Bernhelm Booß-Bavnbek
Giampiero Esposito
Matthias Lesch
Editors

LECTURE NOTES IN PHYSICS 807

New Paths Towards Quantum Gravity

Springer

Lecture Notes in Physics

The Lecture Notes in Physics

The series Lecture Notes in Physics (LNP), founded in 1969, reports new developments in physics research and teaching – quickly and informally, but with a high quality and the explicit aim to summarize and communicate current knowledge in an accessible way. Books published in this series are conceived as bridging material between advanced graduate textbooks and the forefront of research and to serve three purposes:

- to be a compact and modern up-to-date source of reference on a well-defined topic
- to serve as an accessible introduction to the field to postgraduate students and nonspecialist researchers from related areas
- to be a source of advanced teaching material for specialized seminars, courses and schools

Both monographs and multi-author volumes will be considered for publication. Edited volumes should, however, consist of a very limited number of contributions only. Proceedings will not be considered for LNP.

Volumes published in LNP are disseminated both in print and in electronic formats, the electronic archive being available at springerlink.com. The series content is indexed, abstracted and referenced by many abstracting and information services, bibliographic networks, subscription agencies, library networks, and consortia.

Proposals should be sent to a member of the Editorial Board, or directly to the managing editor at Springer:

Christian Caron
Springer Heidelberg
Physics Editorial Department I
Tiergartenstrasse 17
69121 Heidelberg / Germany
christian.caron@springer.com

Bernhelm Booß-Bavnbek
Giampiero Esposito
Matthias Lesch (Eds.)

New Paths Towards Quantum Gravity

Springer

Bernhelm Booß-Bavnbek
Roskilde Universitet
IMFUFA, Institut for Natur
Systemer og Modeller (NSM)
Postboks 260
4000 Roskilde
Denmark
booss@ruc.dk

Giampiero Esposito
Istituto Nazionale di Fisica
Nucleare
Sezione di Napoli
Complesso Universitario di
Monte S. Angelo
Edificio N / Via Cintia
80126 Napoli
Italy

Matthias Lesch
Mathematisches Institut
Universität Bonn
Endenicher Allee 60
53115 Bonn
Germany
ml@matthiaslesch.de
lesch@math.uni-bonn.de

Booß-Bavnbek B., et al. (Eds.), *New Paths Towards Quantum Gravity*, Lect. Notes Phys. 807 (Springer, Berlin Heidelberg 2010), DOI 10.1007/978-3-642-11897-5

Lecture Notes in Physics ISSN 0075-8450 e-ISSN 1616-6361
ISBN 978-3-642-11896-8 e-ISBN 978-3-642-11897-5
DOI 10.1007/978-3-642-11897-5
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010922996

Cover design: Integra Software Services Pvt. Ltd. Pondicherry

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

“We must know - we will know!”

David Hilbert, 8th of September 1930
Address to the Society of German
Scientists and Physicians at Königsberg

“I still need to find help”, said K. “You look for too much help from people you don’t know”, said the priest disapprovingly. “Can you really not see that’s not the help you need?”

Franz Kafka, The Trial, undated
Translation by David Wyllie

Foreword

This volume contains the extended proceedings of the Summer School: *New Paths Towards Quantum Gravity* held in Holbæk, in Denmark. I happened to be the "oldest" participant in the school, managing an invitation thanks to my role as member of the advisory committee, since justly the school was reserved to students and young post-docs. The volume reflects nicely the atmosphere and excellent level of the school. I hope none of my colleagues, either mathematicians or physicists, will feel offended if I say that we really do not know what quantum gravity is, beyond the obvious statement that it should be a theory capable of unifying general relativity and quantum field theory. We are at a stage in which new ideas are produced at a rate such that often they seem to diverge to a point in which no unequivocal ways to decide the best direction is seen. This means that we are in an exciting moment for research in the field.

As always in physics we have two guiding principles, experimental data and mathematical rigour. But both of these come with a novel twist. Experimental results are difficult to get in a direct way for lack of energy, but this does not mean that a good theory does not have immediate check, cosmological observations are a new key tool, but also in the low energy regime there is still a lot to learn. Mathematical rigour is as always a good guide, this volume shows that there is also the need for the development of a new kind of mathematics which should accompany this new physics.

The division of the book into two (overlapping) parts reflects the duality between the physical vision and the mathematical construction. The noncommutative geometry vision of Gracia-Bondìa, the discretization and renormalization one of Ambjørn, Jurkiewicz, and Loll, and the gauge fields path integrals of Reshetikhin give three examples of promising paths towards quantum gravity. I am sure that some parts of all three points of view will be components of the larger picture.

The second part is an overview of the mathematical results. It comprises the write-up of Bouwknegt lectures on cohomology, plus the welcome additions of two contributions not originally present in the school by Zessin on stochastic geometry, and Avramidi's on tools for the effective action in quantum gravity.

One crucial part of the school had been the discussion sections, and reading the various contributions I have seen how this percolated in them, thus enriching them. As a "student" of the school I have witnessed the appreciation that my fellow students have had of the quality of the lectures, and as one of the first readers of the proceedings I am sure that this appreciation will be replicated by the readers of the volume.

Napoli, August 2009 Fedele Lizzi

Preface

In this volume we have collected a package of teaching materials which arouse from a Summer School on *New Paths Towards Quantum Gravity*, held at the Field Laboratory of Roskilde University at Holbæk Bay in Denmark in May 2008 for an audience of 30 Ph.D. students and (mostly young) postdoctoral researchers and disseminated at `http://QuantumGravity.ruc.dk`.

We organized the Summer School out of curiosity. It seemed to us that something radically new was going on: was there a new perception of physics reality evolving? Were there the most delicate advances of our own fields (spectral geometry, global analysis, non-commutative algebras, deformation theory) gaining application and new impulses? Was the interplay between mathematics and physics once again changing? Moreover, we were aware of public concerns regarding asteroid and comet collision, forced black holes, the evidence of accelerated expansion of the universe, the rational explanation of miracles by the supposed Higgs particle. We had read about the personal concerns of outstanding physicists and mathematicians like Lee Smolin [Sm06] and John Baez [Ba07] who warned against a distortion of physics by exaggerated cracked-up expectations. We shared their concern about pretended physics authority borrowed from the intricacy of the involved mathematics. We simply wanted to expand views, open perspectives, and invite curiosity.

We were so lucky to find outstanding and dedicated teachers and discussants with J. Ambjørn, I. Avramidi, P. Bouwknegt, J. Gracia-Bondía, A. Klemm (not represented in this volume–we refer, however, to B. Haghighat and A. Klemm [HaKl08] and the references given there), N. Reshetikhin, and an explanatory supplement by H. Zessin.

The Organization of the Teaching Material

We have grouped the lectures into three parts. Part I is devoted to *Three Physics Visions*, presented by J.M. Gracia-Bondía; J. Ambjørn, J. Jurkiewicz, and R. Loll, and finally N. Reshetikhin. Part II is devoted to *Novel Mathematical Tools* by I.G. Avramidi, P. Bouwknegt, and H. Zessin. The partition is not sharp, as the reader will see. Part III is devoted to *Afterthoughts on the Merger of Mathematics and Physics* presented by one of us (BBB).

To Part I – Three Physics Visions

J.M. Gracia-Bondía takes a global and phenomenological approach to the subject matter. Clearly he wishes preventing students against high-road ideological approaches, all too frequent in this field. As a useful vaccine, as well on their own right, he reviews the main experimental and observational issues relating directly to quantum gravity and gives his personal view. This includes discussion of the Pioneer and flyby anomalies in the Solar System, as well as astrophysical observations, and table-top experiments with cold neutron interferometry, like the Colella–Overhauser–Werner experiment and its progeny. Here he mainly refers to the experiments by Nesvizhevsky and his group, which have allowed detection of discrete quantum gravitational states for the first time. He points out that these experiments already seem to contradict some of the current orthodoxy in quantum gravity.

Throughout, he emphasizes that "quantum gravity" denotes more of a problem than of a theory. According to Gracia-Bondía, there is no such theory, but several competing schemes, mathematically sophisticated as a rule, but underdeveloped in the face of experimental evidence and the aim of unifying gravity with other fundamental interactions.

In the following sections he looks at the (Einstein–Hilbert) action of general relativity as a consequence of gauge theory for quantum tensor fields. He performs a parallel pedagogical exercise to Feynman's in his Pasadena lectures in the 1960s: assuming ignorance of Einstein's general relativity, one arrives anew at it by successive approximation. The method, however, differs from the traditional in several respects:

1. He considers only pure gravity. Coupling to matter is sketched only after the fact, for completeness.
2. This is fully quantum field theoretical in that it recruits the canonical formalism on Fock space and quantum gauge invariance. The main tool is BRS technology, and ghost fields are introduced from the outset. In other words, he treats gravity as any other gauge theory; at some point one puts the Planck constant equal to zero in this quantum model.
3. He uses the causal (or Epstein–Glaser) renormalization scheme.
4. He never invokes the stress–energy tensor.

Next, he gives a general introduction to non-commutative geometry. He explains the Doplicher–Fredenhagen–Robert Gedanken experiment to the effect that spacetime must become "fuzzy" at the Planck length scale, if not before (thus, it is not infinitely divisible). He justifies that on these circumstances commutative and non-commutative manifolds must be treated on the same footing, in the putative functional integral for gravity. This may justify the place granted, in the same context, to the reconstruction theorem of ordinary manifolds from spectral data by Alain Connes. He also explains the most efficient procedure known at present for constructing non-commutative manifolds: isospectral deformation. Non-commutative field theory on Moyal product algebras is also briefly discussed in that section.

Gracia-Bondía devotes some final considerations to the cosmological constant problem and the unimodular theories of gravity.

J. Ambjørn, J. Jurkiewicz, and R. Loll provide quite a different approach. They ask Why do we study two-dimensional quantum gravity? They give two answers: First, one can test quantization procedures for gravity in a simple setting. Second, it has long been known that string theory can be viewed as two-dimensional quantum gravity coupled to matter fields. This particular view of string theory spawned the development of the dynamical triangulation approach to quantum gravity. This method is particularly powerful in two dimensions, since exact non-perturbative solutions can be obtained by loop equations, matrix models, and closely related methods of stochastic geometry. Their lectures span over a wide range, highlighting

1. Computer simulation of four-dimensional quantum gravity.
2. Geometric quantization of the point particles.
3. Geometric quantization of two-dimensional quantum gravity.
4. The combinatorial solution of two-dimensional quantum gravity and its relation to matrix models.
5. Generalized Hartle–Hawking wave functions.
6. Defining geodesic distance in quantum gravity and calculating the so-called two-point functions.
7. The relation to causal dynamical triangulations.

