45}$$

and the static potential has indeed the expected form, $V(R) = -G\, M_1 M_2 / R$. The contribution involving the sum of the two particle masses is $R$ independent, and can therefore be subtracted, if the Wilson line correlation is divided by the averages of the individual single line contribution, as in

$$V(R) = -\lim_{T \to \infty} \frac{1}{T} \log \frac{< W(0,R;\, M_1,M_2) >}{< L(0;\, M_1) >< L(R;\, M_2) >} \sim -\, G\, \frac{M_1 M_2}{R} \ . \tag{8.46}$$

If one is only interested in the spatial dependence of the potential, one can simplify things further and take $M_1 = M_2 = M$. To higher order in the weak field expansion one has to take into account multiple graviton exchanges, contributions from graviton loops and self-energy contributions due to other particles.

How does all this translate to the lattice theory? At this point, the prescription for computing the Newtonian potential for quantum gravity should be clear. For each metric configuration (which is a given configuration of edge lengths on the lattice) one chooses a geodesic that closes due to the lattice periodicity (and there might be many that have this property for the topology of a four-torus), with length $T$ (see Fig. 8.2). One then enumerates all the geodesics that lie at a fixed distance $R$ from the original one, and computes the associated correlation between the Wilson lines. After averaging the Wilson line correlation over many metric configurations, one extracts the potential from the $R$ dependence of the correlation of Eq. (8.46). In general since two geodesics will not be at a fixed geodesic distance from each other in the presence of curvature, one needs to introduce some notion of average distance, which then gives the spatial separations of the sources $R$.

On the lattice one can construct the analog of the Wilson line for one heavy particle,

$$L(x,y,z) \;=\; \exp\{-M\sum_i l_i\} \;, \tag{8.47}$$

where edges are summed in the "$t$" direction, and the path is closed by the periodicity of the lattice in the $t$ direction. One can envision the simplicial lattice as divided up in hypercubes, in which case the points $x,y,z$ can be taken as the remaining labels for the Wilson line.

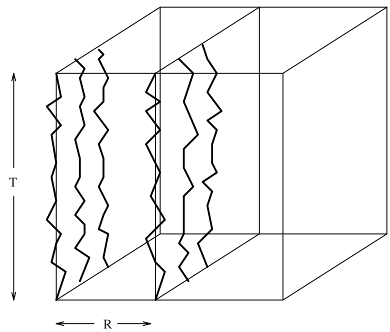



**Fig. 8.2** Correlations between Wilson lines closed by the lattice periodicity.

For a single line one expects

$$< L(x,y,z) > = < \exp\{-M\sum_i l_i\} > \sim e^{-\tilde{M}T} \quad , \tag{8.48}$$

where $T$ is the linear size of lattice in the $t$ direction and $\tilde{M}$ the renormalized mass. The correlation between Wilson lines at average "distance" $R$ is then given by

$$-\frac{1}{T}\,\log\left[\frac{< L(x,y,0)\, L(x,y,R) >}{< L(x,y,0) >< L(x,y,R) >}\right] \underset{T\gg R}{\sim} V(R) \;. \tag{8.49}$$

Numerical studies suggest that the correct qualitative features of the potential emerge close to the critical point. In particular it was found that the potential is attractive close to the critical point, and for two equal mass particles of mass $\mu$ scales, as expected, like the mass squared. As for any correlation in gravity, the accurate determination of the potential as a function of distance $R$ is a more difficult task, since at large distance the correlations are small and the statistical noise becomes large. Still, the first results (Hamber and Williams, 1995) suggest that the potential has more or less the expected classical form in the vicinity of the critical point, both as far as the mass dependence and perhaps even the distance dependence are concerned. In particular it is attractive.

## 8.6 Scaling in the Vicinity of the Critical Point

In practice the correlation functions at fixed geodesic distance are difficult to compute numerically, and therefore not the best route to study the critical properties. But scaling arguments allow one to determine the scaling behavior of correlation functions from critical exponents characterizing the singular behavior of the *free energy* and various local averages in the vicinity of the critical point. In general a divergence of the correlation length $\xi$

$$\xi(k) \equiv \underset{k \to k_c}{\sim} A_\xi \, |k_c - k|^{-\nu} \quad , \tag{8.50}$$

signals the presence of a phase transition, and leads to the appearance of a singularity in the free energy $F(k)$. The scaling assumption for the free energy postulates that a divergent correlation length in the vicinity of the critical point at $k_c$ leads to non-analyticities of the type

$$\begin{aligned} F \equiv -\frac{1}{V} \ln Z = F_{reg} + F_{sing} \\ F_{sing} \sim \xi^{-d} \quad , \end{aligned} \tag{8.51}$$

where the second relationship follows simply from dimensional arguments (the free energy is an extensive quantity). The regular part $F_{reg}$ is generally not determined from $\xi$ by purely dimensional considerations, but as the name implies is a regular function in the vicinity of the critical point. Combining the definition of $\nu$ in Eq. (8.50) with the scaling assumption of Eq. (8.51) one obtains

$$F_{sing}(k) \underset{k \to k_c}{\sim} (\text{const.}) \, |k_c - k|^{d\nu} \quad . \tag{8.52}$$

The presence of a phase transition can then be inferred from non-analytic terms in invariant averages, such as the average curvature and its fluctuation. For the average curvature one obtains

$$\mathscr{R}(k) \underset{k \to k_c}{\sim} A_{\mathscr{R}} \, |k_c - k|^{d\nu - 1} \quad , \tag{8.53}$$

up to regular contributions (i.e. constant terms in the vicinity of $k_c$). An additive constant can be added, but numerical evidence sor far points to this constant being consistent with zero. Similarly one has for the curvature fluctuation

$$\chi_{\mathscr{R}}(k) \underset{k \to k_c}{\sim} A_{\chi_{\mathscr{R}}} \, |k_c - k|^{-(2-d\nu)} \quad . \tag{8.54}$$

At a critical point the fluctuation $\chi$ is in general expected to diverge, corresponding to the presence of a divergent correlation length. From such averages one can therefore in principle extract the correlation length exponent $\nu$ of Eq. (8.50) without having to compute a correlation function.

An equivalent result, relating the quantum expectation value of the curvature to the physical correlation length $\xi$ , is obtained from Eqs. (8.50) and (8.53)

$$\mathscr{R}(\xi) \underset{k \to k_c}{\sim} \xi^{1/\nu - 4} \quad , \tag{8.55}$$

again up to an additive constant. Matching of dimensionalities in this last equation is restored by inserting an appropriate power of the Planck length $l_P = \sqrt{G}$ on the r.h.s..

One can relate the critical exponent $\nu$ to the scaling behavior of correlations at large distances. The curvature fluctuation is related to the connected scalar curvature correlator at zero momentum

$$\chi_{\mathscr{R}}(k) \sim \frac{\int dx \int dy < \sqrt{g} R(x) \sqrt{g} R(y) >_c}{< \int dx \sqrt{g} >} \quad . \tag{8.56}$$

A divergence in the above fluctuation is then indicative of long range correlations, corresponding to the presence of a massless particle. Close to the critical point one expects for large separations $l_0 \ll |x-y| \ll \xi$ a power law decay in the geodesic distance, as in Eq. (8.31),

$$< \sqrt{g} R(x) \sqrt{g} R(y) > \underset{|x-y| \to \infty}{\sim} \frac{1}{|x-y|^{2n}} \quad . \tag{8.57}$$

Inserting the above expression in Eq. (8.56) and comparing with Eq. (8.54) determines the $n$ as $n = d - 1/\nu$. A priori one cannot exclude to possibility that some states acquire a mass away from the critical point, in which case the correlation functions would have the behavior of Eq. (8.32) for $|x-y| \gg \xi$.

## 8.7 Physical and Unphysical Phases

An important alternative to the analytic methods in the continuum is an attempt to solve quantum gravity directly via numerical simulations. The underlying idea is to evaluate the gravitational functional integral in the discretized theory $Z$ by summing over a suitable finite set of representative field configurations. In principle such a method given enough configurations and a fine enough lattice can provide an arbitrarily accurate solution to the original quantum gravity theory.

In practice there are several important factors to consider, which effectively limit the accuracy that can be achieved today in a practical calculation. Perhaps the most important one is the enormous amounts of computer time that such calculations can use up. This is particularly true when correlations of operators at fixed geodesic distance are evaluated. Another practical limitation is that one is mostly interested in the behavior of the theory in the vicinity of the critical point at $G_c$, where the correlation length $\xi$ can be quite large and significant correlations develop both between different lattice regions, as well as among representative field configurations, an effect known as critical slowing down. Finally, there are processes which are not well suited to a lattice study, such as problems with several different length (or energy) scales. In spite of these limitations, the progress in lattice field theory has been phenomenal in the last few years, driven in part by enormous advances in computer technology, and in part by the development of new techniques relevant to the problems of lattice field theories.

The starting point is the generation of a large ensemble of suitable edge length configurations. The edge lengths are updated by a straightforward Monte Carlo algorithm, generating eventually an ensemble of configurations distributed according to the action and measure of Eq. (6.91) (Hamber and Williams, 1984; Hamber, 1984; Berg, 1985); some more recent references are (Beirl et al, 1994; Riedler et al, 1999; Bittner et al, 2002). Further details of the method as applied to pure gravity can be found for example in the recent work (Hamber, 2000) and will not be repeated here.

As far as the lattice is concerned, one starts with the 4-d hypercube of Fig. 7.1 divided into simplices, and then stacks a number of such cubes in such a way as to construct an arbitrarily large lattice, as shown in Fig. 8.3. Other lattice structures are of course possible, including even a random lattice. The expectation is that for long range correlations involving distance scales much larger than the lattice spacing the precise structure of the underlying lattice structure will not matter.

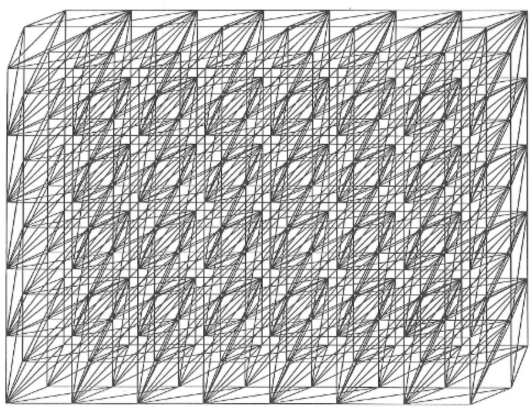


**Fig. 8.3** Four-dimensional hypercubes divided into simplices and stacked to form a four-dimensional lattice.

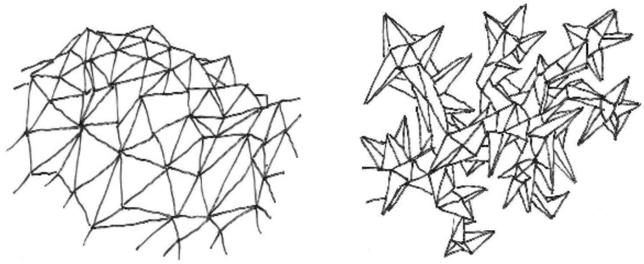


**Fig. 8.4** A pictorial description of the smooth (*left*) and rough (*right*) phases of four-dimensional lattice quantum gravity.

The lattice sizes investigated typically range from $4^4$ sites (3840 edges) to $32^4$ sites (15,728,640 edges). On a dedicated massively parallel supercomputer millions of consecutive edge length configurations can be generated for tens of values of $k$ in a few months time.

Even though these lattices are not very large, one should keep in mind that due to the simplicial nature of the lattice there are many edges per hypercube with many interaction terms, and as a consequence the statistical fluctuations can be comparatively small, unless measurements are taken very close to a critical point, and at rather large separation in the case of the correlation functions or the potential. In addition, extrapolations to the infinite volume limit can be aided by finite size scaling methods, which exploit predictable renormalization group properties of finite size systems.

Usually the topology is restricted to a four-torus, corresponding to periodic boundary conditions. One can perform similar calculations with lattices employing different boundary conditions or topology, but one would expect the universal scaling properties of the theory to be determined exclusively by short-distance renormalization effects. Indeed the Feynman rules of perturbation theory do not depend in any way on boundary terms, although some momentum integrals might require an infrared cutoff.

Based on physical considerations it would seem reasonable to impose the constraint that the scale of the curvature be much smaller than the inverse of the average lattice spacing, but still considerably larger than the inverse of the overall system size. Equivalently, that in momentum space physical scales should be much smaller that the ultraviolet cutoff, but much larger than the infrared cutoff. A typical requirement is therefore that

$$l_0 \ll \xi \ll L \,, \tag{8.58}$$

where $L$ is the linear size of the system, $\xi$ the correlation length related for example to the large scale curvature by $\mathscr{R} \sim 1/\xi^2$, and $l_0$ the lattice spacing. Contrary to ordinary lattice field theories, the lattice spacing in lattice gravity is a dynamical quantity. Thus the quantity $l_0 = \sqrt{<l^2>}$ only represents an average cutoff parameter.

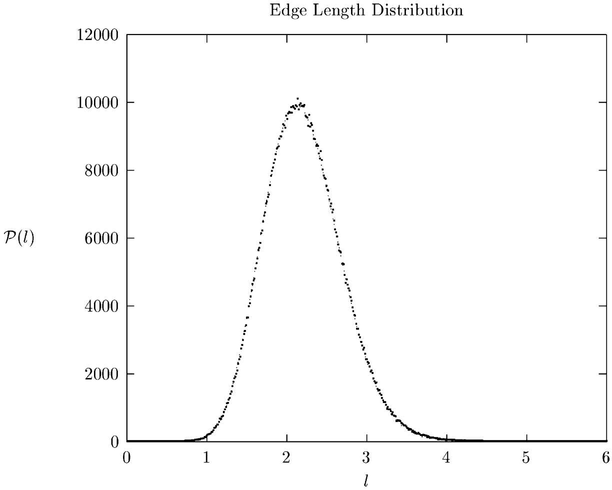



**Fig. 8.5** A typical edge length distribution in the smooth phase for which $k < k_c$, or $G > G_c$. Note that the lattice gravitational measure of Eq. (6.77) cuts off the distribution at small edge lengths, while the cosmological constant term prevents large edge lengths from appearing.

Furthermore the bare cosmological constant $\lambda_0$ appearing in the gravitational action of Eq. (6.91) can be fixed at 1 in units of the cutoff, since it just sets the overall length scale in the problem. The higher derivative coupling $a$ can be set to a value very close to 0 since one ultimately is interested in the limit $a \to 0$, corresponding to the pure Einstein theory.

One finds that for the measure in Eq. (6.77) this choice of parameters leads to a well behaved ground state for $k < k_c$ for higher derivative coupling $a \to 0$. The system then resides in the "smooth" phase, with an effective dimensionality close to four. On the other hand for $k > k_c$ the curvature becomes very large and the lattice collapses into degenerate configurations with very long, elongated simplices (see Fig. 8.4). Fig. 8.5 shows an example of a typical edge length distribution in the well behaved strong coupling phase close to but below $k_c$.

Fig. 8.6 shows the corresponding curvature ($\delta A$ or $\sqrt{g}R$) distribution.

One finds that as $k$ is varied, the average curvature $\mathscr{R}$ is negative for sufficiently small $k$ ("smooth" phase), and appears to go to zero continuously at some finite value $k_c$. For $k > k_c$ the curvature becomes very large, and the simplices tend to collapse into degenerate configurations with very small volumes ($<V> / <l^2>^2 \sim 0$). This "rough" or "collapsed" phase is the region of the usual weak field expansion ($G \to 0$). In this phase the lattice collapses into degenerate configurations with very long, elongated simplices (Hamber, 1984; Hamber and Williams, 1985; Berg, 1985). This phenomenon is usually intepreted as a lattice remnant of the conformal mode instability of Euclidean gravity discussed in Sect. (2.5).

An elementary argument can be given to explain the fact that the collapsed phase for $k > k_c$ has an effective dimension of two. The instability is driven by the Euclidean Einstein term in the action, and in particular its unbounded conformal mode

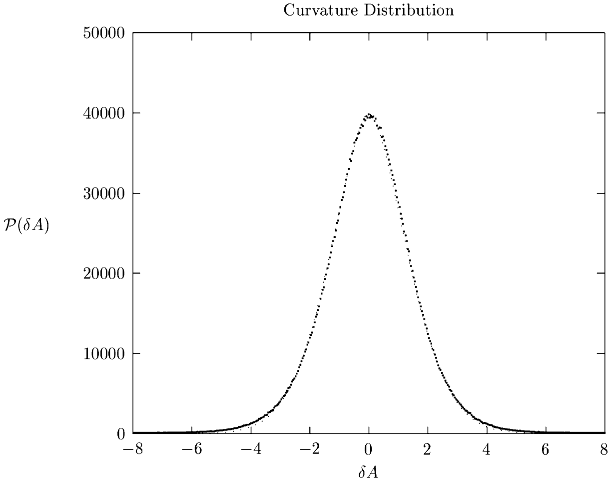



**Fig. 8.6** A typical curvature distribution in the smooth phase for which $k < k_c$, or $G > G_c$. Note that the distribution is peaked around a local curvature value which close to zero.

contribution. As the manifold during collapse reaches an effective dimension of two, the action effectively turns into a topological invariant, unable to drive the instability further to a still lower dimension[1].

Accurate and reproducible curvature data can only be obtained for $k$ below the instability point, since for $k > k_u \approx 0.053$ an instability develops, presumably associated with the unbounded conformal mode. Its signature is typical of a sharp first order transition, beyond which the system ventually tunnels into the rough, elongated phase which is two-dimensional in nature and has no physically acceptable continuum limit. The instability is caused by the appearance of one or more localized singular configuration, with a spike-like curvature singularity. At strong coupling such singular configurations are suppressed by a lack of phase space due to the functional measure, which imposes non-trivial constraints due to the triangle inequalities and their higher dimensional analogs. In other words, it seems that the measure regulates the conformal instability at sufficiently strong coupling.

---

[1] One way of determining coarse aspects of the underlying geometry is to compute the effective dimension in the scaling regime, for example by considering how the number of points within a thin shell of geodesic distance between $\tau$ and $\tau + \Delta$ scales with the geodesic distance itself. For distances a few multiples of the average lattice spacing one finds

$$N(\tau) \underset{\tau \to \infty}{\sim} \tau^{d_v} , \tag{8.59}$$

with $d_v = 3.1(1)$ for $G > G_c$ (the smooth phase) and $d_v \simeq 1.6(2)$ for $G < G_c$ (the rough phase). One concludes that in the rough phase the lattice tends to collapse into a degenerate tree-like configuration, whereas in the smooth phase the effective dimension of space-time is consistent with four. Higher derivative terms affect these results at very short distances, where they tend to make the geometry smoother.

As a consequence, the non-analytic behavior of the free energy (and its derivatives, which include for example the average curvature) has to be obtained by analytic continuation of the Euclidean theory into the metastable branch. This procedure, while perhaps unusual, is formally equivalent to the construction of the continuum theory exclusively from its strong coupling (small $k$, large $G$) expansion

$$\mathscr{R}(k) = \sum_{n=0}^{\infty} b_n k^n \ . \tag{8.60}$$

Ultimately it should be kept in mind that one is really only interested in the *pseudo-Riemannian* case, and not the Euclidean one for which an instability due to the conformal mode is to be expected. Indeed had such an instability not occurred for small enough $G$ one might have wondered if the resulting theory still had any relationship to the original continuum theory.

## 8.8 Numerical Determination of the Scaling Exponents

One way to extract the critical exponent $\nu$ is to fit the average curvature to the form [see Eq. (8.53)]

$$\mathscr{R}(k) \underset{k \to k_c}{\sim} -A_{\mathscr{R}} \, (k_c - k)^{\delta} \quad . \tag{8.61}$$

Using this set of procedures one obtains on lattices of up to $16^4$ sites $k_c = 0.0630(11)$ and $\nu = 0.330(6)$. Note that the average curvature is negative at strong coupling up to the critical point: locally the parallel transport of vectors around infinitesimal loops seems to be described by a lattice version of Euclidean anti-de Sitter space.

Fig. 8.7 shows as an example a graph of the average curvature $\mathscr{R}(k)$ raised to the third power. One would expect to get a straight line close to the critical point if the exponent for $\mathscr{R}(k)$ is exactly $1/3$. The numerical data indeed supports this assumption, and in fact the linearity of the results close to $k_c$ is quite striking. Using this procedure one obtains on the $16^4$-site lattice an esimate for the critical point, $k_c = 0.0639(10)$.

Often it can be advantageous to express results obtained in the cutoff theory in terms of physical (i.e. cutoff independent) quantities. By the latter one means quantities for which the cutoff dependence has been re-absorbed, or restored, in the relevant definition. As an example, an expression equivalent to Eq. (8.53), relating the vacuum expectation value of the local scalar curvature to the physical correlation length $\xi$ , is

$$\frac{< \int dx \sqrt{g} R(x) >}{< \int dx \sqrt{g} >} \underset{G \to G_c}{\sim} \text{const.} \left(l_P^2\right)^{(d-2-1/\nu)/2} \left(\frac{1}{\xi^2}\right)^{(d-1/\nu)/2} , \tag{8.62}$$

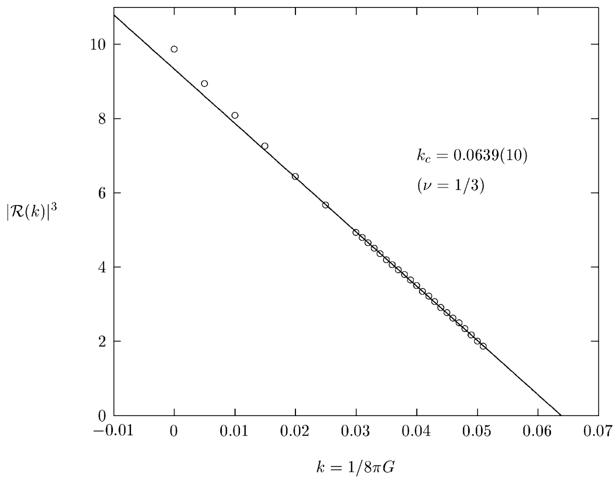



**Fig. 8.7** Average curvature $\mathscr{R}(k)$ on a $16^4$ lattice, raised to the third power. If $\nu = 1/3$, the data should fall on a straight line. The continuous line represents a linear fit of the form $A\,(k_c - k)$. The small deviation from linearity of the transformed data is quite striking.

which is obtained by substituting Eq. (8.50) into Eq. (8.53). The correct dimensions have been restored in this last equation by supplying appropriate powers of the Planck length $l_P = G_{phys}^{1/(d-2)}$, which involves the ultraviolet cutoff $\Lambda$. For $\nu = 1/3$ the result of Eq. (8.62) becomes particularly simple

$$\frac{< \int dx \sqrt{g} R(x) >}{< \int dx \sqrt{g} >} \underset{G \to G_c}{\sim} \text{const.} \; \frac{1}{l_P \, \xi} \; . \tag{8.63}$$

Note that a naive estimate based on dimensional arguments would have suggested the incorrect result $\sim 1/l_P^2$. Instead the above expression actually vanishes at the critical point. This shows that $\nu$ plays the role of an anomalous dimension, determining the magnitude of deviations from naive dimensional arguments.