N. Reshetikhin gives a survey of quantization methods for gauge theories, from quantum mechanics to Yang–Mills and Chern–Simons theories. First, the author formalizes the usual physical concepts in a rigorous mathematical setting, and then he describes the quantization methods, from the canonical one to Faddeev–Popov and BRST methods (also mentioning the BV formalism). In our view, this is a very important area in mathematical physics. Without a rigorous definition of the functional (Feynman) integral the current mathematical status of quantum field theory remains at the naive phenomenological level, which allows one to carry out calculation in perturbative renormalizable theories (which form a set of measure zero in the set of all possible theories) but fails miserably for non-renormalizable theories as well as in strong-coupling regimes, even in renormalizable theories (like in the infrared QCD). Therefore, a thorough understanding of the quantization, in particular, in complicated modern gauge field theories plays a crucial role in the program of developing a satisfying theory of quantum gravitational phenomena.

To Part II – Novel Mathematical Tools

I.G. Avramidi reviews the status of covariant methods in quantum field theory and quantum gravity, in particular, some recent progress in the calculation of the effective action via the heat kernel method. He studies the heat kernel associated with an elliptic second-order partial differential operator of Laplace type acting on smooth sections of a vector bundle over a Riemannian manifold without boundary. He recalls the general knowledge about that topic and develops a manifestly covariant method for computation of the heat kernel asymptotic expansion as well as new algebraic methods for calculation of the heat kernel for covariantly constant background, in particular, on homogeneous bundles over symmetric spaces. That enables one to compute the low-energy non-perturbative effective action.

P. Bouwknegt introduces various, mostly novel, mathematical concepts and explains how they are applied in modern quantum field theory and string theory. Each section begins by discussing a physical example, which suggests a particular mathematical framework which is then subsequently developed. The emphasis is more on concepts than on rigorous mathematical detail and aimed to be a first introduction to various modern mathematical techniques for beginning postgraduate students. It is assumed that the students have basic understanding of differential geometry, algebra, and quantum field theory.

The following topics are treated in detail:

1. Cohomology and differential characters. Using the electromagnetic field as an example, the basic definitions are given of de Rham cohomology, Cech cohomology, and their relation through the Cech–de Rham complex. Various definitions of differential cohomology theories are given, such as Deligne cohomology and Cheeger–Simons differential characters. According to Bouwknegt, an example of a third-order Deligne cohomology class occurs naturally in string theory, where it is known as the Kalb–Ramond (or Neveu–Schwarz) B-field. This B-field is the connecting theme throughout the sections.
2. T-duality. A physical introduction into T-duality is given and the Buscher rules for the T-duality of closed strings in a curved background with S^1 isometry are derived. From here the topological properties of T-duality are derived and put in the context of Gysin sequences for principal circle bundles. Principal torus bundles are discussed in a similar way. According to Bouwknegt, T-duality in that case naturally leads to non-commutative and possibly nonassociative structures.
3. Generalized geometry. An elementary introduction to (Hitchin's) generalized geometry is given, including a basic discussion of the Courant bracket, spinors, generalized complex manifolds, generalized Kähler manifolds, etc. Applications in the context of T-duality are discussed.

H. Zessin provides a short introduction into some recent developments in stochastic geometry. It has one of its origins in simplicial gravity theory and may be considered as an elaboration of some aspects of the presentation by J. Ambjørn and collaborators in Part I. Zessin's aim is to define and construct rigorously point processes on spaces of simplices in Euclidean space in such a way that the configurations of these simplices are simplicial complexes. The main interest then is concentrated on their curvature properties. He illustrates certain basic ideas from a mathematical point of view and recalls the concepts and notations used. He presents the fundamental zero-infinity law of stochastic geometry and the construction of cluster processes based on it. Next, he presents the main mathematical object, i.e. Poisson–Delaunay surfaces possessing an intrinsic random metric structure. He finally discusses their ergodic behaviour and presents the 2-dimensional Regge model of pure simplicial quantum gravity.

Contrary to the Ambjørn–Jurkiewicz–Loll (AJL) contribution, Zessin's lectures are written in a dense mathematical style and strongly formalized. They show how certain concepts of stochastic geometry, which play a role in physics, can be defined in a theoretically clear and rigorous way and how precise results can be derived unambiguously. Clearly, a graduate student in mathematics will perceive Zessin's lectures as concrete and lucid and AJL as rather abstract and difficult to grasp, while a graduate student of physics may have an opposite appreciation.

To Part III – Afterthoughts

As a service to the reader, we reproduce an edited version of the opening remarks by one of us (BBB) to the International Workshop *Quantum Gravity: An Assessment* which followed immediately after the Summer School at the same place. The point was to recall common knowledge on modelling, mathematization, and science history and to address the ethics of "quantum gravity" research. That may help to put the lectures of this volume in a common frame in spite of their scattering and heterogeneity. Some of these considerations were published in [BEL07] in condensed form before the Summer School as a kind of *platform* for the assessment of our endeavour. For a comparison with mathematization in other frontier fields of research we refer to the recent [Bo09].

Acknowledgements The editors are indebted to the sponsoring agencies of the Summer School: The European Science Foundation (ESF) Activity *Quantum Geometry and Quantum Gravity*; The Centre for the Topology and Quantization of Moduli spaces (CTQM) and The Geometry and Mathematical Physics School (GEOMAPS), both located at Aarhus; The Danish Network *Mathematical Physics and Partial Differential Equations*; the Danish Centre in Non-Commutative Geometry located at Copenhagen; and Roskilde University. We thank the students who have participated. Their questions played a significant part for the final shaping of these Lecture Notes. We thank Michael Bohn and Benjamin Himpel for their creativity and order in the thousands of practicalities of the summer school announcement, student selection, and implementation. We thank our scientific advisers who did an extremely important job during the preparation of the Summer School and afterwards by their help in evaluating and editing the received contributions. We wish to thank in particular Irina Araf'eva, John Barrett, Peter Gilkey, Koji Hasegawa, Albrecht Klemm, Fedele Lizzi, Jouko Mickelsson, Werner Nagel, Ryszard Nest, Holger Bech Nielsen, Daniel Sternheimer, and Ruth Williams for extensive consultations. Foremost, however, we thank the lecturers and contributors for their ingenuity and patience in providing the teaching material and accommodating with suggestions and criticism in a friendly and extremely productive way.

References

[Ba07] John Baez, This Week's Finds, Work on quantum gravity has seemed stagnant and stuck for the last couple of years, which is why I've been turning more towards pure math, Week 220 (August 31, 2005), `http://math.ucr.edu/home/baez/week220.html`

[Bo09] Bernhelm Booß-Bavnbek, The mathematization of the individual sciences–revisited, *Journal of Contemporary Mathematical Analysis (Armenian Academy of Sciences)* **44**/2 (April 2009), 81–96, [arXiv:0903.0942v1 [math.HO]]. Russian edition: Izvestiya NAN Armenii. Matematika **44**/2 (2009) 17–38.

[BEL07] Bernhelm Booß-Bavnbek, Giampiero Esposito, Matthias Lesch, Quantum gravity: unification of principles and interactions, and promises of spectral geometry, *SIGMA (Symmetry, Integrability and Geometry: Methods and Applications)* **3** (2007), 098–127, [arXiv:0708.1705v3 [hep-th]].
[HaKl08] Babak Haghighat, Albrecht Klemm, Topological Strings on Grassmannian Calabi-Yau manifolds, JHEP 0901:029-65 (2009) [arXiv:0802.2908v1 [hep-th]].
[Sm06] Lee Smolin, *The Trouble with Physics*, Allen Lane, Penguin Books, London, (2006).

Roskilde, Napoli, Bonn
September 2009

Bernhelm Booß-Bavnbek
Giampiero Esposito
Matthias Lesch

Contents

Chapter 1
Notes on “Quantum Gravity” and Noncommutative Geometry

J.M. Gracia-Bondía

Abstract I hesitated for a long time before giving shape to these notes, originally intended for preliminary reading by the attendees to the Summer School “New paths towards quantum gravity” (Holbaek Bay, Denmark, May 2008). At the end, I decide against just selling my mathematical wares, and for a survey, necessarily very selective, but taking a global phenomenological approach to its subject matter. After all, noncommutative geometry does not purport yet to solve the riddle of quantum gravity; it is more of an insurance policy against the probable failure of the other approaches. The plan is as follows: the introduction invites students to the fruitful doubts and conundrums besetting the application of even classical gravity. Next, the first experiments detecting quantum gravitational states inoculate us a healthy dose of scepticism on some of the current ideologies. In Sect. 1.3 we look at the action for general relativity as a consequence of gauge theory for quantum tensor fields. Section 1.4 briefly deals with the unimodular variants. Section 1.5 arrives at noncommutative geometry. I am convinced that, if this is to play a role in quantum gravity, commutative and noncommutative manifolds must be treated on the same footing, which justifies the place granted to the reconstruction theorem. Together with Sect. 1.3, this part constitutes the main body of the notes. Only very summarily at the end of this section do we point to some approaches to gravity within the noncommutative realm. The last section delivers a last dose of scepticism. My efforts will have been rewarded if someone from the young generation learns to mistrust current mindsets.

1.1 Introduction

“Quantum gravity” denotes a problem, not a theory. There is no theory of quantum gravity. There exist several competing schemes, as mathematically sophisticated and fecund, as a rule, as undeveloped in the face of experimental evidence and of the purported aim of unifying gravity with other fundamental interactions.

J.M. Gracia-Bondía (✉)
Departamento de Física Teórica, Universidad de Zaragoza, Zaragoza 50009, Spain
Departamento de Física, Universidad de Costa Rica, San Pedro 2060, Costa Rica,
jmgb@fis.ucm.es

Gracia-Bondía, J.M.: *Notes on “Quantum Gravity” and Noncommutative Geometry*. Lect. Notes Phys. **807**, 3–58 (2010)
DOI 10.1007/978-3-642-11897-5_1

My account of the subject is unabashedly *low road*. The concept was coined by Glashow in his thought-provoking book [1]. The low road

> is the path from the laboratory to the blackboard, from experiment to theory, from hard-won empirical observations to the mathematical framework in which they are described, explained and ultimately understood. This is the traditional path that science has so successfully followed since the Renaissance.... In each of these cases, scientists built their theories upon a scaffold of experimental data. The Standard Model could not have been invented by theorists, however brilliant, just sitting around and thinking.
>
> Sometimes scientists have followed a different road. The high road tries to avoid the morass of mundane experimental data.

Glashow goes on portraying the invention by Einstein of classical general relativity as the single example of successful pursuit of the high road and exemplifying modern high-roaders with superstring theorists.