Since the critical exponents play such a central role in determining the existence and nature of the continuum limit, it appears desirable to have an independent way of estimating them, which either does not depend on any fitting procedure, or at least analyzes a different and complementary set of data. By systematically studying the dependence of averages on the physical size of the system, one can independently estimate the critical exponents.

Reliable estimates for the exponents in a lattice field theory can take advantage of a comprehensive finite-size analysis, a procedure by which accurate values for the critical exponents are obtained by taking into account the linear size dependence of the result computed in a finite volume $V$.

In practice the renormalization group approach is brought in by considering the behavior of the system under scale transformations. Later the scale dependence is applied to the overall volume itself. The usual starting point for the derivation of the scaling properties is the renormalization group behavior of the free energy $F = -\frac{1}{V}\log Z$

$$F(t,\{u_j\}) \;=\; F_{reg}(t,\{u_j\}) \;+\; b^{-d}\,F_{sing}(b^{y_t}t,\{b^{y_j}u_j\}) \;\;, \tag{8.64}$$

where $F_{sing}$ is the singular, non-analytic part of the free energy, and $F_{reg}$ is the regular part. $b$ is the block size in the RG transformation, while $y_t$ and $y_j (j \geq 2)$ are the relevant eigenvalues of the RG transformation, and $t$ the reduced temperature variable that gives the distance from criticality. One denotes here by $y_t > 0$ the relevant eigenvalue, while the remaining eigenvalues $y_j \leq 0$ are associated with either marginal or irrelevant operators. Usually the leading critical exponent $y_t^{-1}$ is called $\nu$, while the next subleading exponent $y_2$ is denoted $-\omega$. For more details on the method we have to refer to the comprehensive review in (Cardy, 1988).

The correlation length $\xi$ determines the asymptotic decay of correlations, in the sense that one expects for example for the two-point function at large distances

$$< \mathscr{O}(x)\,\mathscr{O}(y) > \underset{|x-y|\,\gg\,\xi}{\sim} e^{-|x-y|/\xi} \quad . \tag{8.65}$$

The scaling equation for the correlation length itself

$$\xi\,(t) \;=\; b\,\xi\,(b^{y_t}t) \;\;, \tag{8.66}$$

implies for $b = t^{-1/y_t}$ that $\xi \sim t^{-\nu}$ with a correlation length exponent

$$\nu = 1/y_t \;\; . \tag{8.67}$$

Derivatives of the free energy $F$ with respect to $t$ then determine, after setting the scale factor $b = t^{-1/y_t}$, the scaling properties of physical observables, including corrections to scaling. Thus for example, the second derivative of the free energy with respect to $t$ yields the specific heat exponent $\alpha = 2 - d/y_t = 2 - d\nu$,

$$\frac{\partial^2}{\partial t^2}\,F(t,\{u_j\}) \;\sim\; t^{-(2-d\nu)} \quad . \tag{8.68}$$

In the gravitational case one identifies the scaling field $t$ with $k_c - k$, where $k = 1/16\pi G$ involves the bare Newton's constant. The appearance of singularities in physical averages, obtained from appropriate derivatives of $F$, is rooted in the fact that close to the critical point at $t = 0$ the correlation length diverges.

The above results can be extended to the case of a finite lattice of volume $V$ and linear dimension $L = V^{1/d}$. The volume-dependent free energy is then written as

$$F(t,\{u_j\},L^{-1}) \;=\; F_{reg}(t,\{u_j\}) \;+\; b^{-d}\,F_{sing}(b^{y_t}t,\{b^{y_j}u_j\},b/L) \;\; . \tag{8.69}$$

For $b = L$ (a lattice consisting of only one point) one obtains the Finite Size Scaling (FSS) form of the free energy [see for example (Brezin and Zinn-Justin, 1985) for a field-theoretic justification]. After taking derivatives with respect to the fields $t$ and $\{u_j\}$, the FSS scaling form for physical observables follows,

$$O(L,t) \;=\; L^{x_O/\nu} \, \left[ \tilde{f}_O(L\,t^{\nu}) \,+\, \mathcal{O}(L^{-\omega}) \right] \quad , \tag{8.70}$$

where $x_O$ is the scaling dimension of the operator $O$, and $\tilde{f}_O(x)$ an arbitrary function.

As an example, consider the average curvature $\mathscr{R}$. From Eq. (8.70), with $t \sim k_c - k$ and $x_O = 1 - 4\nu$, one has

$$\mathscr{R}(k,L) \;=\; L^{-(4-1/\nu)} \, \left[ \tilde{\mathscr{R}}\left( (k_c - k)\, L^{1/\nu} \right) \,+\, \mathcal{O}(L^{-\omega}) \right] \;, \tag{8.71}$$

where $\omega > 0$ is a correction-to-scaling exponent. If scaling involving $k$ and $L$ holds according to Eq. (8.70), then all points should lie on the same universal curve.

Fig. 8.8 shows a graph of the scaled curvature $\mathscr{R}(k)\, L^{4-1/\nu}$ for different values of $L = 4, 8, 16$, versus the scaled coupling $(k_c - k)L^{1/\nu}$. If scaling involving $k$ and $L$ holds according to Eq. (8.71), with $x_{\mathcal{O}} = 1 - 4\nu$ the scaling dimension for the curvature, then all points should lie on the same universal curve. The data is in good agreement with such behavior, and provides a further test on the exponent, which seems consistent within errors with $\nu = 1/3$.

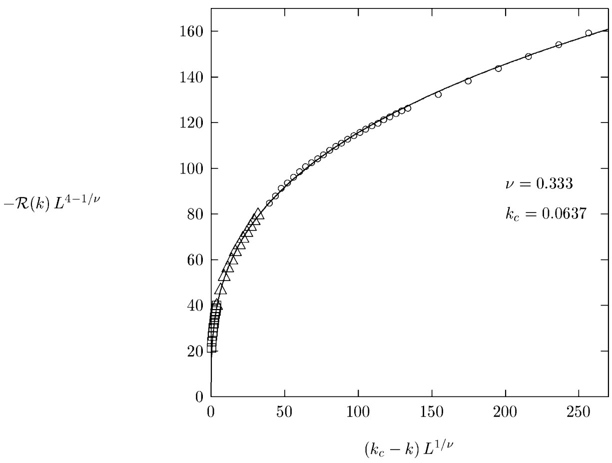



**Fig. 8.8** Finite size scaling behavior of the scaled curvature versus the scaled coupling. Here $L = 4$ for the lattice with $4^4$ sites (□), $L = 8$ for a lattice with $8^4$ sites (△), and $L = 16$ for the lattice with $16^4$ sites (○). Statistical errors are comparable to the size of the dots. The continuous line represents a best fit of the form $a + bx^c$. Finite size scaling predicts that all points should lie on the same universal curve. At $k_c = 0.0637$ the scaling plot gives the value $\nu = 0.333$.

As a second example consider the curvature fluctuation $\chi_{\mathscr{R}}$. From the general Eq. (8.70) one expects in this case, for $t \sim k_c - k$ and $x_O = 2 - 4\nu$,

$$\chi_{\mathscr{R}}(k,L) \;=\; L^{2/\nu-4}\left[\,\tilde{\chi_{\mathscr{R}}}\left((k_c-k)\,L^{1/\nu}\right)\;+\;\mathscr{O}(L^{-\omega})\,\right] \quad , \tag{8.72}$$

where $\omega > 0$ is again a correction-to-scaling exponent. If scaling involving $k$ and $L$ holds according to Eq. (8.70), then all points should lie again on the same universal curve.

The value of $k_c$ itself should depend on the size of the system. One expects

$$k_c(L) \underset{L\to\infty}{\sim} k_c(\infty) + c\,L^{-1/\nu} \;+\;\cdots \tag{8.73}$$

where $k_c(\infty)$ is the infinite-volume limit critical point.

The previous discussion applies to continuous, second order phase transitions. First order phase transitions are driven by instabilities, and are in general not governed by any renormalization group fixed point. The underlying reason is that the correlation length does not diverge at the first order transition point, and thus the system never becomes scale invariant. In the simplest case, a first order transition develops as the system tunnels between two neighboring minima of the free energy. In the metastable branch the free energy acquires a small complex part with a very weak essential singularity in the coupling at the first order transition point (Langer, 1967a,b; Fisher, 1967; Griffiths, 1969). As a consequence, such a singularity is not generally visible from the stable branch, in the sense that a power series expansion in the temperature is unaffected by such a weak singularity. Indeed in the infinite volume limit the singularities associated with a first order transition at $T_u$ become infinitely sharp, a $\theta$- or $\delta$-function type singularity. While the singularity in the free energy at the endpoint of the metastable branch (at say $T_c$) cannot be explored directly, it can be reached by a suitable analytic continuation from the stable side of the free energy branch. A similar situation arises in the case of lattice *QCD* with fermions, where zero fermion mass (chiral) limit is reached by extrapolation (Hamber and Parisi, 1983).

From the best data (with the smallest statistical uncertainties and the least systematic effects) one concludes

$$k_c = 0.0636(11) \qquad \nu = 0.335(9) \quad , \tag{8.74}$$

which suggests $\nu = 1/3$ for pure quantum gravity. Note that at the critical point the gravitational coupling is not weak, $G_c \approx 0.626$ in units of the ultraviolet cutoff. It seems that the value $\nu = 1/3$ does not correspond to any known field theory or statistical mechanics model in four dimensions. For a perhaps related system, namely dilute branched polymers, it is known that $\nu = 1/2$ in three dimensions, and $\nu = 1/4$ at the upper critical dimension $d = 8$, so one would expect a value close to $1/3$ somewhere in between. On the other hand for a scalar field one would have obtained $\nu = 1$ in $d = 2$ and $\nu = \frac{1}{2}$ for $d \geq 4$, which seems excluded.

Table 8.1 provides a short summary of the critical exponents for quantum gravitation as obtained by various perturbative and non-perturbative methods in three and four dimensions. The $2+\varepsilon$ and the truncation method results were discussed

| Method | $\nu^{-1}$ in $d=3$ | $\nu^{-1}$ in $d=4$ |
|---|---|---|
| lattice | 1.67(6) | - |
| lattice | - | 2.98(7) |
| $2+\varepsilon$ | 1.6 | 4.4 |
| truncation | 1.2 | 2.666 |
| exact ? | 1.5882 | 3 |

**Table 8.1** Direct determinations of the critical exponent $\nu^{-1}$ for quantum gravitation, using various analytical and numerical methods in three and four space-time dimensions.

previously in Sects. (3.5) and (3.6), respectively. The lattice model of Eq. (6.91) in four dimensions gives for the critical point $G_c \approx 0.626$ in units of the ultraviolet cutoff, and $\nu^{-1} = 2.98(7)$ which is used for comparison in Table 8.1. In three dimensions the numerical results are consistent with the universality class of the interacting scalar field. The same set of results are compared graphically in Fig. 8.9 and Fig. 8.10 below.

The direct numerical determinations of the critical point $k_c = 1/8\pi G_c$ in $d=3$ and $d=4$ space-time dimensions are in fact quite close to the analytical prediction of the lattice $1/d$ expansion given previously in Eq. (7.150),

$$\frac{k_c}{\lambda_0^{1-2/d}} = \frac{2^{1+2/d}}{d^3} \left[ \frac{\Gamma(d)}{\sqrt{d+1}} \right]^{2/d} . \tag{8.75}$$

The above expression gives for a bare cosmological constant $\lambda_0 = 1$ the estimate $k_c = \sqrt{3}/(16 \cdot 5^{1/4}) = 0.0724$ in $d=4$, to be compared with the numerical result $k_c = 0.0636(11)$ in (Hamber, 2000). Even in $d=3$ one has $k_c = 2^{5/3}/27 = 0.118$, to be compared with the direct determination $k_c = 0.112(5)$ from (Hamber and Williams, 1993). These estimates are compared below in Fig. 8.10.

## 8.9 Renormalization Group and Lattice Continuum Limit

The discussion in the previous sections points to the existence of a phase transition in the lattice gravity theory, with divergent correlation length in the vicinity of the critical point, as in Eq. (8.50)

$$\xi(k) \underset{k \to k_c}{\sim} A_\xi \, |k_c - k|^{-\nu} . \tag{8.76}$$

As described previously, the existence of such a correlation length is usually inferred indirectly by scaling arguments, from the presence of singularities in the free energy $F_{latt} = -\frac{1}{V} \ln Z_{latt}$ as a function of the lattice coupling $k$. Equivalently, $\xi$ could have been computed directly from correlation functions at fixed geodesic distance using

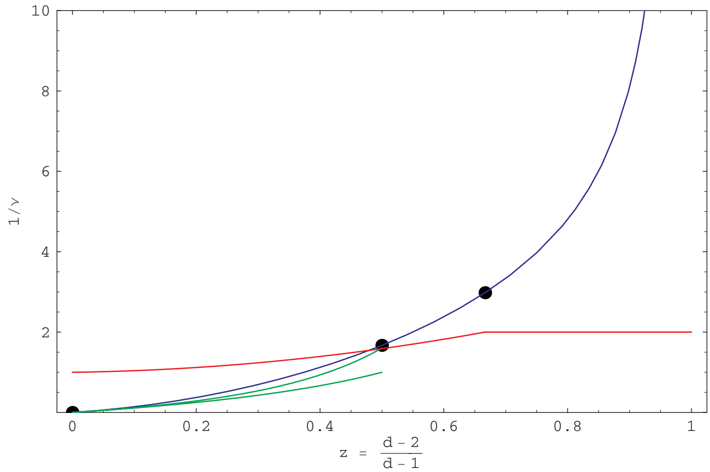



**Fig. 8.9** Universal gravitational exponent $1/\nu$ as a function of the dimension. The abscissa is $z = (d-2)/(d-1)$, which maps $d = 2$ to $z = 0$ and $d = \infty$ to $z = 1$. The larger circles at $d = 3$ and $d = 4$ are the lattice gravity results, interpolated (*continuous curve*) using the exact lattice results $1/\nu = 0$ in $d = 2$, and $\nu = 0$ at $d = \infty$ [from Eq. (7.159)]. The two curves close to the origin are the $2+\varepsilon$ expansion for $1/\nu$ to one loop (*lower curve*) and two loops (*upper curve*). The lower almost horizontal line gives the value for $\nu$ expected for a scalar field theory, for which it is known that $\nu = 1$ in $d = 2$ and $\nu = \frac{1}{2}$ in $d \geq 4$.

the definition in Eq. (8.32), or even from the correlation of Wilson lines associated with the propagation of two heavy spinless particles. The outcome of such large scale numerical calculations is eventually a determination of the quantities $\nu$, $k_c = 1/8\pi G_c$ and $A_\xi$ from first principles, to some degree of numerical accuracy.

In either case one expects the scaling result of Eq. (8.76) close to the fixed point, which we choose to rewrite here in terms of the inverse correlation length $m \equiv 1/\xi$

$$m = \Lambda A_m |k - k_c|^\nu \quad . \tag{8.77}$$

Note that in the above expression the correct dimension for $m$ (inverse length) has been restored by inserting explicitly on the r.h.s. the ultraviolet cutoff $\Lambda$. Here $k$ and $k_c$ are of course still dimensionless quantities, and correspond to the bare microscopic couplings at the cutoff scale, $k \equiv k(\Lambda) \equiv 1/[8\pi G(\Lambda)]$. $A_m$ is a calculable numerical constant, related to $A_\xi$ in Eq. (8.50) by $A_m = A_\xi^{-1}$. It is worth pointing out that the above expression for $m(k)$ is almost identical in structure to the one for the non-linear $\sigma$-model in the $2+\varepsilon$ expansion, Eq. (3.36) and in the large $N$ limit, Eqs. (3.59), (3.60) and (3.64). It is of course also quite similar to $2+\varepsilon$ result for continuum gravity, Eq. (3.121).

The lattice continuum limit corresponds to the large cutoff limit taken at *fixed m*,

$$\Lambda \to \infty \,, \quad k \to k_c \,, \quad m \text{ fixed} \,, \tag{8.78}$$

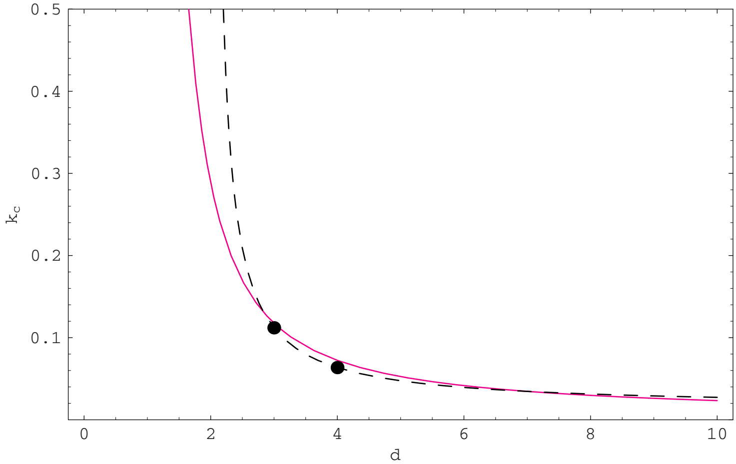



**Fig. 8.10** Critical point $k_c = 1/8\pi G_c$ in units of the ultraviolet cutoff as a function of dimension $d$. The circles at $d = 3$ and $d = 4$ are the lattice results, suitably interpolated (*dashed curve*) using the additional lattice result $1/k_c = 0$ in $d = 2$. The lower continuous curve is the analytical large-$d$ lattice result of Eq. (7.150).

which shows that the continuum limit is reached in the vicinity of the ultraviolet fixed point (see Fig. 8.11). Phrased equivalently, one takes the limit in which the lattice spacing $a \approx 1/\Lambda$ is sent to zero at *fixed* $\xi = 1/m$, which requires an approach to the non-trivial UV fixed point $k \to k_c$. The quantity $m$ is supposed to be a renormalization group invariant, a physical scale independent of the scale at which the theory is probed. In practice, since the cutoff ultimately determines the physical value of Newton's constant $G$, $\Lambda$ cannot be taken to $\infty$. Instead a very large value will suffice, $\Lambda^{-1} \sim 10^{-33} cm$, for which it will still be true that $\xi \gg \Lambda$ which is all that is required for the continuum limit.

For discussing the renormalization group behavior of the coupling it will be more convenient to write the result of Eq. (8.77) directly in terms of Newton's constant $G$ as

$$m = \Lambda \left(\frac{1}{a_0}\right)^{\nu} \left[\frac{G(\Lambda)}{G_c} - 1\right]^{\nu}, \qquad (8.79)$$

with the dimensionless constant $a_0$ related to $A_m$ by $A_m = 1/(a_0 k_c)^{\nu}$. Note that the above expression only involves the dimensionless ratio $G(\Lambda)/G_c$, which is the only relevant quantity here. The lattice theory in principle completely determines both the exponent $\nu$ and the amplitude $a_0$ for the quantum correction. Thus from the knowledge of the dimensionless constant $A_m$ in Eq. (8.77) one can estimate from first principles the value of $a_0$ in Eq. (8.84). Lattice results for the correlation functions at fixed geodesic distance give a value for $A_m \approx 0.72$ with a significant uncertainty, which, when combined with the values $k_c \simeq 0.0636$ and $\nu \simeq 0.335$ given above, gives $a_0 = 1/(k_c A_m^{1/\nu}) \simeq 42$. The rather surprisingly large value for $a_0$

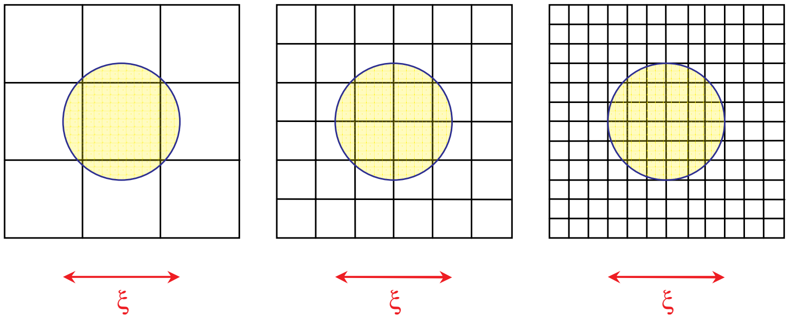



**Fig. 8.11** The lattice quantum continuum limit is gradually approached by considering sequences of lattices with increasingly larger correlation lengths $\xi$ in lattice units. Such a limit requires the existence of an ultraviolet fixed point, where quantum field correlations extend over many lattice spacing.

appears here as a consequence of the relatively small value of the lattice $k_c$ in four dimensions.

The renormalization group invariance of the physical quantity $m$ requires that the running gravitational coupling $G(\mu)$ varies in the vicinity of the fixed point in accordance with the above equation, with $\Lambda \to \mu$, where $\mu$ is now an arbitrary scale,

$$m = \mu \left( \frac{1}{a_0} \right)^{\nu} \left[ \frac{G(\mu)}{G_c} - 1 \right]^{\nu} . \tag{8.80}$$

The latter is equivalent to the renormalization group invariance requirement

$$\mu \frac{d}{d\mu} m[\mu, G(\mu)] = 0 \ , \tag{8.81}$$

provided $G(\mu)$ is varied in a specific way. Indeed this type of situation was already encountered before, for example in Eqs. (3.62) and (3.122). Eq. (8.81) can therefore be used to obtain, if one so wishes, *a* $\beta$-function for the coupling $G(\mu)$ in units of the ultraviolet cutoff,

$$\mu \frac{\partial}{\partial \mu} G(\mu) \ = \ \beta\,[G(\mu)] \ , \tag{8.82}$$

with $\beta(G)$ given in the vicinity of the non-trivial fixed point, using Eq. (8.80), by

$$\beta(G) \equiv \mu \frac{\partial}{\partial \mu} G(\mu) \underset{G \to G_c}{\sim} -\frac{1}{\nu}(G - G_c) + \dots \tag{8.83}$$

The above procedure is in fact in complete analogy to what was done for the non-linear $\sigma$-model, for example in Eq. (3.64). But the last two steps are not really necessary, for one can obtain the scale dependence of the gravitational coupling directly from Eq. (8.80), by simply solving for $G(\mu)$,

$$G(\mu) \ = \ G_c \left[ 1 + a_0 (m^2/\mu^2)^{1/2\nu} + \ldots \right] \ . \tag{8.84}$$

This last expression can be compared directly to the $2+\varepsilon$ result of Eq. (3.117), as well as to the $\sigma$-model result of Eq. (3.22). The physical dimensions of $G$ can be restored by multiplying the above expression on both sides by the ultraviolet cutoff $\Lambda$, if one so desires. One concludes that the above result physically implies gravitational anti-screening: the gravitational coupling $G$ increases with distance.

Note that the last equation only involves the dimensionless ratio $G(\mu)/G_c$, and is therefore unaffected by whether the coupling $G$ is dimensionful (after inserting an appropriate power of the cutoff $\Lambda$) or dimensionless. It simply relates the gravitational coupling at one scale to the coupling at a different scale,

$$\frac{G(\mu_1)}{G(\mu_2)} \ \approx \ \frac{1 + a_0 (m/\mu_1)^{1/\nu} + \ldots}{1 + a_0 (m/\mu_2)^{1/\nu} + \ldots} \ . \tag{8.85}$$

In conclusion, the lattice result for $G(\mu)$ in Eq. (8.84) and the $\beta$-function in Eq. (8.83) are qualitatively similar to what one finds both in the $2+\varepsilon$ expansion for gravity and in the non-linear $\sigma$-model *in the strong coupling phase*.

But there are also significant differences. Besides the existence of a phase transition between two geometrically rather distinct phases, one major new aspect provided by non-perturbative lattice studies is the fact that the weakly coupled small $G$ phase turns out to be pathological, in the sense that the theory becomes unstable, with the four-dimensional lattice collapsing into a tree-like two-dimensional structure for $G < G_c$. While in continuum perturbation theory both phases, and therefore both signs for the coupling constant evolution in Eq. (3.114), seem acceptable (giving rise to both a "Coulomb" phase, and a strong coupling phase), the Euclidean lattice results rule out the small $G < G_c$ branched polymer phase (Hamber, 1984; Berg, 1985). The collapse eventually stops at $d = 2$ because the gravitational action then becomes a topological invariant.

It appears difficult therefore to physically characterize the weak coupling phase based on just the lattice results, which only seem to make sense in the strong coupling phase $G > G_c$. One could envision an approach wherein such a weak coupling phase would be discussed in the framework of some sort of analytic continuation from the strong coupling phase, which would seem possible at least for some lattice results, such as the gravitational $\beta$-function of Eq. (8.83). The latter clearly makes sense on both sides of the transition, just as is the case for Eq. (3.64) for the non-linear $\sigma$-model. In particular Eq. (8.83) implies that the coupling will flow towards the Gaussian fixed point $G = 0$ for $G < G_c$. The scale dependence in this phase will be such that one expects gravitational screening: the coupling $G(\mu)$ gets increasingly weaker at larger distances. But how to remove the geometric collapse to a two-dimensional manifold remains a major hurdle; one could envision an approach where one introduces one more cutoff on the edges at short distances, so that each simplex cannot go below a certain fatness. But if the results so far can be used as a guide, the gradual removal of such a cutoff would then plunge the theory back into a degenerate two-dimensional, and therefore physically unacceptable, geometry.

## 8.10 Curvature Scales

As can be seen from Eqs. (3.79) and (8.21) the path integral for pure quantum gravity can be made to depend on the gravitational coupling $G$ and the cutoff $\Lambda$ only: by a suitable rescaling of the metric, or the edge lengths in the discrete case, one can set the cosmological constant to unity in units of the cutoff. The remaining coupling $G$ should then be viewed more appropriately as the gravitational constant *in units of the cosmological constant* $\lambda$.

The renormalization group running of $G(\mu)$ in Eq. (8.84) involves an invariant scale $\xi = 1/m$. At first it would seem that this scale could take any value, including very small ones based on the naive estimate $\xi \sim l_P$, which would preclude any observable quantum effects in the foreseeable future. But the result of Eqs. (8.62) and (8.63) suggest otherwise, namely that the non-perturbative scale $\xi$ is in fact related to *curvature*. From astrophysical observation the average curvature is very small, so one would conclude from Eq. (8.63) that $\xi$ is very large, and possibly macroscopic. But the problem with Eq. (8.63) is that it involves the lattice Ricci scalar, a quantity related curvature probed by parallel transporting vectors around infinitesimal loops with size comparable to the lattice cutoff $\Lambda^{-1}$. What one would like is instead a relationship between $\xi$ and quantities which describe the geometry on larger scales.

In many ways the quantity $m$ of Eq. (8.80) behaves as a dynamically generated mass scale, quite similar to what happens in the non-linear $\sigma$-model case [Eq. (3.60)], or in the $2+\varepsilon$ gravity case [Eq. (3.118)]. Indeed in the weak field expansion, presumably appropriate for slowly varying fields on very large scales, a mass-like term does appear, as in Eq. (1.79), with $\mu^2 = 16\pi G|\lambda_0| \equiv 2|\lambda|$ where $\lambda$ is the scaled cosmological constant. From the classical field equation $R = 4\lambda$ one can relate the above $\lambda$, and therefore the mass-like parameter $m$, to curvature, which leads to the identification

$$\lambda_{obs} \simeq \frac{1}{\xi^2} \, , \tag{8.86}$$

with $\lambda_{obs}$ the observed small but non-vanishing cosmological constant.

A further indication that the identification of the observed scaled cosmological constant with a mass-like - and thefore renormalization group invariant - term makes sense beyond the weak field limit can be seen for example by comparing the structure of the three classical field equations

$$\begin{aligned} R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R + \lambda\, g_{\mu\nu} &= 8\pi G T_{\mu\nu} \\ \partial^{\mu} F_{\mu\nu} + \mu^2 A_{\nu} &= 4\pi e\, j_{\nu} \\ \partial^{\mu} \partial_{\mu} \phi + m^2 \phi &= \frac{g}{3!} \phi^3 \, , \end{aligned} \tag{8.87}$$

for gravity, QED (massive via the Higgs mechanism) and a self-interacting scalar field, respectively.

A third argument suggesting the identification of the scale $\xi$ with large scale curvature and therefore with the observed scaled cosmological constant goes as follows. Observationally the curvature on large scale can be determined by parallel transporting vectors around very large loops, with typical size much larger than the lattice cutoff $l_P$. In gravity, curvature is detected by parallel transporting vectors around closed loops. This requires the calculation of a path dependent product of Lorentz rotations $\mathbf{R}$, in the Euclidean case elements of $SO(4)$, as discussed in Sect. 6.4. On the lattice, the above rotation is directly related to the path-ordered ($\mathscr{P}$) exponential of the integral of the lattice affine connection $\Gamma^{\lambda}_{\mu\nu}$ via

$$R^{\alpha}{}_{\beta} = \left[ \mathscr{P} e^{\int_{\substack{\text{path} \\ \text{between simplices}}} \Gamma^{\lambda} dx_{\lambda}} \right]^{\alpha}_{\beta} . \tag{8.88}$$

Now, in the strongly coupled gravity regime ($G > G_c$) large fluctuations in the gravitational field at short distances will be reflected in large fluctuations of the $\mathbf{R}$ matrices. Deep in the strong coupling regime it should be possible to describe these fluctuations by a uniform (Haar) measure. Borrowing from the analogy with Yang-Mills theories, and in particular non-Abelian lattice gauge theories with compact groups [see Eq. (3.145)], one would therefore expect an exponential decay of near-planar Wilson loops with area $A$ of the type

$$W(\Gamma) \sim \operatorname{tr}\exp\left[ \int_{S(C)} R^{\cdot}{}_{\cdot\mu\nu} A^{\mu\nu}_{C} \right] \sim \exp(-A/\xi^2) \ , \tag{8.89}$$

where $A$ is the minimal physical area spanned by the near-planar loop. A derivation of this standard result for non-Abelian gauge theories can be found, for example, in the textbook (Peskin and Schroeder, 1995).

In summary, the Wilson loop in gravity provides potentially a measure for the magnitude of the large-scale, averaged curvature, operationally determined by the process of parallel-transporting test vectors around very large loops, and which therefore, from the above expression, is computed to be of the order $R \sim 1/\xi^2$. One would expect the power to be universal, but not the amplitude, leaving open the possibility of having both de Sitter or anti-de Sitter space at large distances (as discussed previously in Sect. 8.8, the average curvature describing the parallel transport of vectors around *infinitesimal* loops is described by a lattice version of Euclidean anti-de Sitter space). A recent explicit lattice calculation indeed suggests that the de Sitter case is singled out, at least for sufficiently strong copuling (Hamber and Williams, 2007). Furthermore one would expect, based on general scaling arguments, that such a behavior would persists throughout the whole strong coupling phase $G > G_c$, all the way up to the on-trivial fixed point. From it then follows the identification of the correlation length $\xi$ with a measure of large scale curvature, the most natural candidate being the scaled cosmological constant $\lambda_{phys}$, as in Eq. (8.86). This relationship, taken at face value, implies a very large, cosmological value for $\xi \sim 10^{28} cm$, given the present bounds on $\lambda_{phys}$. Other closely related possibilities may exist, such as an identification of $\xi$ with the Hubble constant (as measured today), determining the macroscopic expansion rate of the universe via the correspondence $\xi \simeq 1/H_0$. Since this quantity is presumably time-dependent, a possible scenario would be one in which $\xi^{-1} = H_\infty = \lim_{t\to\infty} H(t)$, with $H_\infty^2 = \frac{\lambda}{3}$, for which the horizon radius is $R_\infty = H_\infty^{-1}$.

Since, as pointed out in Sects. (3.5) and (8.3), the gravitational path integral only depends in a non-trivial way on the dimensionless combination $G\sqrt{\lambda_0}$, the physical Newton's constant itself $G$ can be decomposed into non-running and running parts as

$$G = \frac{1}{G\lambda_0} \cdot G^2 \lambda_0 \;\to\; \xi^2 \cdot \left[ (G\sqrt{\lambda_0})\,(\mu^2) \right]^2 \;, \tag{8.90}$$

where we have used $1/G\lambda_0 \sim \xi^2$. The running of the second, dimensionless term in square brackets can be directly deduced from either Eqs. (8.80) or (8.84). Note that there $\lambda_0$ does not appear there explicitly, since originally it was set equal to one by scaling the metric (or the edge lengths).

In conclusion, the modified Einstein equations, incorporating the quantum running of $G$, read

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R + \lambda\, g_{\mu\nu} = 8\pi\, G(\mu)\, T_{\mu\nu} \;, \tag{8.91}$$

with $\lambda \simeq \frac{1}{\xi^2}$, and only $G(\mu)$ on the r.h.s. scale-dependent in accordance with Eq. (8.84). The precise meaning of $G(\mu)$ in a covariant framework will be given later in Sect. 9.2.

## 8.11 Gravitational Condensate

In strongly coupled gravity there appears to be a deep relationship, already encountered previously in non-Abelian gauge theories, between the non-perturbative scale $\xi$ appearing in Eq. (8.84), and the non-perturbative vacuum condensate of Eq. (8.62) and (8.86), which is a measure of curvature. The inescapable conclusion of the results of Eqs. (8.62) and (6.67) is that the scale $\xi$ appearing in Eq. (8.84) is related to curvature, and must be *macroscopic* for the lattice theory to be consistent. How can quantum effects propagate to such large distances and give such drastic modifications to gravity? The answer to this paradoxical question presumably lies in the fact that gravitation is carried by a massless particle whose interactions cannot be screened, on any length scale.

It is worth pointing out here that the gravitational vacuum condensate, which only exists in the strong coupling phase $G > G_c$, and which is proportional to the curvature, is genuinely non-perturbative. One can summarize the result of Eq. (8.86) as

$$\mathscr{R}_{obs} \simeq (10^{-30} eV)^2 \sim \xi^{-2} \;, \tag{8.92}$$

where the condensate is, according to Eq. (8.79), non-analytic at $G = G_c$. A graviton vacuum condensate of order $\xi^{-1} \sim 10^{-30} eV$ is of course extraordinarily small compared to the QCD color condensate ($\Lambda_{\overline{MS}} \simeq 220\, MeV$) and the electro-weak Higgs condensate ($v \simeq 250\, GeV$). One can pursue the analogy with non-Abelian gauge theories further by stating that the quantum gravity theory cannot provide a value for the non-perturbative curvature scale $\xi$: it needs to be fixed by some sort of phenomenological input, either by Eqs. (8.84) or by (8.86). But one important message is that the scale $\xi$ in those two equations is one and the same.

Can the above physical picture be used to provide further insight into the nature of the phase transition, and more specifically the value for $\nu$? We will mention here a simple geometric argument which can be given to support the exact value $\nu = 1/3$ for pure gravity (Hamber and Williams, 2004). First one notices that the vacuum polarization induced scale dependence of the gravitational coupling $G(r)$ as given in Eq. (8.84) implies the following general structure for the quantum corrected static gravitational potential,

$$V(r) = -G(r)\,\frac{mM}{r} \approx -G(0)\,\frac{mM}{r}\left[1 + c\,(r/\xi)^{1/\nu} + \mathscr{O}\left[(r/\xi)^{2/\nu}\right]\right] \ , \qquad (8.93)$$

for a point source of mass $M$ located at the origin and for intermediate distances $l_p \ll r \ll \xi$. One can visualize the above result by stating that virtual graviton loops cause an effective anti-screening of the primary gravitational source $M$, giving rise to a quantum correction to the potential proportional to $r^{1/\nu - 1}$. But only for $\nu = 1/3$ can the additional contribution be interpreted as being due to a close to uniform mass distribution surrounding the original source, of strength

$$\rho_\xi(M) = \frac{3cM}{4\pi\xi^3} \ . \qquad (8.94)$$

Such a simple geometric interpretation fails unless the exponent $\nu$ for gravitation is exactly one third. In fact in dimensions $d \geq 4$ one would expect based on the geometric argument that $\nu = 1/(d-1)$ if the quantum correction to the gravitational potential arises from such a virtual graviton cloud. These arguments rely of course on the lowest order result $V(r) \sim \int d^{d-1}k\, e^{ik\cdot x}/k^2 \sim r^{3-d}$ for single graviton exchange in $d > 3$ dimensions.

Equivalently, the running of $G$ can be characterized as being in part due to a tiny non-vanishing (and positive) non-perturbative gravitational vacuum contribution to the cosmological constant, with

$$\lambda_0(M) = \frac{3cM}{\xi^3} \ , \qquad (8.95)$$

and therefore an associated effective classical average curvature of magnitude $R_{class} \sim G\lambda_0 \sim GM/\xi^3$. It is amusing that for a very large mass distribution $M$, the above expression for the curvature can only be reconciled with the naive dimensional estimate $R_{class} \sim 1/\xi^2$, provided for the gravitational coupling $G$ itself one has $G \sim \xi/M$, which is reminiscent of Mach's principle and its connection with the Lense-Thirring effect (Lense and Thirring, 1918; Sciama, 1953; Feynman, 1962).

# Chapter 9
# Scale Dependent Gravitational Couplings

## 9.1 Renormalization Group and Scale Dependence of *G*

Non-perturbative studies of quantum gravity suggest the possibility that gravitational couplings might be weakly scale dependent due to nontrivial renormalization group effects. This would introduce a new gravitational scale, unrelated to Newton's constant, required in order to parametrize the gravitational running in the infrared region. If one is willing to accept such a scenario, then it seems difficult to find a compelling theoretical argument for why the non-perturbative scale entering the coupling evolution equations should be very small, comparable to the Planck length. One possibility is that the relevant non-perturbative scale is related to the curvature and therefore macroscopic in size, which could have observable consequences. One key ingredient in this argument is the relationship, in part supported by Euclidean lattice results combined with renormalization group arguments, between the scaling violation parameter and the scale of the average curvature.

## 9.2 Effective Field Equations

To summarize the results of the previous section, the result of Eq. (8.84) implies for the running gravitational coupling in the vicinity of the ultraviolet fixed point

$$G(k^2) \;=\; G_c \left[ 1 + a_0 \left( \frac{m^2}{k^2} \right)^{\frac{1}{2\nu}} + O[\,(m^2/k^2)^{\frac{1}{\nu}}\,] \right] \;, \tag{9.1}$$

with $m = 1/\xi$, $a_0 > 0$ and $\nu \simeq 1/3$. Since $\xi$ is expected to be very large, the quantity $G_c$ in the above expression should now be identified with the laboratory scale value $\sqrt{G_c} \sim 1.6 \times 10^{-33} cm$. Quantum corrections on the r.h.s. are therefore quite small as long as $k^2 \gg m^2$, which in real space corresponds to the "short distance" regime $r \ll \xi$.

The interaction in real space is often obtained by Fourier transform, and the above expression is singular as $k^2 \to 0$. The infrared divergence needs to be regulated, which can be achieved by utilizing as the lower limit of momentum integration $m = 1/\xi$. Alternatively, as a properly infrared regulated version of the above expression one can use

$$G(k^2) \simeq G_c \left[ 1 + a_0 \left( \frac{m^2}{k^2 + m^2} \right)^{\frac{1}{2\nu}} + \ldots \right] . \tag{9.2}$$

The last form for $G(k^2)$ will only be necessary in the regime where $k$ is small, so that one can avoid unphysical results. From Eq. (9.2) the gravitational coupling then approaches at very large distances $r \gg \xi$ the finite value $G_\infty = (1 + a_0 + \ldots) G_c$. Note though that in Eqs. (9.1) or (9.2) the cutoff no longer appear explicitly, it is absorbed into the definition of $G_c$. In the following we will be mostly interested in the regime $l_P \ll r \ll \xi$, for which Eq. (9.1) is completely adequate.

The first step in analyzing the consequences of a running of $G$ is to re-write the expression for $G(k^2)$ in a coordinate-independent way. The following methods are not new, and have found over the years their fruitful application in gauge theories and gravity, for example in the discussion of non-local effective actions (Vilkovisky, 1984; Barvinsky and Vilkovisky, 1985). Since in going from momentum to position space one usually employs $k^2 \to -\Box$, to obtain a quantum-mechanical running of the gravitational coupling one should make the replacement

$$G \;\to\; G(\Box) \ , \tag{9.3}$$

and therefore from Eq. (9.1)

$$G(\Box) = G_c \left[ 1 + a_0 \left( \frac{1}{\xi^2 \Box} \right)^{\frac{1}{2\nu}} + \ldots \right] . \tag{9.4}$$

In general the form of the covariant d'Alembertian operator $\Box$ depends on the specific tensor nature of the object it is acting on,

$$\Box \, T^{\alpha\beta\ldots}{}_{\gamma\delta\ldots} = g^{\mu\nu} \nabla_\mu \left( \nabla_\nu \, T^{\alpha\beta\ldots}{}_{\gamma\delta\ldots} \right) . \tag{9.5}$$

Only on scalar functions one has the fairly simple result

$$\Box \, S(x) = \frac{1}{\sqrt{g}} \, \partial_\mu \, g^{\mu\nu} \sqrt{g} \, \partial_\nu \, S(x) \ , \tag{9.6}$$

whereas on second rank tensors one has the already significantly more complicated expression $\Box T_{\alpha\beta} \equiv g^{\mu\nu} \nabla_\mu (\nabla_\nu T_{\alpha\beta})$.

The running of $G$ is expected to lead to a non-local gravitational action, for example of the form

$$I = \frac{1}{16\pi G} \int dx \sqrt{g} \left( 1 - a_0 \left( \frac{1}{\xi^2 \Box} \right)^{1/2\nu} + \dots \right) R \ . \tag{9.7}$$

Due to the fractional exponent in general the covariant operator appearing in the above expression, namely

$$A(\Box) = a_0 \left( \frac{1}{\xi^2 \Box} \right)^{1/2\nu} \ , \tag{9.8}$$

has to be suitably defined by analytic continuation from positive integer powers. The latter can be done for example by computing $\Box^n$ for positive integer $n$ and then analytically continuing to $n \to -1/2\nu$. Alternatively one can make use of the identity

$$\frac{1}{\Box^n} = \frac{(-1)^n}{\Gamma(n)} \int_0^\infty ds\, s^{n-1} \exp(is\Box) \ , \tag{9.9}$$

and later perform the relevant integrals with $n \to 1/2\nu$. Other procedures can be used to define $A(\Box)$, for example based on an integral representation involving the scalar propagator (Lopez Nacir and Mazzitelli, 2007).

It should be stressed here that the action in Eq. (9.7) should be treated as a *classical* effective action, with dominant radiative corrections at short distances ($r \ll \xi$) already automatically built in, and for which a restriction to generally smooth field configurations does make some sense. In particular one would expect that in most instances it should be possible, as well as meaningful, to neglect terms involving large numbers of derivatives of the metric in order to compute the effects of the new contributions appearing in the effective action.

Had one *not* considered the action of Eq. (9.7) as a starting point for constructing the effective theory, one would naturally be led [following Eq. (9.3)] to consider the following effective field equations

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R + \lambda\, g_{\mu\nu} = 8\pi G \left[1 + A(\Box)\right] T_{\mu\nu} \ , \tag{9.10}$$

the argument again being the replacement $G \to G(\Box) \equiv G[1 + A(\Box)]$. Being manifestly covariant, these expressions at least satisfy some of the requirements for a set of consistent field equations incorporating the running of $G$. The above effective field equation can in fact be re-cast in a form similar to the classical field equations

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R + \lambda\, g_{\mu\nu} = 8\pi G \tilde{T}_{\mu\nu} \ , \tag{9.11}$$

with $\tilde{T}_{\mu\nu} = [1 + A(\Box)]\, T_{\mu\nu}$ defined as an effective, or gravitationally dressed, energy-momentum tensor. Just like the ordinary Einstein gravity case, in general $\tilde{T}_{\mu\nu}$ might not be covariantly conserved a priori, $\nabla^\mu \tilde{T}_{\mu\nu} \neq 0$, but ultimately the consistency of the effective field equations demands that it be exactly conserved, in consideration of the Bianchi identity satisfied by the Riemann tensor [a similar problem arises in other non-local modifications of gravity (Barvinsky, 2003)]. The ensuing new covariant conservation law

$$\nabla^\mu \tilde{T}_{\mu\nu} \equiv \nabla^\mu \left[ [1+A(\Box)]\, T_{\mu\nu} \right] = 0 \ , \tag{9.12}$$

can be then be viewed as a *constraint* on $\tilde{T}_{\mu\nu}$ (or $T_{\mu\nu}$) which, for example, in the specific case of a perfect fluid, will imply again a definite relationship between the density $\rho(t)$, the pressure $p(t)$ and the RW scale factor $a(t)$, just as it does in the standard case. Then the requirement that the bare energy momentum-tensor be conserved would imply that the quantum contribution $A(\Box)\, T_{\mu\nu}$ itself be separately conserved. That this is indeed attainable can be shown in a few simple cases, such as the static isotropic solution discussed below. There a "vacuum fluid" is introduced to account for the vacuum polarization contribution, whose energy momentum tensor can be shown to be covariantly conserved. That the procedure is consistent in general is not clear, in which case the present approach should perhaps be limited to phenomenological considerations.