However, we ought to say, string theory in general is a very reasonable bet compared with most "quantum gravity" schemes. What motivates them? From a textbook [2, p. 24] we quote Bergmann:

> Today's theoretical physics is largely built on two giant conceptual structures: quantum theory and general relativity. As the former governs primarily the atomic and subatomic worlds, whereas the latter's principal applications so far have been in astronomy and cosmology, our failure to harmonize quanta and gravitation has not yet stifled progress in either front. Nevertheless, the possibility that there might be some deep dissonance has caused physicists an esthetic unease, and it has caused a number of people to explore avenues that might lead to a quantum theory of gravitation, no matter how many decades away the observations.

Dissonance, we claim, there is not: trees electromagnetically keep growing on the third planet from the Sun, bound by gravity since as far as we can tell. There is theoretical ignorance about a vast region of possible experience unconstrained by evidence. Be that as it may, "aesthetic unease" is about the worst guide for science. Ugliness is in the eye of the beholder. Nobody claims the standard model of particle physics to be beautiful. However, it has survived more than 35 years of determined theoretical and – much more important – empirical assault. It possesses now the beauty of staying power: any scheme whatsoever aiming to replace it needs to manage the Standard Model disguise.

History is a better guide. The clash between classical mechanics and electromagnetism, seemingly leading to catastrophic atomic collapse, was overcome by more profound experiments and the quantum theories designed to explain them. Therefore, we do little of the "dissonance" of the underpinnings of quantum theory and classical gravity, since in all likelihood at least one of those is doomed to perish.

Glashow concludes

> History is on our side (i.e., of the low-roaders). Every few years there has been a world-shaking new discovery in fundamental physics or cosmology... Can anyone really believe that nature's bag of tricks has run out? Have we finally reached the point where there is no longer... a bewildering new phenomenon to observe? Of course not.

Fortunately, even classical gravity is in deep crisis. This opens a number of opportunities. The crisis concerns almost every aspect.

- Cosmic acceleration. In a nutshell, the expansion of the universe seems to be *accelerating* when it should be *braking*. This is the “cosmological constant” or “dark energy” problem. The question is obviously why now? We shall come back to this.
- Galaxy clustering and cosmology. As it turns out, some think the previous to be a pseudo-problem. Wiltshire and coworkers [3–6] have argued that

  > Cosmic acceleration can be understood as an apparent effect, and dark energy as a misidentification of those aspects of cosmological gravitational energy that by virtue of the strong equivalence principle cannot be localized.

  Wiltshire’s proposal is of the “radically conservative” kind. The implication is that we truly do not know how to solve the Einstein equations.

  In a similar vein, current orthodoxy regarding gravitational collapse towards black holes and the “information loss” problem has been also called into question [7].
- The best-tested aspects of the theory are challenged by the solar system anomalies. To begin with, at least since the 1980s it has been known that the trajectories of the *Pioneer* 10 and *Pioneer* 11 past the outer planets’ orbits deviate from the predictions, as though some extra force is tugging at them from the direction of the Sun [8–10].
  The unmodelled blue shift appearing in the *Pioneer* missions data amounts to 10^{-9} cm/s^2; it may not seem much, but it adds now to many thousands of kilometres behind the projected paths. A “covariant” solution to the anomaly seems ruled out – see, for instance, [11]. In desperation, some bold proposals are being made. For instance that, because of the influence of background gravitational sources in the universe on the evolving quantum vacuum [12, 13], astronomical time and time as nowadays measured by atomic clocks might not coincide.
- To this, add the even more surprising and now apparently verified fact (spoken about in hushed ones since 1990, when first noticed in the flight of probe *Galileo* by Earth) that the slingshot manoeuvre of spacecraft delivers (or takes away) more energy than the current theory allows us to expect [14]. A simple empirical formula describes rather accurately the deviations, which translate into a few millimetres a second of extra velocity.
  Both solar system anomalies belong in the category of “unexpected experiments”.
- The existence of (non-baryonic) *dark matter* is better established than that of dark energy, since several lines of evidence point to a relatively low baryon content of the universe.
  However, models do exist that attribute the relatively high acceleration of stars in a typical galaxy, thus the appearance of dark matter, to mysterious deviations from standard gravity. Particularly, Milgrom’s MOND (modified Newton dynamics) model – see [15] and references therein, as well as the discussion in the popularization book [16]. MOND postulates that Newton’s law is modified in very weak acceleration regimes. There is no “respectable” theory behind it as yet. However, as it happens, Milgrom’s hypothesis implies predictions on the surface densities of galaxies and more; these have been pretty much verified till

now. The Milgrom acceleration is pretty close to the cosmic acceleration. It is not very different in order of magnitude from the "acceleration" of the *Pioneers*.
On the other hand, interaction with dark matter might explain the Pioneers' blue shift.

- Taken together, dark matter and energy signal the transition to a new cosmological paradigm. Whether they will emerge as modified gravity (massive graviton or other), new energy components, or pointers to strings and other noncommutative substructures remains to be seen.
- Among the questions of principle that periodically erupt into controversy is the question of the speed of transmission of the gravitational interaction or, if you wish, the lack of aberration of gravity [17].

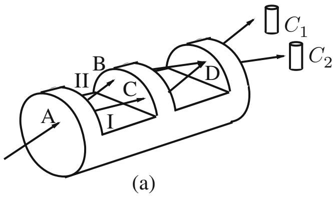
1.2 Gravity and Experiment: Expect the Unexpected

Perhaps the most fundamental question of principle, for our purposes, concerns the role, if any, of the principle of equivalence in the interface of gravity with the quantum world. We begin by that in earnest. Now, there is little in the way of quantum gravity that we can probe in laboratory benches at present. The universe was created with a quarantine: gravity is so weak an interaction that it can only produce measurable effects in the presence of big masses, and this very fact militates against detecting radiative corrections to it. To see quantum effects in pure gravity is far beyond our power. What we can do with some confidence is to envisage quantum systems in classical background gravitational fields, with back-reaction neglected, or approximately treated. In fact, only the interface of nonrelativistic quantum mechanics with Newtonian gravity has been experimentally tested.

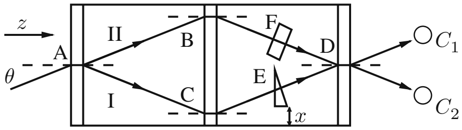
Some wisdom is gained, however, by not discarding a priori such humble beginnings. For this writer, the alpha of quantum gravity is the Colella–Overhauser–Werner (COW for short, from now on) experiment [18]. It tests the equivalence principle. The latter appears in textbooks in slightly different formulations. For some, the "strong" principle says that accelerative and gravitational effects are locally equivalent; the "weak" principle states that inertial masses and gravitational charges are the same (up to a universal constant). Some others use the nomenclature the other way around. In both cases we refer to systems placed in external fields, such that the complicating effects of the gravitational pull by the system itself can be neglected. From the second form it plainly follows that all *classical* masses fall with the same acceleration in a gravity field. Thus, if the initial conditions for those masses coincide, their trajectories will coincide as well: Galileo's uniqueness of free fall. In other words, mass is superfluous to describe particle motions in classical gravity; it all belongs to the realm of kinematics. From this to the assertion [19, p. 334] that

> geometry and gravitation were one and the same thing.

is there but a near-vanishing step.



(a)



(b)

Fig. 1.1 (**a**) In the most common interferometer three "ears" are cut from a perfect crystal, ensuring coherence over it (about 10 cm long). The incident beam is split (by Bragg scattering) at A into two, I and II. These are redirected at B and C and recombine in the last ear. The relative phase at D determines the counting rate at the detectors. (**b**) *Top view* of the interferometer. The relative phase can be changed in a known way by inserting a wedge in one beam at E, the thickness of which can be changed by displacement. The experiment is performed at F. The figures are reprinted with permission from [20]. Copyright 1983 by the American Physical Society. Redrawn by Heine Larsen, Roskilde, (see http://publish.aps.org/copyrightFAQ.html)

So, what does the COW experiment mean for humanity? It and its follow-ups lend support to the equivalence principle. It would have been earth-shaking if they did not, but it is indispensable to reflect on which aspects of current orthodoxy are confirmed, and which ones actually disproved by it.

The COW tool is neutron (and neutral atom) interferometry. A typical neutron interferometer – Fig. 1.1, taken from [20] – is a silicon crystal of length L. The incident beam is split with half-angle θ in the first ear of the apparatus at one extreme, redirected halfway through it, and recombines in the third ear at the other extreme. The neutron wavelength λ_N and the atom spacing in the crystal need to be of the same order, about 10^{-8} cm. Thus the momentum is in the ballpark of $(\hbar/\lambda_N) \sim 10^{-20}$ erg. The neutron is relatively cold: with an inertial mass $m_i \sim 10^{-24}$ g, this implies a velocity $v \sim 10^4$ cm/s; thus a nonrelativistic calculation will do.

A gravitational phase shift is obtained simply by rotating the apparatus about the incident beam, say an angle α, so the acceleration is $g \sin\alpha$, with g the standard acceleration on Earth. The phase shift over one period is of the order of the quotient between the (difference in) potential energy and the kinetic energy of the beam; even with the small velocities involved, this is of the order $\sim 10^{-7}$. Under such conditions, it is not hard to see that the phase difference is given approximately by

$$\frac{\int V\,dt}{\hbar},$$

where V denotes the difference in potential between the higher and the lower unperturbed neutron paths and t is the time.

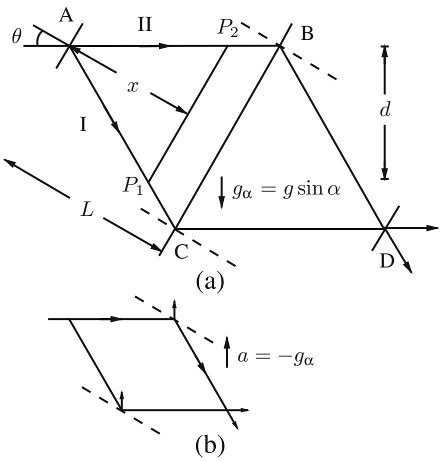


Fig. 1.2 Gravitational perturbation of the beam. (**a**) The interferometer is rotated around the incident beam by an angle α; the beams will be at a different height (equal to $2x \sin\theta$ between equivalent points along the paths), with an effective gravitational field $g_\alpha = g \sin\alpha$ in the interferometer plane. (**b**) In the free-fall system, the neutron beams are unaccelerated, but the interferometer scattering planes appear to be accelerating upwards. The figures are reprinted with permission from [20]. Copyright 1983 by the American Physical Society. Redrawn by Heine Larsen, Roskilde, (see http://publish.aps.org/copyrightFAQ.html)

Now, let x be a rectilinear coordinate along the long diagonal of the rhomb constituted by the two beam's paths. Then the difference of height between the paths is as indicated in Fig. 1.2. The difference in potential is $2mg \sin\alpha x \sin\theta$. Thus we have

$$\frac{\int V\,dt}{\hbar} = \frac{4mg \sin\alpha \sin\theta}{\hbar v \cos\theta} \int_0^L x\,dx = \frac{mgA \sin\alpha}{\hbar v}, \tag{1}$$

with v the mean velocity of the neutrons and A the area of the rhomb, given by half the diagonals' product:

$$A = 2L^2 \tan\theta.$$

Actually the mass appearing in (1) is the gravitational charge; the inertial mass m_i is hidden in the relation between v and the de Broglie wavelength. The shift (1) is around 100 rad, and the resulting fringe pattern easily visible and measurable. (We have neglected the effect of the Earth's rotation, which amounts to less of 2% of the total shift.) It turned out that the neutrons do fall in the Earth's gravity field as predicted by the Schrödinger equation, with m and m_i identified.