Let us make a few additional comments regarding the above effective field equations, in which we will set the cosmological constant $\lambda = 0$ from now on. One simple observation is that the trace equation only involves the (simpler) scalar d'Alembertian, acting on the trace of the energy-momentum tensor

$$R = -8\pi G\, [1+A(\Box)]\, T_\mu{}^\mu \ . \tag{9.13}$$

Finally, to the order one is working here, the above effective field equations should be equivalent to

$$\left(1 - A(\Box) + O(A(\Box)^2)\right) \left(R_{\mu\nu} - \tfrac{1}{2}\, g_{\mu\nu} R\right) = 8\pi G\, T_{\mu\nu} \ , \tag{9.14}$$

where the running of $G$ has been moved over to the "gravitational" side.

## 9.3 Poisson's Equation and Vacuum Polarization Cloud

One of the simplest cases to analyze is of course the static case. The non-relativistic, static Newtonian potential is defined as usual as

$$\phi(r) = (-M) \int \frac{d^3\mathbf{k}}{(2\pi)^3}\, e^{i\mathbf{k}\cdot\mathbf{x}}\, G(\mathbf{k^2})\, \frac{4\pi}{\mathbf{k^2}} \ , \tag{9.15}$$

and therefore proportional to the $3-d$ Fourier transform of

$$\frac{4\pi}{\mathbf{k^2}} \to \frac{4\pi}{\mathbf{k^2}} \left[ 1 + a_0 \left( \frac{m^2}{\mathbf{k^2}} \right)^{\frac{1}{2\nu}} + \dots \right] \ . \tag{9.16}$$

But, as already mentioned, for small $\mathbf{k}$ proper care has to be exercised in providing a properly infrared regulated version of the above expression, which, from Eq. (9.2), reads

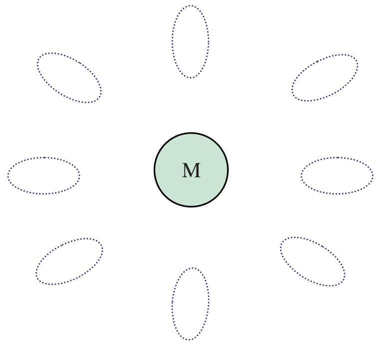



**Fig. 9.1** A virtual graviton cloud surrounds the point source of mass $M$, leading to an anti-screening modification of the static gravitational potential. This antiscreening effect of vacuum fluctuations is quite natural in gravity, since the larger the cloud is, the stronger the gravitational force is expected to be.

$$\frac{4\pi}{(\mathbf{k^2}+\mu^2)} \;\rightarrow\; \frac{4\pi}{(\mathbf{k^2}+\mu^2)}\left[1+a_0\left(\frac{m^2}{\mathbf{k^2}+m^2}\right)^{\frac{1}{2\nu}}+\ldots\right] \;, \tag{9.17}$$

where the limit $\mu \to 0$ should be taken at the end of the calculation.

Given the running of $G$ from either Eqs. (9.2) or (9.1) in the large $\mathbf{k}$ limit, the next step is naturally an attempt at finding a solution to Poisson's equation with a point source at the origin, so that one can determine the structure of the quantum corrections to the static gravitational potential in real space. There are in principle two equivalent ways to compute the potential $\phi(r)$, either by inverse Fourier transform of Eq. (9.16), or by solving Poisson's equation $\Delta\phi = 4\pi\rho$ with the source term $\rho(r)$ given by the inverse Fourier transform of the correction to $G(k^2)$, as given below in Eq. (9.20). The zero-th order term then gives the standard Newtonian $-MG/r$ term, while the correction in general is given by a rather complicated hypergeometric function. But for the special case $\nu = 1/2$ the Fourier transform of Eq. (9.17) is easy to do, the integrals are elementary and the running of $G(r)$ so obtained is particularly transparent,

$$G(r) \;=\; G_\infty\left(1-\frac{a_0}{1+a_0}\,e^{-mr}\right) \;, \tag{9.18}$$

where we have set $G_\infty \equiv (1+a_0)\,G$ and $G \equiv G(0)$. $G$ therefore increases slowly from its value $G$ at small $r$ to the larger value $(1+a_0)\,G$ at infinity. Fig. 9.1 illustrates the anti-screening effect of the virtual graviton cloud. Fig. 9.2 gives a schematic illustration of the behavior of $G$ as a function of $r$.

Another possible procedure to obtain the static potential $\phi(r)$ is to solve directly the radial Poisson equation for $\phi(r)$. This will give a density $\rho(r)$ which can later be used to generalize to the relativistic case. In the $a_0 \neq 0$ case one needs to solve $\Delta\phi = 4\pi\rho$, or in the radial coordinate for $r > 0$

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\phi}{dr}\right) = 4\pi G\rho_m(r) \ , \tag{9.19}$$

with the source term $\rho_m$ determined from the inverse Fourier transform of the correction term in Eq. (9.17). The latter is given by

$$\rho_m(r) = \frac{1}{8\pi} c_\nu a_0 M m^3 (mr)^{-\frac{1}{2}(3-\frac{1}{\nu})} K_{\frac{1}{2}(3-\frac{1}{\nu})}(mr) \ , \tag{9.20}$$

with the constant

$$c_\nu \equiv \frac{2^{\frac{1}{2}(5-\frac{1}{\nu})}}{\sqrt{\pi}\,\Gamma(\frac{1}{2\nu})} \ . \tag{9.21}$$

One can verify that the vacuum polarization density $\rho_m$ has the property

$$4\pi\int_0^\infty r^2\,dr\,\rho_m(r) = a_0 M \ , \tag{9.22}$$

where the standard integral $\int_0^\infty dx x^{2-n} K_n(x) = 2^{-n}\sqrt{\pi}\Gamma\left(\frac{3}{2}-n\right)$ has been used. Note that the gravitational vacuum polarization distribution is singular close to $r=0$, just as in QED, Eq. (3.138).

The $r\to 0$ result for $\phi(r)$ (discussed in the following, as an example, for $\nu = 1/3$) can then be obtained by solving the radial equation for $\phi(r)$,

$$\frac{1}{r}\frac{d^2}{dr^2}\left[r\phi(r)\right] = \frac{2a_0 M G m^3}{\pi} K_0(mr) \ , \tag{9.23}$$

where the (modified) Bessel function is expanded out to lowest order in $r$, $K_0(mr) = -\gamma - \ln\left(\frac{mr}{2}\right) + O(m^2 r^2)$, giving

$$\phi(r) = -\frac{MG}{r} + a_0 M G m^3 \frac{r^2}{3\pi}\left[-\ln\left(\frac{mr}{2}\right) - \gamma + \frac{5}{6}\right] + O(r^3) \ , \tag{9.24}$$

where the two integration constants are matched to the general large $r$ solution

$$\phi(r) \underset{r\to\infty}{\sim} -\frac{MG}{r}\left[1 + a_0\left(1 - c_l\,(mr)^{\frac{1}{2\nu}-1} e^{-mr}\right)\right] \ , \tag{9.25}$$

with $c_l = 1/[\nu\, 2^{\frac{1}{2\nu}}\,\Gamma(\frac{1}{2\nu})]$. Note again that the vacuum polarization density $\rho_m(r)$ has the expected normalization property

$$4\pi\int_0^\infty r^2\,dr\,\frac{a_0 M m^3}{2\pi^2} K_0(mr) = \frac{2a_0 M m^3}{\pi}\cdot\frac{\pi}{2m^3} = a_0 M \ , \tag{9.26}$$

so that the total enclosed additional gravitational charge is indeed just $a_0 M$, and $G_\infty = G_0(1+a_0)$.

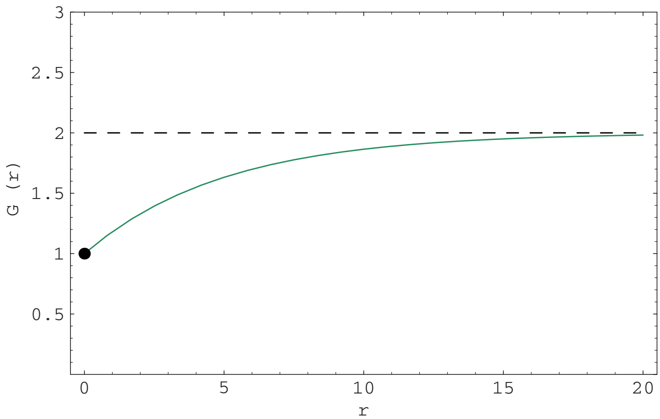



**Fig. 9.2** Schematic scale dependence of the gravitational coupling $G(r)$, from Eq. (9.18) valid for $\nu = 1/2$. The gravitational coupling rises initially like a power of $r$, and later approaches the asymptotic value $G_\infty = (1+a_0)G$ for large $r$. The behavior for other values of $\nu > 1/3$ is similar.

## 9.4 Static Isotropic Solution

The discussion of the previous section suggests that the quantum correction due to the running of $G$ can be described, at least in the non-relativistic limit of Eq. (9.2) as applied to Poisson's equation, in terms of a vacuum energy density $\rho_m(r)$, distributed around the static source of strength $M$ in accordance with the result of Eqs. (9.20) and (9.22).

In general a manifestly covariant implementation of the running of $G$, via the $G(\Box)$ given in Eq. (9.4), will induce a non-vanishing effective pressure term. It is natural therefore to attempt to represent the vacuum polarization cloud by a relativistic perfect fluid, with energy-momentum tensor

$$T_{\mu\nu} \;=\; (\,p+\rho\,)\,u_\mu\,u_\nu + g_{\mu\nu}\,p \;\;, \tag{9.27}$$

which in the static isotropic case reduces to

$$T_{\mu\nu} \;=\; \mathrm{diag}\left[B(r)\,\rho(r), A(r)\,p(r), r^2\,p(r), r^2\,\sin^2\theta\,p(r)\right] \;, \tag{9.28}$$

and gives a trace $T = 3\,p - \rho$. The $tt$, $rr$ and $\theta\theta$ components of the field equations then read

$$-\lambda B(r) + \frac{A'(r)B(r)}{rA(r)^2} - \frac{B(r)}{r^2A(r)} + \frac{B(r)}{r^2} \;=\; 8\pi G B(r)\rho(r) \tag{9.29}$$

$$\lambda A(r) - \frac{A(r)}{r^2} + \frac{B'(r)}{rB(r)} + \frac{1}{r^2} \;=\; 8\pi G A(r)p(r) \tag{9.30}$$

$$-\frac{B'(r)^2 r^2}{4A(r)B(r)^2}+\lambda r^2-\frac{A'(r)B'(r)r^2}{4A(r)^2B(r)}+\frac{B''(r)r^2}{2A(r)B(r)}-\frac{A'(r)r}{2A(r)^2}+\frac{B'(r)r}{2A(r)B(r)} \;=\; 8G\pi r^2 p(r) \;, \tag{9.31}$$

with the $\varphi\varphi$ component equal to $\sin^2\theta$ times the $\theta\theta$ component. Covariant energy conservation $\nabla^\mu T_{\mu\nu}=0$ implies

$$[p(r)+\rho(r)]\,\frac{B'(r)}{2B(r)}+p'(r) \;=\; 0 \;, \tag{9.32}$$

and forces a definite relationship between $B(r)$, $\rho(r)$ and $p(r)$. The three field equations and the energy conservation equation are, as usual, not independent, because of the Bianchi identity. It seems reasonable to attempt to solve the above equations [usually considered in the context of relativistic stellar structure (Misner, Thorne and Wheeler, 1973)] with the density $\rho(r)$ given by the $\rho_m(r)$ of Eq. (9.20). This of course raises the question of how the relativistic pressure $p(r)$ should be chosen, an issue that the non-relativistic calculation did not have to address. One finds that covariant energy conservation in fact completely determines the pressure in the static case, leading to consistent equations and solutions (note that in particular it would not be consistent to take $p(r)=0$).

Since the function $B(r)$ drops out of the $tt$ field equation, the latter can be integrated immediately, giving

$$A(r)^{-1} \;=\; 1-\frac{2MG}{r}-\frac{\lambda}{3}r^2-\frac{8\pi G}{r}\int_0^r dx x^2\rho(x) \;. \tag{9.33}$$

It is natural to identify $c_1=-2MG$, which of course corresponds to the solution with $a_0=0$ ($p=\rho=0$). Next, the $rr$ field equation can be solved for $B(r)$,

$$B(r) \;=\; \exp\left\{c_2-\int_{r_0}^r dy\,\frac{1+A(y)\left(\lambda y^2-8\pi G y^2 p(y)-1\right)}{y}\right\} \;, \tag{9.34}$$

with the constant $c_2$ again determined by the requirement that the above expression for $B(r)$ reduce to the standard Schwarzschild solution for $a_0=0$ ($p=\rho=0$), giving $c_2=\ln(1-2MG/r_0-\lambda r_0^2/3)$. The last task left therefore is the determination of the pressure $p(r)$. One needs to solve the equation

$$p'(r)+\frac{\left[8\pi G r^3 p(r)+2MG-\frac{2}{3}\lambda r^3+8\pi G\int_{r_0}^r dx x^2\rho(x)\right]\,[p(r)+\rho(r)]}{2r\left(r-2MG-\frac{\lambda}{3}r^3-8\pi G\int_0^r dx x^2\rho(x)\right)} \;=\; 0 \;, \tag{9.35}$$

which is usually referred to as the equation of hydrostatic equilibrium. From now on we will focus only the case $\lambda=0$. The last differential equation can be solved for $p(r)$,

$$p_m(r) \;=\; \frac{1}{\sqrt{1-\frac{2MG}{r}}}\left(c_3-\int_{r_0}^r dz\,\frac{MG\rho(z)}{z^2\sqrt{1-\frac{2MG}{z}}}\right) \;, \tag{9.36}$$

where the constant of integration has to be chosen so that when $\rho(r) = 0$ (no quantum correction) one has $p(r) = 0$ as well. Because of the singularity in the integrand at $r = 2MG$, we will take the lower limit in the integral to be $r_0 = 2MG + \varepsilon$, with $\varepsilon \to 0$.

To proceed further, one needs the explicit form for $\rho_m(r)$, which was given in Eq. (9.20),

$$\rho_m(r) \;=\; \frac{1}{8\pi}\, c_\nu\, a_0\, M\, m^3\, (m\, r)^{-\frac{1}{2}(3-\frac{1}{\nu})}\, K_{\frac{1}{2}(3-\frac{1}{\nu})}(m\, r) \;\; . \tag{9.37}$$

The required integrands involve for general $\nu$ the modified Bessel function $K_n(x)$, which can lead to rather complicated expressions for the general $\nu$ case. To determine the pressure, one supposes that it as well has a power dependence on $r$ in the regime under consideration, $p_m(r) = c_p A_0\, r^\gamma$, where $c_p$ is a numerical constant, and then substitute $p_m(r)$ into the pressure equation of Eq. (9.35). This gives, past the horizon $r \gg 2MG$ the algebraic condition

$$(2\gamma - 1)\, c_p M G r^{\gamma-1} - c_p\, \gamma r^\gamma - M G r^{1/\nu-4} \simeq 0 \;\; , \tag{9.38}$$

giving the same power $\gamma = 1/\nu - 3$ as for $\rho(r)$, $c_p = -1$ and surprisingly also $\gamma = 0$, implying that in this regime only $\nu = 1/3$ gives a consistent solution.

The case $\nu = 1/3$ can be dealt with separately, starting from the expression for $\rho_m(r)$ for $\nu = 1/3$

$$\rho_m(r) \;=\; \frac{1}{2\pi^2}\, a_0\, M\, m^3\, K_0(m\, r) \;\; . \tag{9.39}$$

One has for small $r$

$$\rho_m(r) \;=\; -\frac{a_0}{2\pi^2}\, M\, m^3 \left( \ln \frac{m\, r}{2} + \gamma \right) + \dots \tag{9.40}$$

and consequently

$$A^{-1}(r) \;=\; 1 - \frac{2MG}{r} + \frac{4\, a_0\, M G m^3}{3\,\pi}\, r^2 \ln (m\, r) + \dots \tag{9.41}$$

From Eq. (9.35) one can then obtain an expression for the pressure $p_m(r)$, and one finds again in the limit $r \gg 2MG$

$$p_m(r) \;=\; \frac{a_0}{2\pi^2}\, M\, m^3 \ln (m\, r) + \dots \tag{9.42}$$

After performing the required $r$ integral in Eq. (9.34), and evaluating the resulting expression in the limit $r \gg 2MG$, one obtains

$$B(r) \;=\; 1 - \frac{2MG}{r} + \frac{4\, a_0\, M G m^3}{3\,\pi}\, r^2 \ln (m\, r) + \dots \tag{9.43}$$

It is encouraging to note that in the solution just obtained the running of $G$ is the *same* in $A(r)$ and $B(r)$. The expressions for $A(r)$ and $B(r)$ are consistent with a gradual slow increase in $G$ with distance, in accordance with the formula

$$G \to G(r) = G\left(1 + \frac{a_0}{3\pi} m^3 r^3 \ln \frac{1}{m^2 r^2} + \ldots\right) , \tag{9.44}$$

in the regime $r \gg 2MG$, and therefore of course in agreement with the original result of Eqs. (9.1) or (9.2), namely that the classical laboratory value of $G$ is obtained for $r \ll \xi$. Note that the correct relativistic small $r$ correction of Eq. (9.44) agrees roughly in magnitude (but not in sign) with the approximate non-relativistic, Poisson equation result of Eq. (9.24).

There are similarities, as well as some rather substantial differences, with the corresponding QED result of Eq. (3.138). In the gravity case, the correction vanishes as $r$ goes to zero: in this limit one is probing the bare mass, unobstructed by its virtual graviton cloud. On the other hand, in the QED case, as one approaches the source one is probing the bare charge, whose magnitude diverges logarithmically for small $r$.

Finally it should be recalled that neither function $A(r)$ or $B(r)$ are directly related to the relativistic potential for particle orbits, which is given instead by the combination

$$V_{eff}(r) = \frac{1}{2A(r)}\left[\frac{l^2}{r^2} - \frac{1}{B(r)} + 1\right] , \tag{9.45}$$

where $l$ is proportional to the orbital angular momentum of the test particle, as discussed for example in (Hartle, 2003).

The running $G$ term acts in a number of ways as a local cosmological constant term, for which the $r$ dependence of the vacuum solution for small $r$ is fixed by the nature of the Schwarzschild solution with a cosmological constant term. One can therefore wonder what the solutions might look like in $d$ dimensions. In $d \geq 4$ dimensions the Schwarzschild solution to Einstein gravity with a cosmological term is (Myers and Perry, 1986)

$$A^{-1}(r) = B(r) = 1 - 2MGc_d r^{3-d} - \frac{2\lambda}{(d-2)(d-1)} r^2 , \tag{9.46}$$

with $c_d = 4\pi\Gamma(\frac{d-1}{2})/(d-2)\pi^{\frac{d-1}{2}}$, which would suggest, in analogy with the results for $d = 4$ given above that in $d \geq 4$ dimensions only $\nu = 1/(d-1)$ is possible. This last result would also be in agreement with the exact value $\nu = 0$ found at $d = \infty$ in Sect. 7.6.

## 9.5 Cosmological Solutions

A scale dependent Newton's constant will lead to small modifications of the standard cosmological solutions to the Einstein field equations. Here we will provide a brief discussion of what modifications are expected from the effective field equations on the basis of $G(\Box)$, as given in Eq. (9.3), which itself originates in Eqs. (9.2) and (9.1) (Hamber and Williams, 2005a,b).

The starting point is the quantum effective field equations of Eq. (9.10),

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R + \lambda\, g_{\mu\nu} \;=\; 8\pi G\, [1 + A(\Box)]\, T_{\mu\nu} \;, \tag{9.47}$$

with $A(\Box)$ defined in Eq. (9.8). In the Friedmann-Robertson-Walker (FRW) framework these are applied to the standard homogeneous isotropic metric

$$ds^2 \;=\; -dt^2 + a^2(t) \left\{ \frac{dr^2}{1 - k r^2} + r^2 \left( d\theta^2 + \sin^2\theta\, d\varphi^2 \right) \right\} \;. \tag{9.48}$$

It should be noted that there are *two* quantum contributions to the above set of effective field equations. The first one arises because of the presence of a non-vanishing cosmological constant $\lambda \simeq 1/\xi^2$ caused by the non-perturbative vacuum condensate of Eq. (8.86). As in the case of standard FRW cosmology, this is expected to be the dominant contributions at large times $t$, and gives an exponential (for $\lambda > 0$ or cyclic (for $\lambda < 0$) expansion of the scale factor.

The second contribution arises because of the running of $G$ in the effective field equations,

$$G(\Box) \;=\; G\, [1 + A(\Box)] \;=\; G \left[ 1 + a_0 \left( \xi^2 \Box \right)^{-\frac{1}{2\nu}} + \ldots \right] \;, \tag{9.49}$$

for $t \ll \xi$, with $\nu \simeq 1/3$ and $a_0 > 0$ a calculable coefficient of order one [see Eqs. (9.1) and (9.2)].