The experiment appears to confirm both versions of the equivalence principle, since the possibility of describing the problem in the neutron beam reference system as an upward acceleration of the interferometer holds in the Schrödinger equation. This is discussed exhaustively in [21]. Use of the Dirac equation instead makes no practical difference. Anyway, the experiment was repeated in “actually accelerated” interferometers, with the expected result [22].

However, as soon as we try to translate the "weak" principle in *geometrical* terms in the quantum context, we run into trouble. The fact that "trajectories" have not much quantum-mechanical meaning is enough to make us suspicious. Nevertheless, let us for simplicity explore the situation in terms of circular Bohr orbits. (That these are still pertinent concepts is plain to anybody who has done atomic physics with the Wigner phase-space function [23, 24].) Assume a very large mass M bounds a small one m gravitationally into a Bohr atom. For circular orbits with angular velocity ω, Kepler's laws give

$$\omega^2 = \frac{GM}{r^3}, \quad \text{with } r \text{ restricted by} \quad mr^2\omega = n\hbar.$$

Thus

$$E_n = -\tfrac{1}{2}m\omega^2 r^2 = -\frac{G^2M^2m^3}{2\hbar^2 n^2}.$$

Therefore in quantum mechanics one can *tell the mass* of a gravitational bound particle. The explanation for this lies in the very quantization rule

$$[x, p] = i\hbar,$$

which is formulated in phase space. If we define velocity by p/m, we obtain the commutator

$$[x, v] = i\hbar/m.$$

This means that kinematical quantities are functions of $\hbar/m$. In general, it is enough to look at the Schrödinger equation to see that energy eigenvalues go like $mf(\hbar/m)$, or more accurately, $mf(\hbar^2/mm_i)$ for some function f.

Now, if we admit the previous, how does the dependence of the mass disappear in the classical limit? The only possibility is that the quantum number scales with m. This of course makes sense in the semiclassical limit: if particle 1 is heavier than particle 2, we expect its energy levels to be accordingly higher. But for low-lying states geometrical equivalence inevitably breaks down. We have here the curious case of a symmetry generated (rather than broken) by "dequantization". The point was made in [20].

In summary, lofty gravity is treated by quantum mechanics as lightly as lowly electrodynamics. In the classical motion of charged particles, only the parameter e/m appears. This is not interpreted geometrically, since e/m varies from system to system, so nobody thinks it has fundamental significance. When the system is quantized, $\hbar$ comes along in both cases, and in gravity experiments, like the ones described above with states in the continuum, we can tell the mass. Alas, for some this destroys the beauty of the theory. So much that they never mention the fact.

1.2.1 Noncommutative Geometry I

Before examining the consequences of the failure of the geometrical principle, let us see if we can find a way out. To preserve weak equivalence as an exact quantum symmetry, we must take the canonical velocity as a dynamical quantity $\mathfrak{v}$. Then the Hamiltonian is rewritten as follows:

$$H = m(\mathfrak{v}^2/2 + V(x)) = m\mathcal{H}(x, \mathfrak{v}),$$

with V the gravitational potential. If now we quantize the theory in terms of x and $\mathfrak{v}$, we obtain a "quantum gravity" theory respecting the geometrical equivalence principle (although, of course, this flies in the face of the workings of ordinary quantization for other interactions).

Through existence of the constant c of nature, such a quantization method involves the introduction of a fundamental length

$$[x, \mathfrak{v}] = icl_0.$$

This is not quite "noncommutative geometry" in the superficial way it is mostly practised nowadays (the present author is not innocent of such a sin), but resembles it more than a bit. The point we are able to make is twofold: (i) of need the geometrical approach to quantum gravity will be *noncommutative* or will not be; (ii) it is not at all required that l_0 be of the order of Planck's length scale. It has been argued many times, invoking mini-black holes in relation with the incertitude principle and such, that something must happen at that length scale – see [25], for example. But nothing forbids that the critical length be bigger (a string length, for instance), provided it could have escaped detection so far. If and how such fundamental length intervenes is a matter only for experiment to decide.

We return to noncommutative geometry in Sect. 1.5.

1.2.2 Whereto Diffeomorphism Invariance?

The understanding that geometry and gravitation are not to be one and the same thing should be confirmed by some experiment checking (low-lying) states of a quantum system bound by gravity.

Such an experiment – the first ever to observe gravitational quanta – has already taken place [26].

Ultracold neutrons ($v \sim 10$ m/s) are stored in a horizontal vacuum chamber; a mirror is placed below and a non-specular scatterer above. Thus the neutrons find themselves in a sort of gravitational potential well, with a "soft wall" on one side. The Bohr – Sommerfeld formula is good enough to calculate its energy levels associated with vertical motion:

$$E_n = (9m_{\mathrm{N}}/8)^{1/3}\left(\pi\hbar g\left[n-\tfrac{1}{4}\right]\right)^{2/3}.$$

We obtain

$$E_1 \simeq 1.4\ \mathrm{peV} \simeq 10^{-13}\ \mathrm{Ry}. \tag{2}$$

A first remarkable thing is the minuteness of (2). In spite of being so small, quantum effects of gravity have been detected on a table top! However, the main question here is that the difference between masses becomes of a yes/no nature. Suppose that the height of the "slit" formed by the upper and lower walls of the chamber is smaller than 10^{-3} cm. If instead of neutrons one were trying to send through (say) aluminium atoms, they would be observed at the exit. However, that same slit *on Earth* is opaque to neutrons. The following rule of thumb is useful: the energy required to lift a neutron by 10^{-3} cm is classically 1 peV with a good approximation. Accordingly the width of the state (2) can be estimated: the height of the chamber should be bigger than 1.4×10^{-3} cm for neutrons to be observed at the exit. Figure 1.3 illustrates this. The phenomenon has nothing to do with diffraction, since the wavelength of neutrons remains much smaller than the height of the slit; visible light, with a wavelength much bigger than those neutrons, is transmitted.

Bingo! A slit has become a wall, impenetrable. Uniqueness of free fall fails. Gravitation is not just geometry.

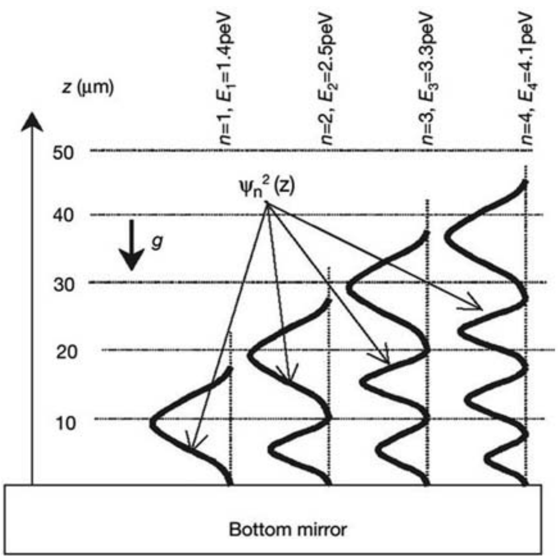


Fig. 1.3 Quantum states are formed in the "potential well" between the Earth's gravity field and the horizontal mirror on bottom. The *vertical axis z* is intended to give an idea about the spatial scale for the phenomenon. The figure is reprinted with permission from V.V. Nesvizhevsky et al., Phys. Rev. D Vol. 67, 102002 (2003). Copyright 2003 by the American Physical Society, (see http://publish.aps.org/copyrightFAQ.html)

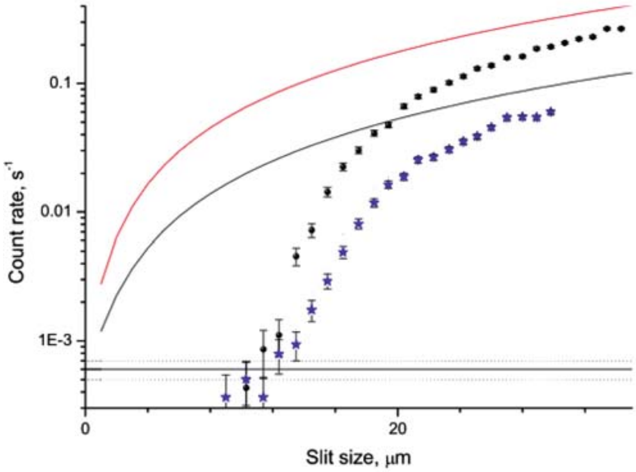


Fig. 1.4 Dependency of the particle flux on the slit size. The *circles* indicate the experimental results [26] for a beam with an average value of 6.5 m/s for the horizontal velocity component. The *stars* show the analogous measurement with 4.9 m/s. The *solid lines* correspond to the classical expectation values for these two experiments. The *horizontal lines* indicate the incertitude in the detector background. The figure is reprinted with permission from [26]

The point is even more forcefully brought home in Fig. 1.4, which describes the actual experimental situation. Put in a different way, at least for interaction with matter, the (geometrical form of the) equivalence principle and the incertitude principle clash. No prizes to guess which must give way.

Surprisingly, our viewpoint is found controversial by some. To put matters into perspective, it is helpful to keep in mind that the equivalence principle is classically expressed by the statements (1) gravitational mass equals inertial mass or (2) the motion of particles in a gravity field is indifferent to their mass. While the COW experiment confirms (1), the second is untrue in the quantum world. Since point particles, paths, and clocks play an apparently essential role in the foundations of general relativity (see the remarks further below), and since it is hard to see how geometry could have come to such a preponderance in dynamics without (2), it would seem the latter is bound to diminish. However, one can argue for an important residual role of geometry in quantum physics, as in the very readable article [27].

(In the current experimental situation, there is not much more than can be done directly to measure quantum jumps in a gravitational field. Present hopes to improve on accuracy of measurement of the quantum states parameters rest on use of storage sources of ultracold neutrons and magnetic field gradients to resonate with the frequency defined by the energy difference of two states [28].)