As a first step in solving the new set of effective field equations, consider first the *trace* of the field equation in Eq. (9.47), written as

$$\left( 1 - A(\Box) + O(A(\Box)^2) \right) R \;=\; 8\pi G T_\mu{}^\mu \;, \tag{9.50}$$

where $R$ is the scalar curvature. Here the operator $A(\Box)$ has been moved over on the gravitational side, so that it now acts on functions of the metric only, using the binomial expansion of $1/[1 + A(\Box)]$. To proceed further, one needs to compute the effect of $A(\Box)$ on the scalar curvature. The d'Alembertian operator acting on scalar functions $S(x)$ is given by

$$\frac{1}{\sqrt{g}}\, \partial_\mu\, g^{\mu\nu} \sqrt{g}\, \partial_\nu\, S(x) \;, \tag{9.51}$$

and for the Robertson-Walker metric, acting on functions of $t$ only, one obtains a fairly simple result in terms of the scale factor $a(t)$

$$-\frac{1}{a^3(t)} \frac{\partial}{\partial t} \left[ a^3(t) \frac{\partial}{\partial t} \right] F(t) \;. \tag{9.52}$$

As a next step one computes the action of $\Box$ on the scalar curvature $R$, which gives

$$-6\left[-2k\ddot{a}(t)-5\dot{a}^2(t)\,\ddot{a}(t)+a(t)\,\ddot{a}^2(t)+3\,a(t)\,\dot{a}(t)\,a^{(3)}(t)+a^2(t)\,a^{(4)}(t)\right]/a^3(t)\ , \tag{9.53}$$

and then $\Box^2$ on $R$ etc. Since the resulting expressions are of rapidly escalating complexity, one sets $a(t)=r_0\,t^{\alpha}$, in which case one has first for the scalar curvature itself

$$R=6\left[\frac{k}{r_0^2\,t^{2\alpha}}+\frac{\alpha\,(-1+2\,\alpha)}{t^2}\right]\ . \tag{9.54}$$

Acting with $\Box^n$ on the above expression gives for $k=0$ and arbitrary power $n$

$$c_n\,6\,\alpha\,(-1+2\,\alpha)\,t^{-2-2n}\ , \tag{9.55}$$

with the coefficient $c_n$ given by

$$c_n\;=\;4^n\frac{\Gamma(n+1)\Gamma(\frac{3\alpha-1}{2})}{\Gamma(\frac{3\alpha-1}{2}-n)}\ . \tag{9.56}$$

Here use has been made of the relationship

$$\left(\frac{d}{dz}\right)^{\alpha}(z-c)^{\beta}\;=\;\frac{\Gamma(\beta+1)}{\Gamma(\beta-\alpha+1)}\,(z-c)^{\beta-\alpha}\ , \tag{9.57}$$

to analytically continue the above expressions to negative fractional $n$ (Samko et al, 1993; Zavada, 1998). For $n=-1/2\nu$ the correction on the scalar curvature term $R$ is therefore of the form

$$\left[1-a_0\,c_\nu\,(t/\xi)^{1/\nu}\right]\cdot 6\,\alpha\,(-1+2\,\alpha)\,t^{-2} \tag{9.58}$$

with

$$c_\nu\;=\;2^{-\frac{1}{\nu}}\,\frac{\Gamma(1-\frac{1}{2\nu})\Gamma(\frac{3\alpha-1}{2})}{\Gamma(\frac{3\alpha-1}{2}+\frac{1}{2\nu})}\ . \tag{9.59}$$

Putting everything together, one then obtains for the trace part of the effective field equations

$$\left[1-a_0\,c_\nu\left(\frac{t}{\xi}\right)^{1/\nu}+O\left((t/\xi)^{2/\nu}\right)\right]\frac{6\,\alpha\,(2\,\alpha-1)}{t^2}\;=\;8\pi G\rho(t) \tag{9.60}$$

The new term can now be moved back over to the matter side in accordance with the structure of the original effective field equation of Eq. (9.47), and thus avoids the problem of having to deal with the binomial expansion of $1/[1+A(\Box)]$. One then has

$$\frac{6\,\alpha\,(2\,\alpha-1)}{t^2}\;=\;8\pi G\left[1+a_0\,c_\nu\left(\frac{t}{\xi}\right)^{1/\nu}+O\left((t/\xi)^{2/\nu}\right)\right]\rho(t)\ , \tag{9.61}$$

which is the Robertson-Walker metric form of Eq. (9.47). If one assumes for the matter density $\rho(t) \sim \rho_0 t^\beta$, then matching powers when the new term starts to take over at larger distances gives the first result

$$\beta = -2 - 1/\nu \ . \tag{9.62}$$

Thus the density decreases *faster* in time than the classical value $(\beta = -2)$ would indicate. The expansion appears therefore to be accelerating, but before reaching such a conclusion one needs to determine the time dependence of the scale factor $a(t)$ (or $\alpha$) as well.

One can alternatively pursue the following exercise in order to check the overall consistency of the approach. Consider $\Box^n$ acting on $T_\mu{}^\mu = -\rho(t)$ instead, as in the trace of the effective field equation of Eq. (9.47)

$$R \ = \ -8\pi G\left[1+A(\Box)\right] T_\mu{}^\mu \ , \tag{9.63}$$

for $\lambda = 0$ and $p(t) = 0$. For $\rho(t) = \rho_0 t^\beta$ and $a(t) = r_0 t^\alpha$ one finds in this case

$$\Box^n\left[-\rho(t)\right] \ \rightarrow \ 4^n(-1)^{n+1} \frac{\Gamma(\frac{\beta}{2}+1)\Gamma(\frac{\beta+3\alpha+1}{2})}{\Gamma(\frac{\beta}{2}+1-n)\Gamma(\frac{\beta+3\alpha+1}{2}-n)} \rho_0 t^{\beta-2n} \ , \tag{9.64}$$

which again implies for $n \rightarrow -1/2\nu$ the value $\beta = -2 - 1/\nu$ as in Eq. (9.62) for large(r) times, when the quantum correction starts to become important (since the l.h.s. of Einstein's equation always goes like $1/t^2$, no matter what the value for $\alpha$ is, at least for k=0).

The next step is to examine the full effective field equations (as opposed to just their trace part) of Eq. (9.13) with cosmological constant $\lambda = 0$,

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R \ = \ 8\pi G\left[1+A(\Box)\right] T_{\mu\nu} \ . \tag{9.65}$$

Here the d'Alembertian operator

$$\Box \ = \ g^{\mu\nu}\nabla_\mu\nabla_\nu \ , \tag{9.66}$$

acts on a second rank tensor,

$$\nabla_\nu T_{\alpha\beta} \ = \ \partial_\nu T_{\alpha\beta} - \Gamma_{\alpha\nu}^\lambda T_{\lambda\beta} - \Gamma_{\beta\nu}^\lambda T_{\alpha\lambda} \equiv I_{\nu\alpha\beta}$$

$$\nabla_\mu\left(\nabla_\nu T_{\alpha\beta}\right) = \partial_\mu I_{\nu\alpha\beta} - \Gamma_{\nu\mu}^\lambda I_{\lambda\alpha\beta} - \Gamma_{\alpha\mu}^\lambda I_{\nu\lambda\beta} - \Gamma_{\beta\mu}^\lambda I_{\nu\alpha\lambda} \ , \tag{9.67}$$

and would thus seem to require the calculation of 1920 terms, of which fortunately many vanish by symmetry. Next one assumes again that $T_{\mu\nu}$ has the perfect fluid form, for which one obtains from the action of $\Box$ on $T_{\mu\nu}$

$$\left(\Box\, T_{\mu\nu}\right)_{tt} \ = \ 6\left[\rho(t) + p(t)\right]\left(\frac{\dot a(t)}{a(t)}\right)^2 - 3\dot\rho(t)\frac{\dot a(t)}{a(t)} - \ddot\rho(t)$$

$$\begin{aligned}
\left(\Box\, T_{\mu\nu}\right)_{rr} &= \frac{1}{1-kr^2}\left\{2\left[\rho(t)+p(t)\right]\dot{a}(t)^2 - 3\,\dot{p}(t)\,a(t)\,\dot{a}(t) - \ddot{p}(t)\,a(t)^2\right\} \\
\left(\Box\, T_{\mu\nu}\right)_{\theta\theta} &= r^2\,(1-kr^2)\,\left(\Box\, T_{\mu\nu}\right)_{rr} \\
\left(\Box\, T_{\mu\nu}\right)_{\varphi\varphi} &= r^2\,(1-kr^2)\,\sin^2\theta\,\left(\Box\, T_{\mu\nu}\right)_{rr} \ ,
\end{aligned} \tag{9.68}$$

with the remaining components equal to zero. Note that a non-vanishing pressure contribution is generated in the effective field equations, even if one assumes initially a pressureless fluid, $p(t) = 0$. As before, repeated applications of the d'Alembertian $\Box$ to the above expressions leads to rapidly escalating complexity, which can only be tamed by introducing some further simplifying assumptions. In the following we will therefore assume that $T_{\mu\nu}$ has the perfect fluid form appropriate for non-relativistic matter, with a power law behavior for the density, $\rho(t) = \rho_0\, t^{\beta}$, and $p(t) = 0$. Thus all components of $T_{\mu\nu}$ vanish in the fluid's rest frame, except the $tt$ one, which is simply $\rho(t)$. Setting $k = 0$ and $a(t) = r_0\, t^{\alpha}$ one then finds

$$\begin{aligned}
\left(\Box\, T_{\mu\nu}\right)_{tt} &= \left(6\,\alpha^2 - \beta^2 - 3\,\alpha\,\beta + \beta\right)\rho_0\, t^{\beta-2} \\
\left(\Box\, T_{\mu\nu}\right)_{rr} &= 2r_0^2\, t^{2\alpha}\,\alpha^2\,\rho_0\, t^{\beta-2} \ ,
\end{aligned} \tag{9.69}$$

which again shows that the $tt$ and $rr$ components get mixed by the action of the $\Box$ operator, and that a non-vanishing $rr$ component gets generated, even though it was not originally present.

Higher powers of the d'Alembertian $\Box$ acting on $T_{\mu\nu}$ can then be computed as well. But a comparison with the left hand (gravitational) side of the effective field equation, which always behaves like $\sim 1/t^2$ for $k = 0$, shows that in fact a solution can only be achieved at order $\Box^n$ provided the exponent $\beta$ satisfies $\beta = -2 + 2n$, or

$$\beta \;=\; -2 - 1/\nu \ , \tag{9.70}$$

as was found previously from the trace equation, Eqs. (9.47) and (9.62). As a result one obtains a much simpler set of expressions, which for general $n$ read

$$\left(\Box^n\, T_{\mu\nu}\right)_{tt} \;\to\; c_{tt}(\alpha,\nu)\,\rho_0\, t^{-2} \ , \tag{9.71}$$

for the $tt$ component, and similarly for the $rr$ component

$$\left(\Box^n\, T_{\mu\nu}\right)_{rr} \;\to\; c_{rr}(\alpha,\nu)\, r_0^2\, t^{2\alpha}\,\rho_0\, t^{-2} \ . \tag{9.72}$$

But remarkably one finds for the two coefficients the simple identity

$$c_{rr}(\alpha,\nu) \;=\; \tfrac{1}{3}\, c_{tt}(\alpha,\nu) \ , \tag{9.73}$$

as well as $c_{\theta\theta} = r^2\, c_{rr}$ and $c_{\varphi\varphi} = r^2\, \sin^2\theta\, c_{rr}$. The identity $c_{rr} = \frac{1}{3}\, c_{tt}$ implies, from the consistency of the $tt$ and $rr$ effective field equations at large times,

$$\alpha \;=\; \frac{1}{2} \ . \tag{9.74}$$

One can find a closed form expression for the coefficients $c_{tt}$ and $c_{rr} = c_{tt}/3$ as functions of $\nu$ which are not particularly illuminating, except for providing an explicit proof that they exist.

As a result, in the simplest case, namely for a universe filled with non-relativistic matter ($p$=0), the effective Friedmann equations then have the following appearance

$$
\begin{aligned}
\frac{k}{a^2(t)} + \frac{\dot{a}^2(t)}{a^2(t)} &= \frac{8\pi G(t)}{3}\rho(t) + \frac{1}{3\xi^2} \\
&= \frac{8\pi G}{3}\left[1 + c_\xi \,(t/\xi)^{1/\nu} + \dots\right]\rho(t) + \tfrac{1}{3}\lambda \ , \qquad (9.75)
\end{aligned}
$$

for the $tt$ field equation, and

$$
\frac{k}{a^2(t)} + \frac{\dot{a}^2(t)}{a^2(t)} + \frac{2\ddot{a}(t)}{a(t)} = -\frac{8\pi G}{3}\left[c_\xi \,(t/\xi)^{1/\nu} + \dots\right]\rho(t) + \lambda \ , \qquad (9.76)
$$

for the $rr$ field equation. The running of $G$ appropriate for the Robertson-Walker metric, and appearing explicitly in the first equation, is given by

$$
G(t) = G\left[1 + c_\xi\left(\frac{t}{\xi}\right)^{1/\nu} + \dots\right] \ , \qquad (9.77)
$$

with $c_\xi$ of the same order as $a_0$ of Eq. (9.1). Note that the running of $G(t)$ induces as well an effective pressure term in the second ($rr$) equation.[1] One has therefore an effective density given by

$$
\rho_{eff}(t) = \frac{G(t)}{G}\rho(t) \ , \qquad (9.78)
$$

and an effective pressure

$$
p_{eff}(t) = \frac{1}{3}\left(\frac{G(t)}{G} - 1\right)\rho(t) \ , \qquad (9.79)
$$

with $p_{eff}(t)/\rho_{eff}(t) = \frac{1}{3}(G(t) - G)/G(t)$. Strictly speaking, the above results can only be proven if one assumes that the pressure's time dependence is given by a power law. In the more general case, the solution of the above equations for various choices of $\xi$ and $a_0$ has to be done numerically. Within the FRW framework, the gravitational vacuum polarization term behaves therefore in some ways (but not all) like a positive pressure term, with $p(t) = \omega\rho(t)$ and $\omega = 1/3$, which is therefore characteristic of radiation. One could therefore visualize the gravitational vacuum polarization contribution as behaving like ordinary radiation, in the form of a dilute virtual graviton gas: a radiative fluid with an equation of state $p = \frac{1}{3}\rho$. But this

[1] We wish to emphasize that we are *not* talking here about models with a time-dependent value of $G$. Thus, for example, the value of $G \simeq G_c$ at laboratory scales should be taken to be constant throughout most of the evolution of the universe.

would overlook the fact that the relationship between density $\rho(t)$ and scale factor $a(t)$ is quite different from the classical case.

The running of $G(t)$ in the above equations follows directly from the basic result of Eq. (9.1), following the more or less unambiguously defined sequence $G(k^2) \to G(\Box) \to G(t)$. For large times $t \gg \xi$ the form of Eq. (9.1), and therefore Eq. (9.77), is no longer appropriate, due to the spurious infrared divergence of Eq. (9.1) at small $k^2$. Indeed from Eq. (9.2), the infrared regulated version of the above expression should read instead

$$G(t) \simeq G \left[ 1 + c_\xi \left( \frac{t^2}{t^2 + \xi^2} \right)^{\frac{1}{2\nu}} + \ldots \right] . \tag{9.80}$$

For very large times $t \gg \xi$ the gravitational coupling then approaches a constant, finite value $G_\infty = (1 + a_0 + \ldots) G_c$. The modification of Eq. (9.80) should apply whenever one considers times for which $t \ll \xi$ is not valid. But since $\xi \sim 1/\sqrt{\lambda}$ is of the order the size of the visible universe, the latter regime is largely of academic interest.

It should also be noted that the effective Friedmann equations of Eqs. (9.75) and (9.76) also bear a superficial degree of resemblance to what might be obtained in some scalar-tensor theories of gravity, where the gravitational Lagrangian is postulated to be some singular function of the scalar curvature (Capozziello et al, 2003; Carroll et al, 2004; Flanagan, 2004). Indeed in the Friedmann-Robertson-Walker case one has, for the scalar curvature in terms of the scale factor,

$$R = 6 \left( k + \dot{a}^2(t) + a(t) \ddot{a}(t) \right) / a^2(t) \ , \tag{9.81}$$

and for $k = 0$ and $a(t) \sim t^\alpha$ one has

$$R = \frac{6 \alpha (2\alpha - 1)}{t^2} \ , \tag{9.82}$$

which suggests that the quantum correction in Eq. (9.75) is, at this level, nearly indistinguishable from an inverse curvature term of the type $(\xi^2 R)^{-1/2\nu}$, or $1/(1 + \xi^2 R)^{1/2\nu}$ if one uses the infrared regulated version. The former would then correspond the to an effective gravitational action

$$I_{eff} \simeq \frac{1}{16\pi G} \int dx \sqrt{g} \left( R + \frac{f \, \xi^{-\frac{1}{\nu}}}{|R|^{\frac{1}{2\nu} - 1}} - 2\lambda \right) \ , \tag{9.83}$$

with $f$ a numerical constant of order one, and $\lambda \simeq 1/\xi^2$. But this superficial resemblance is seen here more as an artifact, due to the particularly simple form of the Robertson-Walker metric, with the coincidence of several curvature invariants not expected to be true in general. In particular in Eqs. (9.75) and (9.76) it would seem artificial and in fact inconsistent to take $\lambda \sim 1/\xi^2$ to zero while keeping the $\xi$ in $G(t)$ finite.

## 9.6 Quantum Gravity and Mach's Principle

The essence of Mach's proposal can be summarized in the statement that faraway galaxies provide a preferential reference frame with respect to which the motion of an observer can be characterized as inertial or not (Mach, 1960). In such a framework acceleration cannot be described in absolute terms, since in the absence of such a distant mass distribution the very notion of acceleration would be meaningless. Water in the Newtonian water bucket rotating by itself in an empty universe could not possibly rise at the sides, since there would be no frame of reference for such an isolated bucket. It would seem possible therefore that the origin of inertial mass itself might be due to some new long range interaction between a test mass and the faraway galaxy distribution (Sciama, 1953; Feynman, 1963).

Over the years, and starting with considerations by Einstein himself, extensive discussion's of Mach's principle and its relation to General Relativity have appeared. Some have argued convincingly that some aspects of Mach's principle are indeed incorporated in the framework of General Relativity, since for example the choice of locally inertial frames is determined form the overall mass distribution, and in particular the Einstein tensor is completely determined by the energy-momentum tensor for matter. On the other hand the Weyl conformal tensor $W_{\mu\nu\lambda\sigma}$ is not determined by matter, and the field equations remain incomplete without the specifications of suitable boundary conditions. Another disturbing aspect is the fact that in a universe without matter, $T^{\mu\nu} = 0$, there still exist flat space solutions $g_{\mu\nu} = \eta_{\mu\nu}$ describing the motion of a test particle in terms of Minkowski dynamics, whereas based on Mach's principle one would expect in this case a complete absence of inertia (as discussed in some detail already in (Pauli, 1958), where he provides some arguments in favor of the existence of a small cosmological constant $\lambda$).

A number of proposals have been put forward to address the issue of how to incorporate some aspects of Mach's proposal in General Relativity. One possibility has been the introduction of boundary terms to account for non-trivial boundary conditions at infinity. Another set of proposals (Sciama, 1953; 1971; Brans and Dicke, 1961; Feynman, 1963) postulate that the value of Newton's constant $G$ is in some way related to a new cosmic field describing inertial "forces", i.e. the natural tendency of massive bodies to resist acceleration. In Sciama's original theory, which is non-relativistic and therefore in many ways incomplete, the key ingredient is the fact that an acceleration of a test particle with respect to the universe should be equivalently described by an acceleration of the universe as a whole, with the test particle at rest.

In the Brans-Dicke relativistic scalar-tensor theory of gravity, $G$ is assumed not to be a constant of nature: instead it is described by a local average of the additional Brans-Dicke scalar field $\phi$, $G^{-1} \simeq \langle \phi \rangle$. Of course even in this extended theory of gravity on needs at some point to specify a consistent set of boundary conditions at infinity. But unfortunately the theory, at least in its original formulation is not favored by current solar system tests, which put stringent bounds on the value of the Brans-Dicke parameter $\omega$ describing the coupling of the cosmic scalar field $\phi$ to gravity.

Nevertheless these arguments are sufficiently concrete to allow specific predictions, the principal one being a relationship between Newton's constant $G$ and the nature of mass distribution in the universe. Feynman in particular argues that the fact that the total energy of the visible universe, of approximate mass $M$ and approximate size $R$, is zero,

$$M - \frac{3}{5} \frac{GM^2}{R} \simeq 0 \ , \tag{9.84}$$

might not be a simple numerical coincidence. After all, in order to cancel completely, the first and second term need to be fine tuned to one part in $10^{80}$ (the approximate numbers of protons in the visible universe). In fact a result of this type should not be seen as entirely surprising, since one knows for example that in the canonical formulation of gravity the Hamiltonian constraint is precisely $H = 0$. If that is indeed the case, then one would expect $G$ to be related to cosmic quantities,

$$G \simeq \frac{R}{M} \ . \tag{9.85}$$

Another separate and puzzling result, coming from cosmology, is that the observed matter density is very close to one, in the appropriate units, $\Omega \approx 1$ (more properly, it is actually a combined density of baryonic, non-baryonic and dark energy components, $\Omega_t = \Omega_\lambda + \Omega_{dm} + \Omega_b$, but we will overlook such subtle distinctions here). For non-relativistic matter one has $\Omega = \rho/\rho_{\text{crit}}$ with $\rho_{\text{crit}} = 3H^2/8\pi G$, which then gives

$$\Omega = \frac{8\pi G \rho}{3 H_0^2} \ , \tag{9.86}$$

where $H_0$ is the value of the Hubble constant today. Since $H_0$ is expected in general to be time-dependent, a possible scenario would be one in which for large times $R^{-1} = H_\infty = \lim_{t\to\infty} H(t)$, with $H_\infty^2 = \frac{\lambda}{3}$, and for which the horizon radius is $R_\infty = H_\infty^{-1}$. Then for $H_0 \sim R$ one has $\Omega \sim 1$ if $2MGR \simeq 1$, which would again relate $G$ to the overall size and mass of the visible universe in accordance with Eq. (9.85).

Perhaps a slightly more convincing argument can be formulated by making use of the Lense-Thirring effect (Lense and Thirring, 1918). There one considers a test particle at rest at the origin, and a thin spherical shell of total mass $M$ rotating rigidly around it, at some distance $R$ and at angular velocity $\Omega$. Originally the problem was solved assuming the fields to be weak and the accelerations to be small, but a more complete analysis (Brill and Cohen, 1966) shows that the test mass at the origin (a Foucault pendulum, for example) will rotate due to frame dragging with an angular velocity given by

$$\omega = \Omega \left[ 1 + \frac{3}{4} \frac{R - 2MG}{MG(1+\beta)} \right]^{-1} \ , \tag{9.87}$$

where $\beta$ is a dimensionless constant that depends on the relative contributions of $T^{ij}$ and $T^{00}$ to the shell's gravitational mass. For small $MG$ and $\beta$ one recovers the post-Newtonian result $\omega/\Omega = 4MG/3R$. The full solution on the other hand gives complete inertial frame dragging $\omega = \Omega$ if one has exactly $G = R/2M$, as Mach would have expected. It seems a general feature of these scenarios that, since Newton's constant is related to the total size and overall mass distribution in the visible universe, it is in fact not expected to be a constant, although the specific way in which $R$ and $M$ would vary over time is not understood.

Turning to quantum gravity, the question arises if any of the above ideas and speculations have any relevance for a quantum version of General Relativity. One can approach the problem from a number of different points of view.

For example, if gravity is indeed embedded in a larger unified theory of all forces in nature (such as supergravity or string theory), it is often stated that such a unifications would become manifest at distances of the order of $l_P = \sqrt{G}$. These claims build on an implicit underlying assumption that the Planck energy $\mu_P = 1/l_P$ is a fundamental energy scale at which all fundamental forces attain the same strength, and where unification therefore has to take place.[2] But if $G$ is related to cosmic quantities such as $R$ and $M$, then the argument for why the unification scale should be at $\mu_P$, and not at some other scale, becomes less compelling. Nevertheless the statement that quantum effects in gravity will presumably become very important at distances of order $l_P$ remains valid, since the latter is one of the fundamental scales appearing in the gravitational Lagrangian.

Recent astrophysical observations suggest our universe is described to a good approximation by General Relativity with a non-vanishing cosmological constant $\lambda$. Within the framework of General Relativity $\lambda$ can be viewed as describing the curvature of empty space, and therefore not associated in any way with the matter distribution. Of course one of the great puzzles of theoretical physics is why the observed cosmological constant is so small in "natural" gravitational units, $\lambda G \sim 10^{-122}$. If indeed Mach's ideas are correct, and gravitational constants such as $G$ are related to cosmic quantities, then one would expect $\lambda$ to be related to the quantities $R$ and $M$. Since $\lambda$ describes the curvature of the vacuum, it cannot involve $M$, which then leaves as the only possibility, by dimensional arguments,

$$\lambda \simeq \frac{1}{R^2} \ . \tag{9.88}$$

The smallness of the observed cosmological constant is then simply a consequence of the product $\lambda G \sim 1/MR$ being a ratio of two vastly different scales, the Compton wavelength of the universe $(1/M)$, and its visible size $R$. From this point of view, one is lead to the paradoxical conclusion that it is actually $\lambda$ (or $R$), and not $G$, that appears in the end as the genuinely gravitational scale, to the extent that only the former makes sense even in purely gravitational, matter-less universe. The hierarchy problem of particle physics would then be shifted to the new question of why $1/R$ ends up being so small (by 41 orders of magnitude) compared to other typical particle physics scales, such as the proton mass.

---

[2] Even though even to this day the precise relationship between the string coupling constant $g$, the string scale $\alpha'$ and the Planck scale $\mu_P$ remains somewhat unclear.

# References


Abbott, L., 1982, Introduction to the Background Field Method, in the proceedings of the 21st Cracow School of Theoretical Physics, Cracow, Poland, 1981, *Acta Phys. Polon.* B **13**, 33.

Abers, E. S., 2004, *Quantum Mechanics* (Prentice Hall, New Jersey, 2004).

Abers, E. S., and B. W. Lee, 1973, Phys. Rept. **9**, 1.

Abrikosov, A. A., L. P. Gorkov and I. E. Dzyaloshinski , 1963, *Methods of Quantum Field Theory in Statistical Physics* (Prentice Hall, New York, 1963).

Aida, T., and Y. Kitazawa, 1997, *Nucl. Phys.* B **491**, 427.

Albeverio, S. A., and R. J. Hoegh-Krohn, 1976, *Mathematical Theory of Feynmann Path Integrals*, *Lecture Notes in Mathematics*, Vol. 523, (Springer-Verlag, Berlin, 1976).

Ambjørn, J., J. Jurkiewicz and R. Loll, 2005, *Phys. Rev.* D **72**, 064014.

Aoki, S., et al., [CP-PACS Collaboration], 2003, *Phys. Rev.* D **67**, 034503.

Arnowitt, R., S. Deser and C. W. Misner, 1960, *Phys. Rev.* **117**, 1595.

Arnowitt, R., S. Deser and C. W. Misner, 1962, "The Dynamics Of General Relativity", in *Gravitation: An Introduction to Current Research*, Louis Witten (ed.) (Wiley, New York, 1962), Chapter 7, 227–265 (gr-qc/0405109).

Ashtekar, A., 1986a, *Phys. Rev. Lett.* **57**, 2244.

Ashtekar, A., 1986b, *Phys. Rev.* D **36**, 1587.

Ashtekar, A., and J. Lewandowski, 2004, *Class. Quant. Grav.* **21**, R53.

Avramidy, I. G., and A. O. Barvinsky, 1985, *Phys. Lett.* **159** B, 269.

Balian, R., J. M. Drouffe and C. Itzykson, 1975, *Phys. Rev.* D **11**, 2104.

Banados, M., C. Teitelboim and J. Zanelli, 1992, *Phys. Rev. Lett.* **69**, 1849.

Barrett, J. W., M. Galassi, W. A. Miller, R. D. Sorkin, P. A. Tuckey and R. M. Williams, 1997, *Int. J. Theor. Phys.* **36**, 815.

Barth, N. H., and S. M. Christensen, 1983, *Phys. Rev.* D **28**, 1876.

Barvinsky, A. O., 1998, *Nucl. Phys.* B **520**, 533.

Barvinsky, A. O., 2003, *Phys. Lett.* B **572**, 109.

Barvinsky, A. O., 2007, *Phys. Rev. Lett.* **99**, 071301.

Barvinsky, A. O., and G. A. Vilkovisky, 1985, *Phys. Rept.* **119**, 1.

Barvinsky, A. O., and C. Kiefer, 1998, *Nucl. Phys.* B **526**, 509.

Becker, K., M. Becker and J. H. Schwarz, 2007, *String Theory and M-Theory: A Modern Introduction* (Cambridge University Press, New York, 2007).

Beirl, W., E. Gerstenmayer, H. Markum and J. Riedler, 1994, *Phys. Rev.* D **49**, 5231.

Bekenstein, J. D., 1973, *Phys. Rev.* D **7**, 2333.

Bekenstein, J. D., 1974, *Phys. Rev.* D **9**, 3292.

Berg, B., 1985, *Phys. Rev. Lett.* **55**, 904.

Berg, B., 1986, *Phys. Lett.* B **176**, 39.

Bern, Z., J. J. Carrasco, L. J. Dixon, H. Johansson, D. A. Kosower and R. Roiban, 2007, *Phys. Rev. Lett.* **98**, 161303.
Bernard, C. W., 1974, *Phys. Rev.* D **9**, 3312.
Bernreuther, W., and F. J. Wegner, 1986, *Phys. Rev. Lett.* **57**, 1383.
Bittner, E., W. Janke and H. Markum, 2002, *Phys. Rev.* D **66**, 024008.
Blyth, W. F., and C. J. Isham, 1975, *Phys. Rev.* D **11**, 768.
Boulware, D. G., and S. Deser, 1975, *Annals Phys.* **89**, 193.
Brans, C., and R. H. Dicke, 1961, *Phys. Rev.* **124**, 925.
Brewin, L. C., and A. P. Gentle, 2001, *Class. Quant. Grav.* **18**, 517.
Brezin, E., and C. Itzykson, 1970, *Phys. Rev.* D **2**, 1191.
Brezin, E., and J. Zinn-Justin, 1976, *Phys. Rev. Lett.* **36**, 691.
Brezin, E., J. Zinn-Justin and J. C. Le Guillou, 1976, *Phys. Rev.* D **14**, 2615.
Brezin, E., J. C. Le Guillou and J. Zinn-Justin, 1976, in *Phase Transitions and Critical Phenomena*, Vol. 6, C. Domb and M. S. Green (eds.) (Academic Press, New York, 1976).
Brezin, E., and J. Zinn-Justin, 1985, *Nucl. Phys.* B **257**, 867.
Brink, L., and J. H. Schwarz, 1977, Nucl. Phys. B **121**, 285.
Brill, D. R., and J. M. Cohen, 1966, Phys. Rev. 143, 1011.
Callan, C. G., 1970, *Phys. Rev.* D **2**, 1541.
Callan, C. G., D. Friedan, E. J. Martinec and M. J. Perry, 1985, Nucl. Phys. B **262**, 593.
Campostrini, M., M. Hasenbusch, A. Pelissetto, P. Rossi and E. Vicari, 2001, *Phys. Rev.* B **63**, 214503.
Capozziello, S., S. Carloni and A. Troisi, 2003, in Recent Research Developments in Astronomy and Astrophysics, RSP/AA/21-03, astro-ph/0303041.
Capper, D. M., G. Leibbrandt and M. Ramon Medrano, 1973, *Phys. Rev.* D **8**, 4320.
Caracciolo, S., and A. Pelissetto, 1988, *Nucl. Phys.* B **299**, 693.
Cardy, J. L. (ed.), 1988, *Finite Size Scaling, Current Physics – Sources and Comments* (North Holland, Amsterdam, 1988).
Cardy, J. L., 1997, *Scaling and the Renormalization Group in Statistical Physics* (Cambridge University press, England, 1997).
Cardy, J. L., and H. W. Hamber, 1980, *Phys. Rev. Lett.* **45**, 499.
Carroll, S. M., V. Duvvuri, M. Trodden and M. S. Turner, 2004, *Phys. Rev.* D **70**, 043528.
Caselle, M., A. D'Adda and L. Magnea, 1989, *Phys. Lett.* B **232**, 457.
Casher, A., F. Englert, N. Itzhaki, S. Massar and R. Parentani, 1997, *Nucl. Phys.* B **484**, 419.
Cheeger, J., W. Müller and R. Schrader, 1982, in *Unified Theories of Elementary Particles*, Heisenberg Symposium, München, 1981, *Springer Lecture Notes in Physics*, P. Breitlohner and H. P. Dürr (eds.) (Springer, New York, 1982).
Cheeger, J., W. Müller and R. Schrader, 1984, *Comm. Math. Phys.* **92**, 405.
Christ, N. H., R. Friedberg and T. D. Lee, 1982, *Nucl. Phys.* B **202**, 89.
Christensen, S. M., and M. J. Duff, 1978, *Phys. Lett.* B **79**, 213.
Christensen, S. M., M. J. Duff, G. W. Gibbons and M. Roček, 1980, *Phys. Rev. Lett.* **45**, 161.
Collins, P. A., and R. M. Williams, 1973, *Phys. Rev.* D **7**, 965.
Collins, P. A., and R. M. Williams, 1974, *Phys. Rev.* D **10**, 353.
Coxeter, H., 1948, *Regular Polytopes* (Methuen and Co. Ltd., London, 1948).
Coxeter, H., 1974, *Regular Complex Polytopes* (Cambridge University press, England, 1974).
Cremmer, E., and J. Scherk, 1977, *Nucl. Phys.* B **127**, 259.
Cremmer, E., and B. Julia, 1978, *Phys. Lett.* B **80**, 48.
Creutz, M., 1977, *Phys. Rev.* D **15**, 1128.
Curci, G., and G. Veneziano, 1987, *Nucl. Phys.* B **292**, 555.
Damour, T., 2006, *Experimental test of Gravitational Theory*, in Review of Particle Physics, W.-M. Yao et al., J. Phys. G 33, 1 (2006).
Damour, T., and A. M. Polyakov, 1994, *Nucl. Phys.* B **423**, 532.
Damour, T., F. Piazza and G. Veneziano, 2002, *Phys. Rev.* D **66**, 046007.
Damour, T., 2006, Experimental Tests of Gravitational Theory, in the 2006 Review of Particle Properties, W. M. Yao, et al., *J. Phys.* G **33**, 1.

Das, A., 1977, *Phys. Rev.* D **15**, 2805.
Das, A., M. Kaku and P. K. Townsend, 1979, *Phys. Lett.* **81** B, 11.
David, F., 1982, *Nucl. Phys.* B **209**, 433.
David, F., 1985, *Nucl. Phys.* B **257**, 45.
Deser, S., and B. Zumino, 1976, *Phys. Lett.* B **62**, 335.
Deser, S., R. Jackiw and S. Templeton, 1982, *Annals Phys.* **140**, 372.
Deser, S., R. Jackiw and G. 't Hooft, 1984, *Annals Phys.* **152**, 220.
Deser, S., and R. Jackiw, 1984, *Annals Phys.* **153**, 405.
DeWitt, B. S., 1962, *J. Math. Phys.* **3**, 1073.
DeWitt, B. S., 1964, "Dynamical Theory of Groups and Fields", in *Relativity, Groups and Topology*, 1963, *Les Houches Lectures*, B. de Witt and C. de Witt (eds.) (Gordon and Breach, New York 1964).
DeWitt, B. S., 1967a, *Phys. Rev.* **160**, 1113.
DeWitt, B. S., 1967b, *Phys. Rev.* **162**, 1195.
DeWitt, B. S., 1967c, *Phys. Rev.* **162**, 1239.
Dilkes, F. A., J. Duff, J. T. Liu and H. Sati, 2001, *Phys. Rev. Lett.* **87**, 041301.
Dirac, P. A. M., 1958, *Proc. Roy. Soc. Lond.* A **246**, 333.
Dirac, P. A. M., 1959, *Phys. Rev.* **114**, 924.
Dolan, L., and R. Jackiw, 1974, *Phys. Rev.* D **9**, 3320.
Domb, C., and H. S. Green, 1976, *Phase Transitions and Critical Phenomena*, Vol. 3 (Academic Press, London, 1976).
Dowker, J. S., and R. Critchley, 1976, *Phys. Rev.* D **13**, 3224.
Drouffe, J. M., and C. Itzykson, 1978, *Phys. Rept.* **38**, 133.
Drouffe, J. M., G. Parisi and N. Sourlas, 1979, *Nucl. Phys.* B **161**, 397.
Drummond, I. T., 1986, *Nucl. Phys.* B **273**, 125.
Dubal, M. R., 1989a, *Class. Quant. Grav.* **6**, 141.
Dubal, M. R., 1989b, *Class. Quant. Grav.* **6**, 1925.
Faddeev, L. D., and V. N. Popov, 1967, Phys. Lett. B **25**, 29.
Faddeev, L., and V. Popov, 1974, *Sov. Phys. Usp.* **16**, 777 *Usp. Fiz. Nauk.* **109**, 427 (1974).
Fayet, P., and S. Ferrara, 1977, *Phys. Rept.* **32**, 249.
Feinberg, G., R. Friedberg, T. D. Lee, and H. C. Ren, 1984, Nucl. Phys. B **245**, 343.
Feo, A., 2003, Nucl. Phys. Proc. Suppl. **119**, 198.
Ferrara, S., 1984, *Phys. Rept.* **105**, 5.
Fetter, L., and J. D. Walecka, 1971, *Quantum Theory of Many Particle Systems* (McGraw-Hill, Boston, 1971).
Feynman, R. P., 1948, *Rev. Mod. Phys.* **20**, 367–387.
Feynman, R. P., 1950, *Phys. Rev.* **80**, 440.
Feynman, R. P., 1962, *Lectures on Gravitation*, 1962–1963, F. B. Morinigo and W. G. Wagner (eds.) (California Institute of Technology, Pasadena, 1971).
Feynman, R. P., 1963, *Act. Phys. Pol.* **24**, 697.
Feynman, R. P., and A. Hibbs, 1965, *Quantum Mechanics and Path integrals* (McGraw-Hill, New York, 1965).
Fisher, M. E., 1967, *Physics* **3**, 255.
Flanagan, E. E., 2004, *Phys. Rev. Lett.* **92**, 071101.
Foerster, D., H. B. Nielsen and M. Ninomiya, 1980, *Phys. Lett.* B **94**, 135.
Font, A., and S. Theisen, 2005, *Lect. Notes Phys.* **668**, 101.
Forger, K., B. A. Ovrut, S. J. Theisen and D. Waldram, 1996, *Phys. Lett.* B **388**, 512.
Fradkin, E. S., and G. A. Vilkovisky, 1973, *Phys. Rev.* D **8**, 4241.
Fradkin, E. S., and A. A. Tseytlin, 1981, *Phys. Lett.* **104** B, 377.
Fradkin, E. S., and A. A. Tseytlin, 1985, *Phys. Lett.* B **160**, 69.
Frampton, P., 2000, *Gauge Field Theories* (Wiley, New York, 2000).
Fredenhagen, K., and R. Haag, 1990, *Commun. Math. Phys.* **127**, 273.
Freedman, D. Z., P. van Nieuwenhuizen and S. Ferrara, 1976, *Phys. Rev.* D **13**, 3214.

Fröhlich, J., 1981, I. H. E. S. preprint 1981, in *Non-Perturbative Quantum Field Theory: Mathematical Aspects and Applications*, Selected Papers (World Scientic, Singapore, 1992), 523–545.
Gaillard, M. K., and B. Zumino, 2008, Supersymmetry and Superstring Phenomenology, hep-th:0805.3726.
Galassi, M., 1993, *Phys. Rev.* D **47**, 3254–3264.
Gallavotti, G., 1985, *Rev. Mod. Phys.* **57**, 471.
Gastmans, R., R. Kallosh and C. Trufn, 1978, *Nucl. Phys.* B **133**, 417.
Gawedzki, K., and A. Kupiainen, 1985b, *Phys. Rev. Lett.* **55**, 363.
Gentle, A. P., and W. A. Miller, 1998, *Class. Quant. Grav.* **15**, 389.
Gentle, A. P., 2002, *Gen. Rel. Grav.* **34**, 1701.
Gibbons, G. W., and S. W. Hawking, 1977, *Phys. Rev.* D **15**, 2752.
Gibbons, G. W., S. W. Hawking and M. J. Perry, 1978, *Nucl. Phys.* B **138**, 141.
Gibbons, G. W., and M. J. Perry, 1978, *Nucl. Phys.* B **146**, 90.
Giddings, S., J. Abbott and K. Kuchar, 1984, *Gen. Rel. Grav.* **16**, 751.
Giddings, S. B., D. Marolf and J. B. Hartle, 2006, *Phys. Rev.* D **74**, 064018.
Gilkey, P., 1975, *J. Diff. Geom.* **10**, 601–618.
Giulini, D., 1995, *Phys. Rev.* D **51**, 5630.
Glimm, J., and A. Jaffe, 1974, *Phys. Rev. Lett.* **33**, 440–442.
Glimm, J., and A. Jaffe, 1981, *Quantum Mechanics – A Functional Integral Point of View* (Springer-Verlag, Berlin, 1981).
Goldberg, J. N., 1958, *Phys. Rev.* **111**, 315.
Goroff, M. H., and A. Sagnotti, 1985, *Phys. Lett.* B **160**, 81; *Nucl. Phys.* B **266**, 709.
Goto, T., 1971, *Prog. Theor. Phys.* **46**, 1560.
Grifths, R. B., 1969, *Phys. Rev. Lett.* **23**, 17.
Grisaru, M. T., 1977, *Phys. Lett.* B **66**, 75.
Grisaru, M. T., P. van Nieuwenhuizen and J. A. M. Vermaseren, 1976, *Phys. Rev. Lett.* **37**, 1662.
Gross, D. J., and A. Neveu, 1974, *Phys. Rev.* D **10**, 3235.
Gross, D. J., 1976, in *Les Houches 1975, Proceedings, Methods In Field Theory*, Amsterdam 1976, 141–250.
Gross, D. J., 1984, *Phys. Lett.* B **138**, 185.
Gross, D. J., and V. Periwal, 1988, *Phys. Rev. Lett.* **60**, 2105.
Gross, D. J., J. A. Harvey, E. J. Martinec and R. Rohm, 1985, Phys. Rev. Lett. **54**, 502.
Guida, R., and J. Zinn-Justin, 1998, *J. Phys.* A **31**, 8103.
Gullstrand, A., 1922, *Arkiv. Mat. Astron. Fys.* **16**(8), 1–15.
Halliwell, J. J., and J. B. Hartle, 1990, *Phys. Rev.* D **41**, 1815.
Halliwell, J. J., and J. Louko, 1989a, *Phys. Rev.* D **39**, 2206.
Halliwell, J. J., and J. Louko, 1989b, *Phys. Rev.* D **40**, 1868.
Halliwell, J. J., and J. Louko, 1989c, *Phys. Rev.* D **42**, 3997.
Hamber, H. W., 1984, in *Critical Phenomena, Random Systems and Gauge Theories*, 1984 Les Houches Summer School, Session XLIII (North Holland, Amsterdam, 1984).
Hamber, H. W., 1994, *Phys. Rev.* D **50**, 3932.
Hamber, H. W., 2000, *Phys. Rev.* D **61**, 124008.
Hamber, H. W., and S. Liu, 1995, *Phys. Lett.* B **357**, 51.
Hamber, H. W., and G. Parisi, 1983, *Phys. Rev.* D **27**, 208.
Hamber, H. W., and R. M. Williams, 1984, *Nucl. Phys.* B **248**, 392.
Hamber, H. W., and R. M. Williams, 1985a, *Phys. Lett.* B **157**, 368.
Hamber, H. W., and R. M. Williams, 1986, *Nucl. Phys.* B **269**, 712.
Hamber, H. W., and R. M. Williams, 1995, *Nucl. Phys.* B **435**, 361.
Hamber, H. W., and R. M. Williams, 1999, *Phys. Rev.* D **59**, 064014.
Hamber, H. W., and R. M. Williams, 2005a, *Phys. Rev.* D **72**, 044026.
Hamber, H. W., and R. M. Williams, 2005b, *Phys. Rev.* D **75**, 084014.
Hamber, H. W., and R. M. Williams, 2006, *Phys. Rev.* D **73**, 044031.
Hamber, H. W., and R. M. Williams, 2007, *Phys. Rev.* D **76**, 084008.
Hartle, J. B., 1984, *Phys. Rev.* D **29**, 2730.