Among the numerous works on "quantum gravity" that make much of the classical geometry aspects of gravitation, a good representative is the homonymous book [29]. Its philosophical position is staked out at the outset:

> the question we have to ask is: what we have learned about the world from quantum mechanics and from general relativity?... What we need is a conceptual scheme in which the insights obtained with general relativity and quantum mechanics fit together.
>
> This view is not the majority view in theoretical physics, at present. There is consensus that quantum mechanics has been a conceptual revolution, but many do not view general relativity in the same way... According to this opinion, general relativity should not be taken too seriously as a guidance for theoretical developments.
>
> I think that this opinion derives from a confusion: the confusion between the specific form of the Einstein – Hilbert (EH) action and the modification of the notions of space and time engendered by general relativity.

We are pleased to vote with the bread-and-butter majority here. The trouble is the non-geometrical cast of quantum *dynamics*. Since we know not the shape of things to come, the task is not so much to "fit general relativity with quantum mechanics together" as to – slowly and painstakingly – extend our knowledge to quantum and gravitational phenomena simultaneously taking place. It is somewhat saddening that the COW experiment and its successors are not found in the reference list of [29]; nor are they mentioned in the history of quantum gravity given as an appendix in that book – which is more in the "history of ideas" mould. In fact the sphere of ideas around the proper interpretation of the COW experiment hails back to Wigner, who, long ago, had explained keenly the quantum limitations of the *concepts* of general relativity [30], concluding

> the essentially non-microscopic nature of the general relativistic concepts seems to us inescapable.

In otherwise mathematically subtle and full of gems [29], as in the works of other practitioners of quantum gravity, the warning goes unmentioned, as well as unheeded.

To summarize, a generous dose of salt is in order when dealing with "quantum gravity" claims. Without necessarily enjoying the quarantine, we should go most carefully about breaking it. Not only "large fragments of the physics community" but also thoughtful mathematicians like Yuri Manin advise a useful scepticism, in the respect of taking as physical what is just product of mathematical skill:

> Well-founded applied mathematics generates prestige which is inappropriately generalized to support quite different applications. The clarity and precision of mathematical derivations here are in sharp contrast to the uncertainty of the underlying relations assumed. In fact, similarity of the mathematical formalism involved tends to mask the differences in the differences in the scientific extra-mathematical status...mathematization cannot introduce rationality in a system where it is absent...or compensate for a deficit of knowledge.

This is very timely quoted in [31].

1.3 Gravity from Gauge Invariance in Field Theory

From our standpoint, the action for gravitational interactions is more important than speculative "background independency" in a "final unified theory". Moreover, the pure gravity EH action can be rigorously derived from the theory of quantum fields:

a simple lesson, often forgotten. We proceed to that in this section. (As a historical note, for once the Einstein – Hilbert surname is right on the mark: independently Hilbert and Einstein gave the new equations of gravitation in the dying days of November 1915.)

1.3.1 Preliminary Remarks

The book [32], containing lectures by Feynman on gravitation given at Caltech in 1962–1963, deals with the *perturbative* approach to classical gravity, to wit, with the self-consistent theory of a massless spin-2 field (we may call it graviton). The foreword of this book (by John Preskill and Kip S. Thorne) is recommended reading. There the unfolding of (earlier) variants of the same idea by Kraichnan and Gupta is narrated as well, with references to the original literature. The main aspect in Kraichnan – Gupta – Feynman arguments is that a geometrical theory is obtained from flat-spacetime physics by using consistency requirements. Later work by Deser and Ogivetsky and Polubarinov in the same spirit is also remarkable.

The distinctively non-geometrical flavour is welcome here, where we regard the geometrical approach as suspect. An excellent review with references of the classical path from the action for such field to the EH action is found in the recent book [33, Chap. 3].

Weinberg's viewpoint in 1964 [34] is also very instructive and deserves mention. On the basis of properties of the $\mathbb{S}$-matrix, he proves that gravitons must couple to all forms of energy in the same way. He moreover shows that any particle with inertial mass m_{i} and energy E has, apart from Newton's constant, an effective gravitational charge

$$2E - m_{\mathrm{i}}^2/E.$$

For $E = m_{\mathrm{i}}$, one recovers the usual equivalence result, while for $m_{\mathrm{i}} = 0$ one obtains $2E$, which gives the correct result for the deflection of light. (Also, a graviton must respond to an external gravity field with the same charge.)

In this section we perform a parallel exercise to Feynman's: assuming ignorance of Einstein's general relativity, we arrive again at the EH action by successive approximation. Our method has little to do with the "effective Lagrangians" approach and differs from traditional ones mentioned above in at least one of several respects:

- We consider only pure gravity. Coupling to matter is sketched after the fact, just for completeness.
- It is fully quantum field theoretical, in that recruits the *canonical formalism* on Fock space and quantum gauge invariance. Our main tool is BRS technology, and ghost fields are introduced from the outset. In other words, we treat gravity as any other gauge theory in the quantum regime; we obtain a quantum theory of the gravitational field, in which at some point we put $\hbar = 0$.

- We use the causal (or Epstein – Glaser) renormalization scheme [35], relying on the (perturbative expansion in the coupling parameter of the) $\mathbb{S}$-matrix. This entails a slight change of interpretation, in regard to renormalization, with respect to standard thinking; we briefly discuss the matter at the end of Sect. 1.3.6. Epstein – Glaser renormalization is specially appropriate for gravity issues since it does not rely on translation invariance.
- We never invoke the stress – energy tensor.

In some sense we close a circle opened as well by Feynman in the early 1960s [36], where he first realized that unitarity at (one-)loop graph calculations demanded ghost fields, for gravity as well as for Yang – Mills theory. Through well-known work by DeWitt, Slavnov, Taylor, Fadeev and Popov, and Lee and Zinn-Justin, this would eventually lead to BRS symmetry by the mid-1970s.

We mainly follow [57, 38]. The remote precedent for the last paper is an outstanding old article by Kugo and Ojima [39].

1.3.2 Exempli Gratiae

In order to make clear the strategy, we briefly recall here the similar treatment for (massive and massless) electrodynamics. Suppose we wish to effect the quantization of spin-1 particles by means of real vector fields. The question is how to eliminate the unphysical degrees of freedom, since a vector field has four independent components, while a spin-1 particle has three helicity states, or two if it is massless.

A standard procedure is to impose the constraint $\partial^\mu A_\mu =: (\partial \cdot A) = 0$. However, this is known to lead to the Proca Lagrangian (density), which has very bad properties. Also, under quantization, use of Proca fields entails giving up covariant commutators of the disarmingly simple form found for neutral scalar fields:

$$[A^\mu(x), A^\nu(y)] = i\eta^{\mu\nu} D(x-y), \qquad (A^\mu)^+ = A^\mu, \tag{3}$$

with η the Minkowski metric and D the Jordan – Pauli propagator. We would like to keep them instead. The Klein – Gordon equations

$$(\Box + m^2)A^\mu = 0 \tag{4}$$

we would like to keep as well. Now, it is certainly impossible to realize (3) and (4) on Hilbert space if by $+$ we understand the ordinary involution. However, it is possible to do it through the introduction of a distinguished symmetry η (that is, an operator both self-adjoint and unitary) called the Krein operator. Whenever such a Krein operator is considered, the η-conjugate O^+ of an operator O with adjoint $O^\dagger$ is

$$O^+ := \eta O^\dagger \eta.$$

Let $(\cdot, \cdot)$ denote the positive definite scalar product in H. Then

$$\langle \cdot, \cdot \rangle := (\cdot, \eta \cdot)$$

yields an “indefinite scalar product”, and the definition of O^+ is just that of the adjoint with respect to $\langle \cdot, \cdot \rangle$. Then A will be self-conjugate.

The massive vector field model is known to be a gauge theory [40] if we introduce the auxiliary (scalar) Stückelberg field B (say with the same mass m) and gauge transformations of the form

$$\begin{aligned} \delta A^\mu(x) &= \eta^{\mu\nu} \partial_\nu \theta(x) = \partial^\mu \theta(x), \\ \delta B(x) &= m\theta(x). \end{aligned}$$

The trick now is to use the unphysical parts $\partial \cdot A$, B plus the ghosts u and anti-ghost $\tilde{u}$ to construct the BRS operator

$$Q = \int_{x^0=\text{const}} d^3x \, (\partial \cdot A + mB) \overleftrightarrow{\partial_0} \, u,$$

whose action should reproduce the gauge variations (where commutators $[.,.]_-$ or anticommutators $[.,.]_+$ are taken according to whether the ghost number of the varied field is even or odd):

$$\begin{aligned} sA^\mu(x) &= [Q, A^\mu(x)]_\pm = i\partial^\mu u(x), \\ sB(x) &= [Q, B(x)]_\pm = imu(x), \\ su(x) &= [Q, u(x)]_\pm = 0, \\ s\tilde{u}(x) &= [Q, \tilde{u}(x)]_\pm = -i\big(\partial \cdot A(x) + mB(x)\big). \end{aligned} \tag{5}$$

With these relations one easily proves 2-nilpotency modulo the field equation:

$$2Q^2 = i \int_{x^0=\text{const}} d^3x \, \Box u \, \overleftrightarrow{\partial_0} \, u + im^2 \int_{x^0=\text{const}} d^3x \, u \, \overleftrightarrow{\partial_0} \, u = 0.$$

Thus the right-hand side of (5) is coboundary fields. With the help of nilpotency, the finite gauge variations for the same fields of (5) are easily computed. The supercharge Q is conserved. The massless limit is not singular in this formalism: for photons, we just put $m = 0$, and B drops out of the picture.

1.3.3 The Free Lagrangian

A rank 2 tensor field under the Lorentz group decomposes into the direct sum of four irreducible representations, corresponding to traceless symmetric tensors, a scalar field, and self-dual and anti-self-dual tensors. We group the first two into a symmetric tensor field $h \equiv \{h^{\mu\nu}\}$ with arbitrary trace. Let us introduce as well

$$\varphi := h^\rho_\rho, \quad H \equiv \{H^{\mu\nu}\} := \{h^{\mu\nu} - \tfrac{1}{4}\eta^{\mu\nu}\varphi\}, \quad \text{thus} \quad H^\rho_\rho = 0.$$

(We wish to keep h to denote the whole tensor, and so we do not use the standard notation for its trace.) Again the question is how to eliminate the superfluous degrees of freedom in the description of a spin-2 relativistic particle, which possesses only two helicity states. A fortiori we do not want to follow for the graviton the path of enforcing constraints that was discarded for photons.

For a free graviton one may settle on the Lagrangian

$$\mathcal{L}^{(0)} = \tfrac{1}{2}(\partial_\rho h^{\alpha\beta})(\partial^\rho h_{\alpha\beta}) - (\partial_\rho h^{\alpha\beta})\left(\partial_\beta h^\rho_\alpha\right) - \tfrac{1}{4}(\partial_\rho\varphi)(\partial^\rho\varphi). \tag{6}$$

Of course this choice is not unique. The more general Lorentz-invariant action quadratic in the derivatives of h is of the form

$$\int d^4x \,[a(\partial_\rho h^{\alpha\beta})(\partial^\rho h_{\alpha\beta}) + b(\partial_\rho h^{\alpha\beta})\left(\partial_\beta h^\rho_\alpha\right) + c(\partial_\rho\varphi)(\partial^\sigma h_{\rho\sigma}) + d(\partial_\rho\varphi)(\partial^\rho\varphi)].$$

The frequently invoked Fierz – Pauli Lagrangian [41] is of this type, with $a = \frac{1}{4}, b = -\frac{1}{2}, c = \frac{1}{2}, d = -\frac{1}{4}$. The signs are conventionally chosen in both cases so that the first term has a positive coefficient. The Euler – Lagrange equations corresponding to (6)

$$\partial_\gamma \frac{\partial\mathcal{L}^{(0)}}{\partial(\partial_\gamma h_{\alpha\beta})} = 0$$

yield at once

$$\Box h^{\alpha\beta} - \partial_\gamma\partial^\beta h^{\alpha\gamma} - \partial_\gamma\partial^\alpha h^{\beta\gamma} - \tfrac{1}{2}\eta^{\alpha\beta}\Box\varphi = 0. \tag{7}$$

This form is essentially equivalent to the Fierz – Pauli equation, but more convenient here. (For a critique of the Fierz – Pauli framework, consult [42].)

1.3.4 A Canonical Setting

A crucial point is the invariance of the Lagrangian $\mathcal{L}^{(0)}$ – thus of (7) – under gauge transformations

$$\delta h^{\alpha\beta} = \lambda(\partial^\alpha f^\beta + \partial^\beta f^\alpha - \eta^{\alpha\beta}(\partial\cdot f)) = \lambda b^{\alpha\beta\rho}_\tau \partial_\rho f^\tau, \tag{8}$$

where

$$b^{\alpha\beta\rho}_\tau := \eta^{\alpha\rho}\delta^\beta_\tau + \eta^{\beta\rho}\delta^\alpha_\tau - \eta^{\alpha\beta}\delta^\rho_\tau,$$

for arbitrary $f = (f^\alpha)$. This entails

$$\delta\varphi = -2\lambda(\partial \cdot f). \tag{9}$$

To verify this invariance, with an obvious notation, and up to total derivatives,

$$\begin{aligned}
\delta\mathcal{L}_I^{(0)} &= -\delta h_{\alpha\beta}\Box h^{\alpha\beta}, \\
\delta\mathcal{L}_{II}^{(0)} &= \delta h_{\alpha\beta}\partial^\rho \left(\partial^\alpha h_\rho^\beta + \partial^\beta h_\rho^\alpha\right), \\
\delta\mathcal{L}_{III}^{(0)} &= \tfrac{1}{2}\delta\varphi\Box\varphi.
\end{aligned}$$

One finishes the argument by use of (8) and (9).

That tensor b will reappear often. Classically, one could specify here the *transverse gauge* condition:

$$\partial_\beta(h^{\alpha\beta} + \delta h^{\alpha\beta}) = 0. \tag{10}$$

(In the gravity literature a so-called de Donder gauge condition is more frequently used.) The last equation is obtained at once if f^α solves

$$\lambda\Box f^\alpha = -\partial_\beta h^{\alpha\beta} =: -\,(\partial \cdot h)^\alpha,$$
$$\text{then (7) reduces to} \quad \Box h = 0.$$

As advertised, we refrain from quotient by imposing gauge conditions. In our BRS-like treatment, the elimination of the many extra degrees of freedom takes place cohomologically, rather than by use of constraints. The fields are promoted to (by now still free) normally ordered quantum fields. Clearly, in this approach we need to add to $\mathcal{L}^{(0)}$ the gauge-fixing and free ghost terms:

$$\mathcal{L}_{\text{free}} = \mathcal{L}^{(0)} + \tfrac{1}{2}(\partial\cdot h)\cdot(\partial\cdot h) - \tfrac{1}{2}(\partial_\mu\tilde{u}_\nu + \partial_\nu\tilde{u}_\mu)(\partial^\mu u^\nu + \partial^\nu u^\mu - \eta^{\mu\nu}(\partial\cdot u)). \tag{11}$$

One quantizes h in the most natural way:

$$[h^{\alpha\beta}(x), h^{\mu\nu}(y)] = i b^{\alpha\beta\mu\nu}\, D(x-y), \tag{12}$$

and therefore the propagators for H, φ are given by

$$\begin{aligned}
[H^{\alpha\beta}(x), H^{\mu\nu}(y)] &= i\left(\eta^{\alpha\mu}\eta^{\beta\nu} + \eta^{\alpha\nu}\eta^{\beta\mu} - \tfrac{1}{2}\eta^{\alpha\beta}\eta^{\mu\nu}\right) D(x-y), \\
[\varphi(x), \varphi(y)] &= -8i\, D(x-y), \\
[\varphi(x), H^{\mu\nu}(y)] &= 0.
\end{aligned}$$

Also, for the fermionic ghosts we have the anticommutation relations

$$[u^\alpha(x), u^\beta(y)] = i g^{\alpha\beta} D(x-y). \tag{13}$$

All other anticommutators vanish. The new Euler – Lagrange equations give rise now to the simplest possible, ordinary wave equations for all fields considered.

$$\Box h = 0, \quad \Box u = 0, \quad \Box \tilde{u} = 0.$$

We can prove directly consistency of rules (12) and (13), analogous to (3) and (4), by constructing a explicit representation in a Fock – Krein space. The reader will see this in a later section.

Let us now introduce the BRS operator

$$Q = \int_{x^0=\mathrm{const}} d^3x \, (\partial \cdot h)^\alpha \overleftrightarrow{\partial_0} \, u_\alpha = \int_{x^0=\mathrm{const}} d^3x \, \big((\partial \cdot H)^\alpha + \tfrac{1}{4}\partial^\alpha \varphi\big) \overleftrightarrow{\partial_0} \, u_\alpha; \quad (14)$$

where $(\partial \cdot h)^\alpha$ denotes the divergence $\partial_\beta h^{\alpha\beta}$, which in view of (10) is unphysical, and u_α is the fermionic (vector) ghost field. The associated gauge variations are as follows:

$$\begin{aligned} sh^{\mu\nu} &= [Q, h^{\mu\nu}] = i b^{\mu\nu\rho}_{\tau} \partial_\rho u^\tau = i(\partial^\mu u^\nu + \partial^\nu u^\mu - \eta^{\mu\nu}(\partial \cdot u)), \\ su &= [Q, u]_+ = 0, \\ s\tilde{u} &= [Q, \tilde{u}]_+ = -i(\partial \cdot h)^\mu. \end{aligned} \quad (15)$$

Note that the action of the coboundary operator is dictated by the variation (8). Other important coboundaries like

$$s\varphi = i(\partial \cdot u); \qquad s(\partial \cdot h)^\mu = 0$$

follow from (15) on-shell. Again the supercharge Q is 2-nilpotent and conserved.

1.3.5 What to Expect

We make a temporary halt to examine whether with our choices in Sect. 1.3.3 we are on the right track, after all. Let $g := (g_{\alpha\beta})$ denote the metric tensor and R the Ricci curvature. As hinted above, for this writer the EH action (with c=1 and without the “cosmological constant”)

$$S_{\mathrm{EH}} = -\frac{1}{16\pi G} \int d^4x \, \sqrt{-\det g} \, R = -\frac{1}{16\pi G} \int d^4x \, \mathfrak{g}^{\mu\nu} R_{\mu\nu}$$

constitutes the alpha and omega of gravitation theory. Here G is Newton’s constant, equal to $\hbar/m^2_{\mathrm{Planck}}$. We recall

$$\begin{aligned}
\Gamma^{\alpha}_{\beta\gamma} &= \tfrac{1}{2} g^{\alpha\mu}(\partial_\gamma g_{\beta\mu} + \partial_\beta g_{\gamma\mu} - \partial_\mu g_{\beta\gamma}), \quad \text{thus} \\
\partial_\alpha g^{\mu\nu} &= -\Gamma^{\mu}_{\gamma\alpha} g^{\gamma\nu} - \Gamma^{\nu}_{\gamma\alpha} g^{\gamma\mu} \quad \text{(vanishing covariant derivative)}, \\
R_{\mu\nu} &= \partial_\alpha \Gamma^{\alpha}_{\mu\nu} - \partial_\nu \Gamma^{\alpha}_{\mu\alpha} + \Gamma^{\beta}_{\mu\nu}\Gamma^{\alpha}_{\beta\alpha} - \Gamma^{\beta}_{\mu\alpha}\Gamma^{\alpha}_{\beta\nu}, \\
R &= g^{\alpha\beta} R_{\alpha\beta}.
\end{aligned} \tag{16}$$

It is convenient to have a special notation for

$$\Gamma_\mu := \Gamma^{\alpha}_{\mu\alpha} = \tfrac{1}{2} g^{\alpha\gamma}\partial_\mu g_{\alpha\gamma} = \frac{\partial_\mu(\det g)}{2\det g} = \partial_\mu\Big(\log\sqrt{-\det g}\Big).$$

We have employed that the minors of $g_{\alpha\beta}$ in $\det g$ are equal to $\det g\, g^{\alpha\beta}$. Finally, the Goldberg tensor 1-density

$$\mathfrak{g}^{\alpha\beta} := \sqrt{-\det g}\, g^{\alpha\beta}$$

is – quite canonically, according to [43, Sect. 2.1] – a hero of our story.