Hartle, J. B., 1985, *J. Math. Phys.* **26**, 804.
Hartle, J. B., 1986, *J. Math. Phys.* **27**, 287.
Hartle, J. B., 2003, *Gravity: An Introduction to Einstein's General Relativity*, (Pearson Education, 2003).
Hartle, J. B., and S. W. Hawking, 1976, *Phys. Rev.* D **13**, 2188.
Hartle, J. B., and S. W. Hawking, 1983, *Phys. Rev.* D **28**, 2960.
Hartle, J. B., W. A. Miller and R. M. Williams, 1997, *Class. and Quant. Gravity* **14**, 2137.
Hartle, J. B., S. W. Hawking and T. Hertog, 2008, arXiv:0803.1663 [hep-th].
Hasslacher, B., and M. J. Perry, 1981, *Phys. Lett.* B **103**, 21.
Hawking, S. W., 1975, Commun. Math. Phys. **43**, 199; [Erratum-ibid. **46**, 206 (1976)].
Hawking, S. W., 1976, *Phys. Rev.* D **13**, 191.
Hawking, S. W., 1977, *Commun. Math. Phys.* **55**, 133.
Hawking, S. W., 1978a, *Phys. Rev.* D **18**, 1747.
Hawking, S. W., 1978b, *Nucl. Phys.* B **144**, 349.
Hawking, S. W., 1979, in *General Relativity – An Einstein Centenary Survey*, S. W. Hawking and W. Israel (eds.) (Cambridge University Press, England, 1979).
Hawking, S. W., and J. C. Luttrell, 1984a, *Nucl. Phys.* B **247**, 250.
Hawking, S. W., and J. C. Luttrell, 1984b, *Phys. Lett.* B **143**, 83.
Hawking, S. W., and Z. C. Wu, 1985, *Phys. Lett.* B **151**, 15.
Hawking, S. W., and D. N. Page, 1986a, *Nucl. Phys.* B **264**, 185.
Hawking, S. W., and D. N. Page, 1986b, *Nucl. Phys.* B **298**, 789.
Hawking, S. W., and T. Hertog, 2002, *Phys. Rev.* D **65**, 103515.
Heisenberg, W., 1938, *Z. Phys.* **110**, 251.
Hikami, S., and E. Brezin, 1978, *J. Phys.* A **11**, 1141.
Hughes, R. J., 1981, *Nucl. Phys.* B **186**, 376.
Immirzi, G., 1996, *Class. Quantum Grav.* **13**, 2385–2393.
Itzykson, C., 1983, "Fields On A Random Lattice", in *Progress in Gauge Field Theory* (Cargése, France, September 1–15, 1983).
Itzykson, C., and M. Bander, 1983, in *Non-linear equations in Field theory*, Meudon Seminar (Springer, Verlag, 1983).
Itzykson, C., and J. -B., Zuber, 1980, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
Jevicki, A., and M. Ninomiya, 1985, *Phys. Lett.* B **150**, 115.
Julve, J., and M. Tonin, 1978, *Nuovo Cimento* **46** B, 137.
Kamenshchik, A., C. Kiefer and B. Sandhofer, 2007, *Phys. Rev.* D **76**, 064032.
Kaplan, D. B., and M. B. Wise, 2000, JHEP **0008**, 037.
Kawai, H., and M. Ninomiya, 1990, *Nucl. Phys.* B **336**, 115.
Kawai, H., Y. Kitazawa and M. Ninomiya, 1993a, *Nucl. Phys.* B **393**, 280.
Kawai, H., Y. Kitazawa and M. Ninomiya, 1993b, *Nucl. Phys.* B **404**, 684.
Khatsymovsky, V., 1991, *Class. Quantum Grav.* **8**, 1205–1216.
Kibble, T. W. B., 1961, *J. Math. Phys.* **2**, 212.
Kiefer, C., 2008, in the Proceedings of the Conference Quantum Gravity: Challenges and Perspectives, Bad Honneff, Hermann Nicolai (ed.), *Class. Quant. Grav.* **xx**.
Kleinert, H., 2000, *Phys. Lett.* A **277**, 205.
Kleinert, H., 2006, Path Integrals in Quantum Mechanics, Statistics and Polymer Physics (World Scientic, Singapore, 2006).
Kogut, J. B., and L. Susskind, 1975, *Phys. Rev.* D **11**, 395.
Kogut, J. B., 1983, Rev. *Mod. Phys.* **55**, 775.
Kraus, P., and F. Wilczek, 1995, Nucl. Phys. B **433**, 403.
Kojima, S. I., N. Sakai and Y. Tanii, 1994, Int. J. Mod. Phys. A **9**, 5415.
Kondo, K., 1984, *Prog. Theor. Phys.* **72**, 841.
Kosterlitz, J. M., and D. J. Thouless, 1973, *J. Phys.* CC **6**, 1181.
Kruskal, M. D., 1960, Phys. Rev. **119**, 1743–1745.
Kuchař, K., 1970, *J. Math. Phys.* **11**, 3322.
Kuchař, K., 1976, *J. Math. Phys.* **17**, 777.

Kuchař, K., 1992, "Time and Interpretations of Quantum Gravity", in the Proceedings of the 4th Canadian Conference on General Relativity and Relativistic Astrophysics, G. Kunstatter, D. Vincent and J. Williams (eds.) (World Scientic, Singapore, 1992).

Lafontaine, J., 1986, Seminaire Bourbaki, 38'eme annee, 1985–1986, n. 664.

Lanczos, C., 1938, *The Annals of Mathematics* **39**, 842.

Landau, L. D., and E. M. Lifshitz, 1980, *Statistical Physics*, 3rd edition (Pergamon press, Oxford, 1980).

Langer, J. S., 1967a, *Phys. Rev. Lett.* **21**, 973.

Langer, J. S., 1967b, *Ann. Phys.* **41**, 108.

Lauscher, O., and M. Reuter, 2002, *Class. Quant. Grav.* **19**, 483.

Le Guillou, J. C., and J. Zinn-Justin, 1989, *J. Phys.* (France) **50**, 1395.

Lee, T. D., 1983, "Discrete Mechanics", in *Erice International School of Subnuclear Physics*, vol. 21 (Plenum Press, New York 1985).

Lehto, M., H. B. Nielsen and M. Ninomiya, 1987, *Nucl. Phys.* B **289**, 684.

Lense, J., and H. Thirring, 1918, *Physik. Zeitsch.* **19**, 156.

Leutwyler, H., 1964, *Phys. Rev.* **134**, 1155.

Linde, A. D., 1998, Phys. Rev. D **58**, 083514.

Lipa, J. A., et al., 2003, *Phys. Rev.* B **68**, 174518.

Litim, D., 2004, *Phys. Rev. Lett.* **92**, 201301.

Loll, R., 2007, *Class. Quant. Grav.* **25**, 114006 (2008).

Lopez Nacir, D., and F. D. Mazzitelli, 2007, *Phys. Rev.* D **75**, 024003.

Lund, F., and T. Regge, 1974, Princeton preprint (unpublished).

MacDowell, S. W., and F. Mansouri, 1977a, *Phys. Rev. Lett.* **38**, 739.

MacDowell, S. W., and F. Mansouri, 1977b, *Phys. Rev. Lett.* **38**, 1376.

Mach, E., 1960, *The Science of Mechanics: A Critical and Historical Account of its Development* (Open Court Pub. Co., LaSalle, IL, 1960) (First published in 1906).

Mannion, C., and J. G. Taylor, 1981, *Phys. Lett.* **100** B, 261.

Meijering, J. L., 1953, *Philips Research Reports* **8**, 270.

Miller, W. A., and J. A. Wheeler, 1985, *Nuovo Cimento* **8**, 418–34.

Misner, C. W., 1957, *Rev. Mod. Phys.* **29**, 497.

Misner, C. W., K. S. Thorne, and J. A. Wheeler, 1973, *Gravitation*, (Freeman, San Francisco, 1973).

Modanese, G., 1993, *Phys. Rev.* D **47**, 502.

Modanese, G., 1995, *Nucl. Phys.* B **434**, 697.

Morris, T. R., 1994a, *Phys. Lett.* B **329**, 241.

Morris, T. R., 1994b, *Phys. Lett.* B **334**, 355.

Morris, T. R., and M. D. Turner, 1998, *Nucl. Phys.* B **509**, 637.

Münster, G., 1981, *Nucl. Phys.* B **190**, 439.

Münster, G., 1982, *Nucl. Phys.* B **200**, 536.

Myers, R. C., and M. J. Perry, 1986, *Annals Phys.* **172**, 304.

Nambu, Y., 1969, Proc. Intern. Conf. on Symmetries and Quark Models, Wayne State University.

Nicolai, H., and P. K. Townsend, 1981, *Phys. Lett.* B **98**, 257.

Nicolai, H., 1984, *Supersymmetric Functional Integration Measures*, Lectures given at 1984 NATO ASI Inst. on Supersymmetry, Bonn West Germany, August 20–31.

Nielsen, N. K., 1981, *Am. J. Phys.* **49**, 1171.

Ochiai, H., and K. Sato, 2000, *Prog. Theor. Phys.* **104**, 483.

Osterwalder, K., and R. Schrader, 1972, *Phys. Rev. Lett.* **29**, 1423–1425.

Osterwalder, K., and R. Schrader, 1973, *Comm. Math. Phys.* **31**, 83–112.

Osterwalder, K., and R. Schrader, 1975, *Comm. Math. Phys.* **42**, 281–305.

Page, D. N., 2002, arXiv:hep-th/0610121.

Painlevé, P., 1921, *C. R. Acad. Sci.* (Paris) **173**, 677–680.

Parikh, M. K., and F. Wilczek, 2000, *Phys. Rev. Lett.* **85**, 5042.

Parisi, G., 1975, *Nucl. Phys.* B **100**, 368.

Parisi, G., 1979, Phys. Rept. **49**, 215.
Parisi, G., 1981, *Statistical Field Theory* (Benjamin Cummings, New York, 1981).
Parisi, G., 1985, *Nucl. Phys.* **254**, 58.
Parisi, G., 1992, in the proceedings of the Conference on Chiral Gauge Theories (Rome, 1992) *Nucl. Phys. Proc. Suppl.* **29** BC, 247.
Pauli, W., 1958, *Teoria della Relativitá* (Boringhieri, Torino, 1958) English edition: *Theory of Relativity* (Dover Publications, 1956).
Peres, A., 1962, *Nuov. Cim.* **26**, 53.
Perez, A., 2003, Class. Quant. Grav. **20**, R43.
Peskin, M. E., and D. V. Schroeder, 1995, *An Introduction to Quantum Field Theory*, sec. 22.1 (Addison Wesley, Reading, Massachusetts, 1995).
Piran, T., and R. M. Williams, 1986, *Phys. Rev.* D **33**, 1622.
Polyakov, A. M., 1975, *Phys. Lett.* **59** B, 79.
Polyakov, A. M., 1979, *Phys. Lett.* B **82**, 247.
Polyakov, A. M., 1981a, *Phys. Lett.* B **103**, 207.
Polyakov, A. M., 1981b, *Phys. Lett.* B **103**, 211.
Polyakov, A. M., 1986, *Nucl. Phys.* B **268**, 406.
Polyakov, A. M., 1989, *Gauge Fields and Strings* (Oxford University Press, London,1989).
Ponzano, G. and T. Regge, 1968, in *Spectroscopic and Group Theoretical Methods in Physics*, F. Bloch, S. G. Cohen, A. De-Shalit, S. Sambursky and I. Talmi (eds.) (North-Holland, Amsterdam, 1968), 1–58.
Porter, J. D., 1987a, *Class. Quantum Grav.* **4**, 375.
Porter, J. D., 1987b, *Class. Quantum Grav.* **4**, 391.
Porter, J. D., 1987c, *Class. Quantum Grav.* **4**, 651.
Ramond P., 1990, *Field Theory: A Modern Primer*, revised printing (Benjamin/Cummings, Reading, MA, 1990).
Ray, D. B., and I. M. Singer, 1971, *Adv. Math.* **7**, 145.
Regge, T., 1961, *Nuovo Cimento* **19**, 558.
Reuter, M., 1998, *Phys. Rev.* D **57**, 971.
Reuter, M., and F. Saueressig, 2002, *Phys. Rev.* D **65**, 065016.
Reuter, M., 2008, in the Proceedings of the Conference Quantum Gravity: Challenges and Perspectives, Bad Honneff, Hermann Nicolai (ed.), *Class. Quant. Grav.* **xx**.
Richardson, J. L., 1979, *Phys. Lett.* B **82**, 272.
Riedler, J., W. Beirl, E. Bittner, A. Hauke, P. Homolka and H. Markum, 1999, *Class. Quant. Grav.* **16**, 1163.
Roček, M., and R. M. Williams, 1981, *Phys. Lett.* **104** B, 31.
Roček, M., and R. M. Williams, 1984, *Z. Phys.* C **21**, 371.
Roček, M., and R. M. Williams, 1985, *Class. Quantum Grav.* **2**, 701–706.
Rothe, H. J., 2005, *Lattice gauge theories: An Introduction*, World Sci. Lect. Notes Phys. **74**, 1.
Sakata, S., H. Umezawa and S. Kamefuchi, 1952, *Prog. Theor. Phys.* **7**, 377.
Salam, A., and J. Strathdee, 1978, *Phys. Rev.* D **18**, 4480.
Samko, S. G., A. A. Kilbas and O. I. Marichev, 1993, *Fractional Integrals and Derivatives: Theory and Application* (Gordon and Breach, Amserdam, 1993).
Scherk, J., and J. H. Schwarz, 1974, *Nucl. Phys.* B **81**, 118.
Schleich, K., 1985, *Phys. Rev.* D **32** 1889.
Schleich, K., 1987, *Phys. Rev.* D **36**, 2342.
Schwarz, J. H., 1982, *Phys. Rept.* **89**, 223.
Schwinger, J. S., 1951, *Phys. Rev.* **82**, 664.
Sciama, D. W., 1953, Monthly Notices of the Royal Astronomical Society 113, 34–42.
Sciama, D. W., 1971, *Modern Cosmology* (Cambridge University Press, England, 1971).
Singer, I. M., and J. A. Thorpe, 1967, *Lecture Notes in Elementary Topology and Geometry* (Scott, Foresman and Co., Glenview, Ill., 1967).
Slavnov, A. A., 1972, *Teor. Mat. Fiz.* **13**, 174.
Smolin, L., 1979, *Nucl. Phys.* B **148**, 333.

Smolin, L., 2003, “*Quantum Theory and Symmetries*”, in *Cincinnati*, 655–682: [hepth/0408048].
Sorkin, R. D., 1975a, *Phys. Rev.* D **12**, 385–396.
Sorkin, R. D., 1975b, *J. Math. Phys.* **16**, 2432.
Sorkin, R. D., 1996, arXiv:gr-qc/9701056.
Stelle, K. S., 1977, *Phys. Rev.* D **16**, 953.
Streater, R. F., and A. S. Wightman, 2000, *PCT, Spin and Statistics and All That* (Princeton University Press, Princeton, NJ, 2000).
Strominger, A., 1981, *Phys. Rev.* D **24**, 3082.
Symanzik, K., 1969, “Euclidean Quantum Field Theory”, in *Varenna Lectures*, R. Jost (ed.) (Academic Press, New York, 1969).
Symanzik, K., 1970, *Commun. Math. Phys.* **18**, 227.
Symanzik, K., 1983a, *Nucl. Phys.* B **226**, 187.
Symanzik, K., 1983b, *Nucl. Phys.* B **226**, 205.
’t Hooft, G., and M. Veltman, 1974, *Ann. Inst. Poincaré* **20**, 69.
’t Hooft, G., 1978, in *Recent Developments in Gravitation*, *Cargése Lecture notes* 1978, M. Levy and S. Deser (eds.) (Springer, Nato Science Series, 1979).
’t Hooft, G., 1985, *Nucl. Phys.* B **256**, 727.
’t Hooft, G., 2002, “Perturbative Quantum Gravity, Erice Lecture Notes”, in *Subnuclear Physics Series*, Vol. 40, A. Zichichi (ed.) (World Scientic, Singapore, 2002).
Taylor, T. R., and G. Veneziano, 1988, Phys. Lett. B **213**, 450.
Thiemann, T., 2007a, “Modern Canonical Quantum Gravity” (Cambridge University Press, 2007).
Thiemann, T., 2007b, *Lect. Notes Phys.* **721**, 185.
Tomboulis, E. T., 1984, *Phys. Rev. Lett.* **52**, 1173.
Tsao, H. S., 1977, Phys. Lett. B **68**, 79.
Tuckey, P. A., 1989, *Class. Quantum Grav.* **6**, 1–21.
Unruh, W. G., 1976, Phys. Rev. D **14**, 870.
Utiyama, R., and B. DeWitt, 1962, *J. Math. Phys.* **3**, 608.
Uzan, J. P., 2003, *Rev. Mod. Phys.* **75**, 403.
van Dam, H., and M. J. G. Veltman, 1970, *Nucl. Phys.* B **22**, 397.
van de Ven, A. E. M., 1992, *Nucl. Phys.* B **378**, 309.
van Nieuwenhuizen, P., 1973, *Nucl. Phys.* B **60**, 478.
Veltman, M., 1975, *Quantum Theory of Gravitation*, in *Methods in Field Theory*, *Les Houches Lecture notes session* XXVIII.
Vilenkin, A., 1998, *Phys. Rev.* D **58**, 067301.
Vilenkin, N. Ja., and A. U. Klimyk, 1993, *Representations of Lie Groups and Special Functions*, vol. 2, (Kluwer Academic Publishers, Dordrecht, 1993).
Vilkovisky, G. A., 1984, *Nucl. Phys.* B **234**, 125.
Voronoi, M. G., 1908, *J. Reine Angew. Math.* 134, 198–287.
Weinberg, S., 1972, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, (J. Wiley, 1972).
Weinberg, S., 1979, “Ultraviolet divergences in quantum gravity”, in *General Relativity – An Einstein Centenary Survey*, S. W. Hawking and W. Israel (eds.) (Cambridge University Press, Cambridge, 1979).
Wess, J., and J. Bagger, 1992, *Supersymmetry and Supergravity* (Princeton University Press, Princeton, 1992).
West, P., 1978, *Phys. Lett.* B **76**, 569.
Weyl, H., 1918, *Raum, Zeit, Materie – Vorlesungen über die Allgemeine Relativitätstheorie* (Springer, Berlin, 1918); English edition: *Space, Time, Matter* (Dover, New York 1952).
Wheeler, J. A., 1964, “Geometrodynamics and the Issue of the Final State”, in *Relativity, Groups and Topology*, B. DeWitt and C. DeWitt (eds.) (Gordon and Breach, New York, 1964) 463–500.
Whitt, B., 1984, *Phys. Lett.* B **145**, 176.
Williams, R. M., and G. F. R. Ellis, 1981, *Gen. Rel. Grav.* **13**, 361.
Williams, R. M., and G. F. R. Ellis, 1984, *Gen. Rel. Grav.* **16**, 1003.
Williams, R. M., and P. A. Tuckey, 1990, Class. Quant. Grav. **7**, 2055.

Williams, R. M., 1986, *Class. Quantum Grav.* **3**, 853–869.
Wilson, K. G., 1971a, *Phys. Rev.* B **4**, 3174.
Wilson, K. G., 1971b, *Phys. Rev.* B **4**, 3184.
Wilson, K. G., and M. E. Fisher, 1972, *Phys. Rev. Lett.* **28**, 240.
Wilson, K. G., 1973, *Phys. Rev.* D **7**, 2911.
Wilson, K. G., 1974, *Phys. Rev.* D **10**, 2445.
Wilson, K. G., 1975, *Rev. Mod. Phys.* **47**, 773.
Wong, C. Y., 1971, *J. Math. Phys.* **12**, 70.
Zakharov, V. I., 1970, *JETP Lett.* **12**, 312; *Pisma Zh. Eksp. Teor. Fiz.* **12**, 447 (1970).
Zavada, P., 1998, *Commun. Math. Phys.* **192**, 261 [funct-an/9608002].
Zinn-Justin, J., 2002, *Quantum Field Theory and Critical Phenomena*, fourth edition (Oxford University Press, New York, 2002).
Zinn-Justin, J., 2005, *Path Integrals in Quantum Mechanics* (Oxford University Press, Oxford, 2005).