Let us define $\lambda = 4\sqrt{2\pi G}$ (essentially the inverse of Planck's mass, in natural units). Since our approach to S_{EH} is perturbative, we need to rewrite the corresponding Lagrangian $\mathcal{L}_{\mathrm{EH}}$ as a series in the coupling constant λ. An old trick in classical gravity – see for instance [44, Sect. 93] – is to split off a divergence from $\mathcal{L}_{\mathrm{EH}}$ by using

$$\begin{aligned}
\mathfrak{g}^{\mu\nu}\partial_\alpha\Gamma^{\alpha}_{\mu\nu} &= \partial_\alpha\left(\mathfrak{g}^{\mu\nu}\Gamma^{\alpha}_{\mu\nu}\right) - \Gamma^{\alpha}_{\mu\nu}\partial_\alpha(\mathfrak{g}^{\mu\nu}), \\
\mathfrak{g}^{\mu\nu}\partial_\nu\Gamma_\mu &= \partial_\nu(\mathfrak{g}^{\mu\nu}\Gamma_\mu) - \Gamma_\mu\partial_\nu(\mathfrak{g}^{\mu\nu}).
\end{aligned}$$

With the help of previous equations, one finds

$$\mathfrak{g}^{\alpha\beta} R_{\alpha\beta} = H - \partial_\gamma\left(\mathfrak{g}^{\mu\gamma}\Gamma_\mu - \mathfrak{g}^{\mu\nu}\Gamma^{\gamma}_{\mu\nu}\right) =: H - \partial^\gamma D_\gamma, \tag{17}$$

where

$$H = \mathfrak{g}^{\alpha\beta}\left(\Gamma^{\gamma}_{\alpha\rho}\Gamma^{\rho}_{\beta\gamma} - \Gamma^{\rho}_{\alpha\beta}\Gamma_\rho\right).$$

The key step in our identification comes now: to make the contact between quantum field theory and general relativity, we postulate

$$\mathfrak{g}^{\mu\nu} = \eta^{\mu\nu} + \lambda h^{\mu\nu}. \tag{18}$$

Remark that do *not* assume h to be small in any sense. In (17) we separate the part of the vector D containing negative powers of λ:

$$D_\gamma = \frac{1}{\lambda}\left(\tfrac{1}{2}\partial_\gamma\varphi + \partial^\rho h_{\gamma\rho}\right) + D^{(0)}_\gamma. \tag{19}$$

The inverse matrix $\mathfrak{g}_{\mu\nu}$ with $\mathfrak{g}^{\mu\rho}\mathfrak{g}_{\rho\nu} = \delta^\mu_\nu$ formally becomes a series

$$\mathfrak{g}_{\mu\nu} = \eta_{\mu\nu} - \lambda h_{\mu\nu} + \lambda^2 h_{\mu\gamma} h^\gamma_\nu - \lambda^3 h_{\mu\gamma} h^\gamma_\tau h^\tau_\nu + \cdots . \tag{20}$$

Substituting this expression in the new form of the action $(2/\lambda^2) \int d^4x\, H$, we obtain a series as well:

$$\mathcal{L} = \sum_0^\infty \lambda^n \mathcal{L}^{(n)}. \tag{21}$$

(Actually, the Neumann series (20) is somewhat suspect, in view of convergence problems and other technical difficulties. One could se the Cayley – Hamilton theorem to obtain an exact expression for $(\mathfrak{g}_{\mu\nu})$.) The lowest order, at any rate, is indeed of order λ^0 in view of the two derivatives inside H, and it is seen to coincide with the free model of Sect. 1.3.3. For completeness and use later on, we also report the three-graviton and four-graviton couplings:

$$\begin{aligned}
\mathcal{L}^{(1)} &= \big(-\tfrac14 \partial_\rho \varphi \partial_\sigma \varphi + \tfrac12 \partial_\rho h^{\alpha\beta} \partial_\sigma h_{\alpha\beta} + \partial_\gamma h^\alpha_\rho \partial_\alpha h^\gamma_\sigma \big) h^{\rho\sigma}, \\
\mathcal{L}^{(2)} &= -h_{\alpha\beta} h^\rho_\beta (\partial_\nu h^{\alpha\mu})(\partial_\mu h^{\beta\nu}) - \tfrac12 h_{\rho\sigma} h^\rho_\beta (\partial_\alpha h^{\rho\beta})(\partial_\alpha \varphi) \\
&\quad - \tfrac14 h_{\nu\mu} (\partial_\alpha h^{\nu\mu}) h_{\sigma\rho} (\partial^\alpha h^{\sigma\rho}) + \tfrac12 h_{\nu\mu} (\partial_\alpha h^{\nu\mu}) h^{a\beta} (\partial_\beta \varphi) \\
&\quad + h_{\beta\rho} h^\beta_\sigma (\partial_\mu h^{\rho\alpha}) \left(\partial^\mu h^\sigma_\alpha\right) - h_{\alpha\rho} \left(\partial_\mu h^\rho_\sigma\right) (\partial_\nu h^{\alpha\rho}) h^{\mu\nu} \\
&\quad + \tfrac12 h_{\alpha\rho} h_{\beta\sigma} (\partial_\mu h^{\alpha\sigma})(\partial^\mu h^{\beta\rho}).
\end{aligned} \tag{22}$$

1.3.6 Causal Gauge Invariance by Brute Force

Interacting fields in Epstein – Glaser formalism are made out of free fields. The starting point for the analysis is the functional $\mathbb{S}$-matrix in the Dyson representation under the form of a power series:

$$\mathbb{S}(g) = 1 + T = 1 + \sum_{n=1}^\infty \frac{1}{n!} \int dx_1 \dots dx_n\, T_n(x_1, \dots, x_n) g(x_1) \cdots g(x_n). \tag{23}$$

The theory is constructed basically by using causality and Poincaré invariance of the scattering matrix to determine the form of the time-ordered products T_n. Only those fields that already are present in T_1 should appear in T_n. The adiabatic limit on the "coupling functions" $g(x) \uparrow 1$ is supposedly taken afterwards.

Causal gauge invariance (CGI) is formulated by the fact that $sT_n = [Q, T_n]_\pm$ must be a divergence, keeping in mind that T_n and T'_n are equivalent if they differ by coboundaries.

In particular, first-order CGI means

$$sT_1(x) = i(\partial \cdot T_{1/1})(x).$$

For T_1, let us try a general Ansatz containing cubic terms in the fields and leading to a renormalizable theory. At our disposal there are three field sets: $h, u, \tilde{u}$. The most general coupling with vanishing ghost number *without derivatives* is of the form

$$a\varphi^3 + b\,\varphi h_{\nu\mu}h^{\nu\mu} + c\,h_{\mu\nu}h^{\nu}_{\gamma}h^{\gamma\mu} + (u \cdot \tilde{u})\varphi + e\,h_{\nu\mu}u^{\nu}\tilde{u}^{\mu}.$$

Correspondingly, with ghost number one since the action of the BRS operator increases ghost number by one, we can have (with an obvious simplified notation)

$$T^{\mu}_{1/1} = a'u^{\mu}\varphi^2 + b'u^{\mu}h \cdot h + c'(u \cdot h)^{\mu}\varphi + d'u^{\alpha}h_{\alpha\beta}h^{\beta\mu} + e'u(u \cdot \tilde{u}).$$

Forlorn hope. It must be

$$s(\partial \cdot T_{1/1}) = 0.$$

This condition has only the trivial solution $T_{1/1} = 0$.

Since one cannot form scalars with one derivative, we are forced to consider cubic couplings with two derivatives. This is the root of "non-normalizability" (in Epstein – Glaser jargon) of gravitation. There are 12 possible combinations in T_1 involving only h with two derivatives, and 21 combinations in T_1 involving $h, u, \tilde{u}$, with two derivatives and zero total ghost number. At the end of the day, one obtains $T_1 = T_1^h + T_1^u$, with T_1^h *uniquely* proportional to $\mathcal{L}^{(1)}$ (modulo physically irrelevant divergences), and

$$T_1^u = a\big(- u^{\alpha}(\partial_{\beta}\tilde{u}_{\rho})\partial_{\alpha}h^{\beta\rho} + (\partial_{\beta}u^{\alpha}\partial_{\alpha}\tilde{u}_{\rho} - \partial_{\alpha}u^{\alpha}\partial_{\beta}\tilde{u}_{\rho} + \partial_{\rho}u^{\alpha}\partial_{\beta}\tilde{u}_{\alpha})h^{\beta\rho}\big).$$

The calculations are excruciatingly long, and of little interest. They, as well as the explicit expression of $T_{1/1}$, can be found in [37], to which we remit. By the way, had we tried to use

$$g_{\mu\nu} = \eta_{\mu\nu} + \lambda h_{\mu\nu}$$

instead of (18), then T_1^h turns out much more complicated – even after elimination of a host of divergence couplings.

More intrinsically interesting are the calculations of CGI at second order, also done in [37], which indeed reproduce $\mathcal{L}^{(2)}$. For the higher order analysis, one needs some (rather minimal) familiarity with the Epstein – Glaser method to inductively renormalize (i.e. to define) the time-ordered products T_n based on splitting of distributions. This requires use of antichronological products corresponding to the expansion of the inverse $\mathbb{S}$-matrix. If we write the inverse power series

$$\mathbb{S}^{-1}(g) = 1 + \sum_{1}^{\infty} \frac{1}{n!} \int d^4x_1 \dots \int d^4x_n \, \overline{T}_n(x_1, \dots, x_n) \, g(x_1) \dots g(x_n),$$

then we have $\overline{T}_{|N|}(N) = \sum_{k=1}^{n} (-)^k \sum_{\uplus_{j=1}^k I_j = N} T_{|I_1|}(I_1), \dots, T_{|I_k|}(I_k)$, where the disjoint union is over (non-empty) blocks I_j. For instance, the second-order term $\overline{T}_2(x_1, x_2)$ in the expansion of $\mathbb{S}^{-1}(g)$ is given by

$$\overline{T}_2(x_1, x_2) = -T_2(x_1, x_2) + T_1(x_1)T_1(x_2) + T_1(x_2)T_1(x_1).$$

The inductive step is performed using the totally advanced and totally retarded products. For instance, at the lower orders:

$$\begin{aligned} A_2(x_1, x_2) &= \overline{T}_1(x_1)T_1(x_2) + T_2(x_1, x_2) = T_2(x_1, x_2) - T_1(x_1)T_1(x_2), \\ R_2(x_1, x_2) &= T_1(x_2)\overline{T}_1(x_1) + T_2(x_1, x_2) = T_2(x_1, x_2) - T_1(x_2)T_1(x_1), \\ A_3(x_1, x_2, x_3) &= \overline{T}_1(x_1)T_2(x_2, x_3) + \overline{T}_1(x_2)T_2(x_1, x_3) + \overline{T}_2(x_1, x_2)T_1(x_3) \\ &\quad + T_3(x_1, x_2, x_3), \\ R_3(x_1, x_2, x_3) &= T_1(x_3)\overline{T}_2(x_1, x_2) + T_2(x_1, x_3)\overline{T}_1(x_2) + T_2(x_2, x_3)\overline{T}_2(x_1) \\ &\quad + T_3(x_1, x_2, x_3). \end{aligned} \tag{24}$$

By the induction hypothesis $D_{n+1} := R_{n+1} - A_{n+1}$ depends only on known quantities. Moreover D_{n+1} has causal support. If we can find a way to extract its retarded or the advanced part, that is, to split D_{n+1}, then we can calculate $T_{n+1}(x_1, \dots, x_{n+1})$.