# Index


$1/N$ expansion, 82
$2+\varepsilon$ expansion for gravity, 84
$3+1$ formulation of gravity, 107, 183
$O(2)$ symmetric spin system, 70, 78
$O(N)$ $\sigma$-model, 68, 79, 82
$SL(2,C)$ group, 217
$SO(3,1)$ group, 217
$SO(d)$ group, 175
$SU(N)$ gauge theories, 71, 127, 213, 248
$SU(N)$ generators, 131
$U(N)$-invariant self-coupled fermion model, 83
$\beta$-function for gauge theories, 98
$\beta$-function for gravity, 92, 274, 298
$\beta$-function for the $O(N)$ $\sigma$-model, 71, 81
$\varepsilon$ expansion, 67, 77
$d$-dimensional gravity, 11, 21, 260
$d$-simplex, 170
$n$-bein, 106, 171, 211, 214, 217
$n$-dimensional Riemannian manifold, 21
$p$-forms, 172

accelerated expansion, 317
Acknowledgements, xiii
action for a scalar field, 203
ADM formulation, 107
affine connection, 6, 188, 248, 250
anomalous dimension, 77, 274, 294
anti-De Sitter space, 260
anti-periodic boundary condition, 165
anti-screening, 308
anti-screening in QCD, 98
approximate renormalization group, 96
astrophysical observations, 300
asymptotic freedom, 30, 71, 98
asymptotic safety, 67
asymptotic series, 77
average curvature, 276, 300, 302
average edge length, 235
average hinge bivector, 256
average lattice spacing, 235, 264
average volume, 276
average volume of a simplex, 264
axial gauge, 129
axiomatic field theory, 58

background field method, 17
background gauge fixing, 86
background metric, 17, 29, 86, 145, 227
Baker-Hausdorff formula, 129
bare cosmological constant, 84, 180, 249, 262, 277, 288
bare Green's functions, 75
barycentric subdivision, 178, 199
Bianchi identities, 7, 85, 103, 109, 186
black hole entropy, 158
black hole evaporation, 166
black hole horizon, 154, 163
black hole radiance, 154, 158
black hole temperature, 155, 158
Bogoliubov transformation, 159
bordered determinant, 172
Borel transform, 47, 77
Bose distribution, 162
bosonic string, 24, 40
branched polymer, 93, 273
Brans-Dicke theory, 321
Brownian motion, 58

Calabi-Yau compactification, 51
Callan-Symanzik equations, 75
canonical commutation relations, 114
canonical form of the field equations, 104
canonically conjugate field, 127, 129

Cauchy problem, 103, 107
central charge, 40
change of basis in Fock space, 159
choice of coordinates, 172
choice of time coordinate, 107, 120
chromo-electric field, 128
classical field equations, 2, 10, 183, 301
classical gravity on a lattice, 183
closed Friedman universe, 119
closed strings, 50
closed surface, 247
collapsed lattice, 288
color singlets, 134
compact Euclidean four-geometries, 143
compact Euclidean metrics, 118
compact groups, 220
compact variables, 133
compactification assumption, 51
complex periodic time, 165
complex time, 58, 118, 163
conformal anomaly, 24
conformal gauge, 24
conformal instability, 64, 227, 264, 273
conformal invariance, 48
conformal mode, 64, 91
connected correlation function, 278
conserved energy momentum tensor, 1
constant curvature metric, 119, 235, 238, 260
constraint equations, 103, 107
constraints on metric fluctuation, 55, 86, 191
continuum limit, 134, 295
conventions, xii
coordinate conditions, 103, 110
coordinate independence, 306
corrections to purely thermal emission, 162
correlation between Wilson lines, 281
correlation function, 74, 100, 278
correlation length, 74, 134, 280, 295
correlation length exponent $\nu$, 76, 81, 91, 94, 95, 97, 269, 274, 290
correlations between curvatures, 245, 248, 258, 274, 278, 285
cosmological constant, 9, 10, 29, 63, 84, 106, 138, 180, 218, 249, 260, 262, 300, 302, 315, 321
cosmological constraints, 300
cosmological solutions, 314
cosmological wavefunction, 141
counting of microstates, 158
coupling renormalization, 87
covariant derivative, 106, 211, 215
critical coupling, 80, 82, 83, 90, 94, 264, 267, 284, 290, 294
critical dimension, 24
critical dimension of strings, 47, 50
critical exponent $\nu$, 76
critical exponents, 77, 78
critical point, 92, 264, 267, 284, 294, 295
cross polytope, 263
curvature, 173, 260
curvature correlation, 245, 248, 258, 274, 276, 278, 285
curvature fluctuation, 267, 276
curvature scales, 300, 302
curvature squared terms, 24, 198, 218, 234, 275, 320
cutoff, 16, 31, 48, 52, 56, 61–63, 68, 70, 71, 74, 139, 192, 195, 220, 242, 264, 267, 274, 278, 281, 287, 295, 296, 306
cutoff function, 97

d'Alambertian operator, 317
De Sitter space, 260
decoupling, 220
DeDonder gauge, 12, 225
deficit angle, 171, 262
degenerate configurations, 288
delta function, 18, 62, 69
density of states, 66
derivative expansion, 98
derivatives of the free energy, 276
determinant of the supermetric, 192
DeWitt measure, 59, 192, 196, 242, 249
DeWitt supermetric, 59, 114, 135, 192
diagrammtic expansion, 11
diffeomorphism constraint, 112, 118, 120
diffeomorphism invariance, 8, 184, 196, 209, 220, 234
dihedral angle, 171, 262
dihedral volume, 199
dilaton gravity, 48, 51
dimensional expansion, 67, 84
dimensional regularization, 11, 18
dimensional transmutation, 71, 300
dimensional transmutation in gauge theories, 98
dimensionful coupling constant, 16
Dirac $\gamma$-matrices, 34, 37, 176, 211, 214, 217
Dirac operator, 30, 153, 211
Dirac spin matrices, 34, 37, 215, 217
Dirichlet boundary conditions, 148
distance in function space, 114
dual lattice, 176, 178
dual of a simplex, 173, 176
dual subdivision, 199
dust, 123
dynamical triangulations, 219

effective action, 152, 221, 306
effective charge, 98
effective cosmological constant, 189, 260, 302
effective density, 320
effective field equations for gravity, 302, 305, 314
effective gravitational action, 320
effective pressure, 320
eigenfunctions, 126, 133
einbein, 195
Einstein tensor, 6
Einstein-Cartan theory, 217
Einstein-Hilbert action, 6, 11, 84, 108, 179, 217
electric field, 133, 152
electromagnetic interaction, 1
electrons, 154
elementary parallel transport loop, 173
energy momentum tensor, 1, 10, 112, 308
energy of the vacuum, 142
entropy, 267
entropy of a black hole, 158
Euclidean action, 56, 58, 179
Euclidean De Sitter space, 260
Euclidean Feynman path integral, 58, 63, 190, 274
Euclidean metric, 13
Euclidean rotation, 58
Euclidean Schwarzschild metric, 163
Euclidicity postulate, 58
Euler characteristic, 24, 47, 181, 198
exact lattice identities, 277
excluded volume effects, 267
expansion for critical exponents, 77
exponent $\nu$, 269, 274, 284, 290, 294
external electric field, 152
extrinsic curvature, 111
extrinsic geometry of surfaces, 47

Faddeev-Popov determinant, 145
Faddeev-Popov ghosts, 11, 86
failure of perturbation theory, 20, 67
Fermi distribution, 162
Fermi fields, 83, 106, 175, 211
ferromagnet, 70
Feynman $i\varepsilon$ prescription, 13, 58, 152, 157, 165
Feynman path integral, 55, 63, 84, 117, 143, 164, 190, 274
Feynman path integral for a non-relativistic particle, 56, 142
Feynman propagator, 12, 143, 225
Feynman rules for gravity, 11, 13
field equations, 2, 10, 103, 183
finite differences, 70
finite size scaling, 291
finite termperature field theory, 164
finite theories, 33, 37
first order formulation, 104, 108, 250
flat background, 86
flat Euclidean space, 59
flat hypercubic lattice, 216
foliation, 111
fractal dimension, 58, 167, 268
fractal horizon, 167
frame dragging, 321
free field behavior, 76
Friedman equations, 119, 320
Friedman-Robertson-Walker metric, 314
functional determinant, 59, 148
functional integral, 55, 180, 191, 209
functional measure, 47, 55, 59, 79, 150, 164, 191, 196, 207, 223, 242, 249, 277
functional supermetric, 59

gauge condition, 2, 127, 145, 147
gauge dependence, 85
gauge fields, 213
gauge fixing term, 11, 86
gauge group volume, 220
gauge invariance, 2, 8
gauge symmetry, 98
gauge theories, 98, 220, 248, 269
Gauss's law, 109, 131
Gauss-Bonnet identity, 13, 24, 46, 181
Gaussian fixed point in gravity, 98
Gaussian integral, 18, 144
general coordinate invariance, 8
general coordinate transformation, 8, 174, 196, 209
generators of the Lorentz group, 33, 106, 175
genus of a surface, 47
geodesic distance, 207, 258, 278
ghost field, 12, 15, 28, 44, 49, 225
ghost loop, 87
ghost-graviton vertex, 14
global $O(2)$ model, 71, 78
global $O(N)$ model, 68, 79
global $U(N)$ model, 83
glueballs, 100, 134
gluons, 134
gravitational action, 1, 6, 24, 37
gravitational anti-screening, 89, 285, 299, 305, 314
gravitational Compton scattering, 6
gravitational condensate, 300, 302, 315
gravitational correlation, 276, 295
gravitational correlation length, 321

gravitational exponent $\nu$, 94, 95, 269, 274, 284, 290, 294
gravitational field, 1
gravitational functional integral, 84
gravitational functional measure, 59
gravitational Wilson loop, 176, 188, 248, 301
gravitatonal vacuum polarization cloud, 308
gravitino, 37, 48, 214
graviton, 1
graviton correlations, 277
graviton exchange, 15
graviton helicity, 6
graviton loop, 15, 87
graviton mass, 9
graviton propagator, 3, 12, 28, 225
graviton self-interaction, 6, 14
gravity in $d$ dimensions, 11, 21, 260
gravity on a flat hypercubic lattice, 216
Green's functions, 70, 74, 75, 86
Gross-Neveu model, 83

Haar measure, 130, 189, 218, 220, 255
hadrons, 134
Hamilton-Jacobi equations, 115
Hamiltonian constraint, 112, 118
Hamiltonian for gauge theories, 127, 129
Hamiltonian for gravity, 111, 135
Hamiltonian formulation for gravity, 61
harmonic coordinate condition, 103
harmonic gauge, 2, 12, 225
harmonic oscillator, 43, 153
harmonic oscllator, 126
Hausdorff dimension, 167, 268
Hawking radiation, 154
Hawking temperature, 155, 158, 162, 165
Heisenberg ferromagnet, 70
heterotic string, 50
high temperature series, 244
higher derivative gravity, 19, 24, 30, 123, 198, 234, 275
higher derivative theory, 288
Hilbert space, 114
hinge, 172
hinge bivector, 177
Hirzebruch signature, 24, 198
holonomies in gauge theories, 100
homogeneous isotropic model, 119
horizon, 154, 163, 311, 321
hypercubic lattice, 70, 225
hyperspherical harmonics, 150

icosahedral tessellation, 236
imaginary time, 58, 118, 163
in and out vacua, 158
infinitesimal coordinate transformation, 8
infinitesimal loop in gauge theories, 100
infinities, 16
infrared cutoff, 287
infrared fixed points, 82
infrared regulator, 308
infrared renormalons, 47, 77, 96
infrared stability of gauge theories, 220
initial value problem, 103, 107
inner product of wave functionals, 114
integrated luminosity, 166
integration contours in the path integral, 65
intrinsic curvature, 111, 171
invariant correlation function, 100, 278
invariant local averages, 276
inverse of the DeWitt supermetric, 114
Ising model, 24, 41

Kaluza-Klein compactification, 51
kinematic singularities, 87
Kruskal coordinates, 154

lack of perturbative renormalizability, 16
Lagrangian for a point particle, 56
Landau diamagnetism, 39
Laplacian, 207
lapse and shift functions, 107
large $N$ expansion, 79
large scale curvature, 188, 248, 252, 260, 300, 302
large-$d$ expansion for gravity, 260, 294
lattice, 56, 61–63, 66
lattice action, 179
lattice Bianchi identities, 186
lattice continuum limit, 134, 295
lattice curvature, 171
lattice cutoff, 63, 139
lattice degrees of freedom, 171, 191, 225
lattice DeWitt measure, 196, 242, 249
lattice diffeomorphism invariance, 184, 196, 209, 220, 232, 234
lattice Fermions, 175, 211
lattice functional measure, 277
lattice gauge fixing, 225
lattice gauge theory, 100, 129, 187, 213, 220, 248
lattice graviton, 225
lattice Hamiltonian for gauge theories, 129
lattice Hamiltonian for gravity, 135
lattice Laplacian, 207
lattice measure, 191
lattice metric, 135, 170
lattice path integral, 190, 274
lattice Riemann tensor, 201

lattice supermetric, 135
lattice weak field expansion, 225
lattice Wheeler-DeWitt equation, 135
lattice zero modes, 227, 228, 234, 237, 264
least action principle, 56
Legendre transform, 96
Lehman representation, 134
length of a curve, 195
Lense-Thirring effect, 321
LEP experiment, 98
Lie derivative, 8
light-cone gauge, 40
linear potential, 134
linearized gravity, 1, 27, 110, 147, 225
Liouville field, 24, 47, 90
Lipschitz-Killing curvatures, 198
local averages, 276
local curvature, 173
local gauge invariance, 2, 8, 98, 184, 196, 209, 220, 234, 240
local rotations, 8
local scale transformation, 8
local shears, 8
logarithmic divergence, 314
long distance regime, 280
long range correlations, 220
loop bivector, 256
loop quantization, vii, 107
Lorentz frames, 105, 173
Lorentz group, 33, 106, 217

Mach's principle, 321
magnetic field, 133
Majorana spinors, 33, 48
mass gap, 74, 82, 84, 134
mass renormalization, 35
massless spin two field, 1
matter fields, 90, 112, 119, 123, 203
mean field theory, 254, 260
measure parameter, 62
method of steepest descent, 144
metric, 1
metric fluctuation, 55, 86, 191
metric perturbation, 145
metric signature, xii, 11
mini black holes, 166
minimal surface, 176, 188, 248
minisuperspace, 118
minisuperspace models, 122
modified field equations, 302, 305, 314
modified Newtonian potential, 15, 302, 314
modified static isotropic solution, 311
momentum constraint, 112, 118
momentum cutoff, 31, 62, 74, 220, 306

N=8 supergravity, 39
naive expectation for the running of $G$, 93
Nambu-Goto string, 40
negative curvature, 260
Newton's constant, 1, 10, 84
Newtonian limit, 22, 121
Nielsen-Hughes formula, 39
non-abelian gauge theories, 98
non-analytic contributions, 77
non-analyticity in the partition function, 245, 284, 295
non-compact groups, 220
non-linear $\sigma$-model, 68, 79, 298
non-linear nature of gravity, 6
non-local effective action, 306
non-perturbative scale $\Lambda_{\overline{MS}}$, 98
non-perturbative vacuum contribution, 302
non-relativistic matter, 123
non-relativistic path integral, 56, 142
non-renormalizable theories, 16, 67
non-rotating black hole, 154
non-trivial ultraviolet fixed point, 67, 76, 84, 93, 295
numerical simulations, 273

observables, 276
octahedral tessellation, 236
on-shell counterterm, 19
one-dimensional Coulomb problem, 66
one-dimensional gravity, 195, 209
one-loop beta function, 39
one-loop divergences, 16, 31
one-loop divergences for gravity, 19, 39, 91
one-loop effective action, 31
one-loop renormalization group $\beta$-function, 73, 89, 98
open strings, 50
operator ordering ambiguity, 40, 114, 120, 123
order parameter, 77
origin of inertia, 321
oscillatory wavefunction, 123

Padé approximants, 244
Painlevé coordinates, 156
pair creation, 152, 154
Palatini formulation, 104, 180
Palatini identity, 105
parallel transport, 173, 188, 248
particle orbits, 314
particle thermal emission, 155, 158, 162
partition function, 135
path ordered exponential, 100
Pauli paramagnetism, 39
Pauli-Fierz mass term, 9

perfect fluid, 311
periodic complex time, 165
perpendicular bisectors, 178
perturbatively non-renormalizable theories, 67
phase transition, 67, 82, 92, 96, 295
phases of gravity, 92, 273, 285
physical degrees of freedom, 103, 127
physical phase, 285
piecewise linear space, 169, 198
planar loop, 251
planar rotator, 70
Planck length, 29, 167
plane wave solutions, 4, 23, 227
plaquette, 130, 187, 244, 253
Poincaré group, 33
Poisson bracket, 110, 111
Poisson's equation, 22, 308
Polyakov string, 40
polygonal area, 179
polygonal random surface, 267
polyhedral cells, 199
polyhedron, 169
polyogonal path, 176
position representation, 114, 135, 306
positive curvature, 260
positrons, 154
preface, vii
primary constraint, 127
primordial black holes, 166
problem of time, 114
propagator, 3, 12, 28, 45, 143, 207, 225
pseudo-Riemannian geometry, 11

QCD $\beta$-function, 101
quadratic Lagrangian, 11
quantized flux, 133
quantum averages, 276
quantum black hole, 167
Quantum Chromodynamics (QCD), 98, 127, 129
quantum commutator, 111
Quantum Electrodynamics (QED), 98, 109, 128, 133, 152, 167, 281, 301, 310, 314
quantum fluctuations, 66
quantum gravity and Mach's principle, 321
quantum horizon, 167
quantum mechanical amplitude, 150
quantum operators, 114
quantum spin variables, 219
quantum state of the universe, 118
quantum-mechanical amplitude, 56
quark confinement, 41, 134

radiative correction to the potential, 15
radiative correction to the static potential, 281, 314
random surfaces, 24, 267
random triangulation, 268
random variables, 266
random walk, 268
Rarita-Schwinger field, 37, 214
references, 325
Regge action, 179
Regge discretization, 169
Regge equations, 183
regularized path integral, 190
relativistic pressure, 311
relativistic string, 41
renormalization group, 75, 89, 98, 274, 291, 295, 305, 314
renormalization group $\beta$-function, 73, 89, 92, 98
renormalization group invariant, 75, 76, 81, 94, 298
renormalization of Newton's constant, 87
renormalization of the cosmological constant, 87
renormalized cosmological constant, 260
renormalized Green's functions, 75
renormalons, 77, 96
reparametrization invariance, 184, 196, 209, 220, 234
rescaling of the metric, 84, 196, 277
resummation of the perturbative series, 79
Ricci tensor, 6
Riemann squared terms, 24, 198
Riemann tensor, 6, 201, 217
Riemann tensor at a hinge, 178
Riemannian geometry, 13
Rindler coordinates, 163
Robertson-Walker metric, 314
rotation around a hinge, 177
rotation group, 175
rotations, 8, 173, 188, 211, 248, 252
running coupling constant, 73, 89, 305, 314
running coupling in gauge theories, 98
running Newton's constant, 16, 89, 298, 305, 310, 314, 320
running of the fine structure constant, 98

S-matrix, 158
saddle points, 80, 165
scalar curvature, 6
scalar field, 24, 61, 90, 119, 125, 203, 269, 301
scalar field action, 11, 13
scalar field loop, 15
scalar field propagator, 207
scalar matter loop, 87

scalar-graviton vertex, 14
scalar-tensor theories, 27, 320, 321
scale dependence of Newton's constant, 305, 310, 314
scale dependent couplings, 32, 82
scale dependent Newton's constant, 314, 320
scale factor, 121, 317
scale invariance, 94
scale invariant measure for gravity, 59
scale transformation, 8, 291
scale transformations of the metric, 64, 84, 196, 277
scaling behavior, 284, 290, 294, 295
scaling dimension, 76, 82, 284, 290, 294, 295
scaling properties of the functional integral, 84, 196, 277
scaling violation parameter $\Lambda_{\overline{MS}}$, 98
Schrödinger equation, 114
Schwarzschild metric, 154, 163
Schwarzschild solution, 311
screening, 300
screening in QED, 98
second order phase transition, 67
secondary constraint, 127
self-adjointness requirement, 119
self-coupled fermion model, 83
self-dual connection, 107
self-interacting scalar field, 125, 301
self-similar geometry, 167
semiclassical expansion, 115, 144, 146, 152, 156, 158
semiclassical gravity, 141
shears, 8
short distance cutoff, 56, 61, 152
short distance fluctuations in the metric, 167
short distance regime, 280
short range correlations, 220
signature, xii, 11
signature of the DeWitt supermetric, 114
simlicial gravity, 169
simpicial complex, 169
simplex, 191, 260
simplicial gravity action, 179
simplicial manifold, 263
singular part of the free energy, 245, 284, 295
singularity avoidance, 125
singularity in the scale factor, 122
size of visible universe, 10, 321
skeleton, 169
small breaking of gauge invariance, 220
source terms, 112
space shuttle experiment, 78
spacelike hypersurface, 104
spacetime lattice, 56, 61–63
spacetime supersymmetry, 33
spectrum of the Laplacian, 209
spin connection, 105
spin foam, 219
spin group, 175
spin of the graviton, 6
spin one gauge mode, 110
spin projection operators, 27, 110
spin rotation, 176, 211
spin two ghost, 28
spin zero tachyon, 28
Spin$(d)$ group, 175
spin-3/2 field, 37, 214
spinor field, 175, 211
stable ground state, 273
state of minimum excitation, 142
static isotropic metric, 154
static isotropic solution, 311
static Newtonian potential, 281, 308, 314
static potential in gauge theories, 98, 134, 310
statistical field theory, 59
statistical mechanics Hamiltonian, 164
Stefan-Boltzmann law, 166
string model for hadrons, 41
string tension, 248
string theory, vii, 24, 25, 40, 48
strong coupling expansion, 135, 242, 248
strong coupling phase, 288
strong electric fields, 152
sub-simplex, 191
sum over histories, 55
sum over saddle points, 165
sum over topologies, 47
sum rules, 277
super-conformal gauge, 50
super-Liouville theory, 50
supercharges, 50
superconductors, 70
superficial degree of divergence, 16
superfluid Helium transition, 70, 78
supergravity, vii, 37, 91, 214
supermetric, 59, 118, 135, 192
supersymmetric charges, 35
supersymmetric strings, 48
supersymmetry, 33
surface of the cross polytope, 263
surface self-intersection, 268
susy Yang-Mills, 36

tachyon instability, 48
ten-dimensional string theory, 51
tetrad, 105, 106, 171, 211, 214, 217
tetrahedral tessellation, 235
thermal emission, 158, 162

thermodynamic analogy, 158, 165
Thirring interaction, 90
three-dimensional gravity, 21, 182
three-graviton vertex, 14
three-metric, 104, 136
three-sphere, 146
tiling of minimal surface, 188, 248
time coordinate, 107, 115, 120
time evolution, 103, 107
torsion, 175
total mass of the visible universe, 321
transfer matrix, 130
transverse traceless modes, 4, 110, 147, 225
triangle inequalities, 138, 140, 171, 191
trivial infrared fixed points, 82
triviality in supersymmetric theories, 39
triviality of scalar field theory, 125, 269, 294
truncated renormalization group, 96
truncation of degrees of freedom, 125
tunneling amplitude, 152, 154, 156
two-dimensional gravity, 21, 24, 181, 234
two-loop divergences for gravity, 91
two-phase structure, 82, 92, 264, 284, 294
two-point function, 74, 278

Uehling radiative corrrection, 98
ulraviolet cancellations, 39
ultra-local terms, 61
ultralocal term, 18, 62, 69
ultraviolet behavior, 16
ultraviolet cutoff, 16, 31, 48, 52, 56, 63, 68, 70, 71, 74, 150, 189, 192, 195, 220, 264, 274, 278, 281, 287, 295, 296, 306
ultraviolet fixed point, 30, 98, 220, 295, 305
ultraviolet regularization, 150
uniformly accelerated observer, 163
unitarity in higher derivative gravity, 28
universal exponent $\nu$, 94, 95, 269, 274, 284, 290, 294
universal scaling properties, 268, 274, 284, 290, 294, 295
unphysical degrees of freedom, 127
unphysical phase, 285
Unruh effect, 163

vacumm polarization in QED, 98
vacuum contribution, 302
vacuum energy density, 10, 311
vacuum instability, 47
vacuum polarization, 30
vacuum state, 133, 156, 159, 163
vacuum to vacuum amplitude, 142
van Dam - Veltman discontinuity, 9
vierbein, 105, 106, 171, 214, 217
Virasoro algebra, 40
virtual graviton cloud, 308, 314
virtual particle pair, 154
visible universe, 10
volume dependence, 291
volume of a face, 172
volume of a simplex, 171
Voronoi loop, 173
Voronoi polyhedra, 199
Voronoi volume, 205

Ward identities, 2, 185, 232
wave function of the universe, 141
wave functional, 114, 117, 141, 148
weak coupling phase, 288, 300
weak field expansion, 1, 225
Weyl conformal tensor, 24, 202
Weyl spinors, 33
Weyl symmetry, 48
Wheeler-DeWitt equation, 114, 135, 141
Wick rotation, 13, 58, 118, 163, 165
Wiener paths, 58
Wilson line, 281
Wilson loop, 100, 188, 248, 301
WKB method, 115, 144, 146, 152, 156, 158
world lines, 281
world sheet coordinates, 40
world-sheet supersymmetry, 48

Yang-Mills theories, 71, 84, 98, 127, 248

zero energy Schrödinger equation, 114, 141
zero lattice spacing limit, 295
zero modes, 11, 227, 228, 234, 237, 264
zeta function regularization, 19, 148
zig-zag motion of a quantum particle, 58, 66