Consider then $D_2(x, y) = [T_1(x), T_1(y)]$, the first causal distribution to be split. We have thus

$$\begin{aligned} sD_2(x, y) &= [sT_1(x), T_1(y)] + [T_1(x), sT_1(y)] \\ &= i\partial^x_\mu \left[T^\mu_{1/1}(x), T_1(y)\right] + i\partial^y_\mu \left[T_1(x), T^\mu_{1/1}(y)\right], \end{aligned} \tag{25}$$

so that D_2 *is* gauge invariant, and the issue is how to preserve gauge invariance in the renormalization or distribution splitting. That is, we must split D_2 and the commutators – without the derivatives – in the previous equation and then gauge invariance:

$$sR_2(x, y) = i\partial^x_\mu R^\mu_{2/1}(x) + i\partial^y_\mu R^\mu_{2/2}(y)$$

can only be (and is) violated for $x = y$, that is, by derivative terms in $\delta(x - y)$. That is to say, if *local* renormalization terms $N_2, N^\mu_{2/1}, N^\mu_{2/2}$ can be found in such a way that

$$s(R_2(x,y)+N_2(x,y)) = i\partial^x_\mu\left(R^\mu_{2/1}+N^\mu_{2/1}\right) + i\partial^y_\mu\left(R^\mu_{2/2}+N^\mu_{2/2}\right),$$

with an obvious notation, then CGI to second order holds.

When computing in practice, one is liable to find identities in distribution theory like

$$\begin{aligned} &\partial^x_\mu[A(x)B(y)\delta(x-y)] + \partial^y_\mu[A(y)B(x)\delta(x-y)] \\ &= \partial_\mu A(x)\,B(x)\delta(x-y) + A(x)\,\partial_\mu B(x)\delta(x-y) \qquad (26) \\ \text{and}\quad &A(x)B(y)\partial^x_\mu\delta(x-y) + A(y)B(x)\partial^y_\mu\delta(x-y) \\ &= A(x)\,\partial_\mu B(x)\delta(x-y) - \partial_\mu A(x)\,B(x)\delta(x-y). \qquad (27) \end{aligned}$$

We make the following observation: since

$$A(x)B(y)\delta(x-y) = A(x)B(x)\delta(x-y),$$

it must be

$$\partial^x_\mu\big(A(x)B(y)\delta(x-y)\big) = \partial^x_\mu\big(A(x)B(x)\delta(x-y)\big),$$

which forces

$$B(y)\partial^x_\mu\delta(x-y) = B(x)\partial^x_\mu\delta(x-y) + \partial_\mu B(x)\delta(x-y). \qquad (28)$$

We are able to prove both (26) and (27) from (28).

$$\begin{aligned} &\partial^x_\mu[A(x)B(y)\delta(x-y)] + \partial^y_\mu[A(y)B(x)\delta(x-y)] \\ &= \partial_\mu A(x)\,B(x)\delta(x-y) + A(x)B(y)\partial^x_\mu\delta(x-y) \\ &\quad + \partial_\mu A(x)\,B(x)\delta(x-y) - A(y)B(x)\partial^x_\mu\delta(x-y) \\ &= \partial_\mu A(x)\,B(x)\delta(x-y) + A(x)B(y)\partial^x_\mu\delta(x-y) \\ &\quad - A(x)B(x)\partial^x_\mu\delta(x-y) = \partial_\mu A(x)\,B(x)\delta(x-y) + A(x)\,\partial_\mu B(x)\delta(x-y), \end{aligned}$$

where we have used (28) twice. Analogously,

$$\begin{aligned} &A(x)B(y)\partial^x_\mu\delta(x-y) + A(y)B(x)\partial^y_\mu\delta(x-y) = A(x)B(x)\partial^x_\mu\delta(x-y) \\ &\quad + A(x)\partial_\mu B(x)\delta(x-y) - A(y)B(x)\partial^x_\mu\delta(x-y) \\ &\quad = A(x)\,\partial_\mu B(x)\delta(x-y) - \partial_\mu A(x)\,B(x)\delta(x-y), \end{aligned}$$

using (28) twice again.

Again after excruciatingly long calculations, by the sketched method one recovers the four-graviton couplings (22), plus terms with ghosts that we omit. Nevertheless, the road seems barred in that, in order to rederive the EH Lagrangian, one would have to perform an infinite number of calculations. Put in another way, we could never finish ascertaining that the EH Lagrangian fulfils CGI. (In a (re)normalizable theory it would be enough to verify CGI till third order, but this is not the case here.) For the latter, a better way can be contrived, though. Leaving aside the question of uniqueness (in spite of "folk theorems", uniqueness there is not: see Sect. 1.4), one can jump to the conclusion that $\mathcal{L}_{\mathrm{EH}}$ does satisfy CGI. In the next section, we describe a simple, short, and rigorous argument for this.

Before pursuing, we take stock: a classical Lagrangian is extracted from a quantum theory because, for all computations, naturally starting at T_2, only *tree diagrams* are considered. *Par ce biais-ci* the limit $\hbar \downarrow 0$ is taken. Of course, it is legitimate to perform the CGI analysis on graphs containing loops. In that way, the appropriate radiative corrections to S_{EH} are obtained, although this is not for the fainthearted. See [45] for the graviton self-energy; discrepancies between the coefficients of those corrections are still found in the literature. Anomalies are lurking there as well.

A last comment is in order: we have not tackled the matter of (re)normalizability of the theory, which is in terms of the T_n is a bit involved. Suffice here to say that the conclusion is similar to that of standard arguments (on the basis of the dimensionality of G, for instance). It is true that in causal (re)normalization, there are no ultraviolet divergences as such. There is a problem of correct definition of distributions involved in the perturbative expansion of the $\mathbb{S}$-matrix. The price of a "non-normalizable" theory like Einstein's is an infinite number of normalization constants in the process of that definition. This is not automatically so damning (also in regard of the discussion in the previous section), since perhaps they could be fixed by experiments, or have unobservable consequences. At any rate, the famous one-loop finiteness result by 't Hooft and Veltman – consult for instance the discussion in [46, Sect. III] – means that, at next order in pure gravity, no (new normalization constants and thus no) new geometrical invariants are introduced: another rule of the godly quarantine.

1.3.7 CGI at All Orders: Going for It

We rely in the following on a theorem by Dütsch [38]: BRS invariance of a Lagrangian, depending only on the fields and their first derivatives and carrying nonnegative powers of the couplings, implies local conservation of the BRS current. The latter implies CGI in the Heisenberg representation for tree graphs, and this result is kept in passing to time-ordered products. BRS invariance means precisely that the action of the BRS operator on the Lagrangian is a divergence, without use of the field equations. This admitted, the proof of CGI for the EH Lagrangian – modified like in formula (17) – by means of the BRS formulation of gravity by Kugo and Ojima [39] is simplicity itself.

In (our version of) that formulation, one keeps (18) and uses new gauge variations. The coboundary operator now is of the form

$$s = s_0 + \lambda s_1.$$

Here s_0 acts exactly like s of (5) and

$$\begin{aligned} s_1 h^{\mu\nu} &= i \left(h^{\mu\rho} \delta^\nu_\tau + h^{\nu\rho} \delta^\mu_\tau \right) \partial_\rho u^\tau - i \partial_\tau (h^{\mu\nu}) u^\tau, \\ s_1 u &= -i (u \cdot \partial) u, \\ s_1 \tilde{u} &= 0. \end{aligned} \tag{29}$$

Sotto voce we are introducing here the Lie derivative of $(g^{\mu\nu})$ with respect to the ghost vector field, thus diffeomorphism invariance. The new Lagrangian, complete with gauge-fixing and ghost terms, is

$$\mathcal{L}_{\text{total}} = -H + \mathcal{L}_{\text{gf}} + \mathcal{L}_{\text{ghost}} = -H + \tfrac{1}{2}(\partial \cdot h) \cdot (\partial \cdot h) + \tfrac{i}{2}(\partial_\nu \tilde{u}_\mu + \partial_\mu \tilde{u}_\nu) s h^{\mu\nu}.$$

Of course $\mathcal{L}_{\text{total}}$ is not diffeomorphism invariant. Compare (11). Note that

$$\mathcal{L}_{\text{gf}} = \mathcal{L}^{(0)}_{\text{gf}} = -\tfrac{1}{2}(s\tilde{u})^2,$$

while $\mathcal{L}_{\text{ghost}}$ has terms of order λ. From this,

$$s^2 h = 0, \quad s^2 u = 0, \quad s^2 \tilde{u}_\mu = -\frac{\delta S_{\text{total}}}{\delta \tilde{u}^\mu},$$

vanishing on-shell. It is known from [39] that

$$s\mathcal{L}_{\text{EH}} = -i\lambda \partial \cdot (u \mathcal{L}_{\text{EH}}),$$

and since, with

$$F^\alpha := (\partial^\rho h_{\beta\rho}) s h^{\alpha\beta} \quad \text{we have} \quad s(\mathcal{L}_{\text{gf}} + \mathcal{L}_{\text{ghost}}) = i \partial \cdot F,$$

it would seem that BRS invariance is checked, and we are done. Actually $\mathcal{L}_{\text{EH}}$ does not fulfil the conditions of Dütsch's theorem. However, we can use (17) and (19) to conclude. Indeed

$$-sH = -i\lambda \partial \cdot (u \mathcal{L}_{\text{EH}}) - i \partial \cdot (sD) + \frac{i}{\lambda} \partial \cdot (\Box u - \partial(\partial \cdot u)).$$

The last vector is conserved, but the point is that it cancels the term of the form

$$\frac{1}{\lambda} s\left(\tfrac{1}{2}\partial_\gamma \varphi + \partial^\rho h_{\gamma\rho}\right)$$

in sD. Then

$$s(\mathcal{L}_{\text{total}}) = s(-H + \mathcal{L}_{\text{gf}} + \mathcal{L}_{\text{ghost}}) = -i\partial \cdot \Big(\lambda u \mathcal{L}_{\text{EH}} + sD - F - \frac{\Box u}{\lambda} + \frac{\partial(\partial \cdot u)}{\lambda}\Big);$$

that is

$$s(\mathcal{L}_{\text{total}}) = \partial \cdot I \quad \text{with } I \text{ of the form} \quad I = \sum_{k=0}^{\infty} \lambda^k I^{(k)},$$

and all is well. (The funny and revealing thing in all this is that the parts in $1/\lambda^2$ and $1/\lambda$ in the EH Lagrangian do not contribute to the equations of motion.)

It is instructive to compare the tensor and vector cases. In order to see the parallel, one ought to replace (the massless version of) formulae (5) by

$$\begin{aligned} sA_a^\mu(x) &= iD_{ab}^\mu u_b(x), \\ su_a(x) &= -\tfrac{i}{2} g f_{abc} u_b u_c, \\ s\tilde{u}_a(x) &= -i\big(\partial \cdot A_a(x)\big). \end{aligned}$$

Like there, it is plain that the action of the BRS operator increases ghost number by one. Here f_{abc} denotes the structure constants of a Yang – Mills model, and D is the corresponding covariant derivative.

1.3.8 Details on Quantization and Graviton Helicities

The reader might be curious to know how the physical degrees of freedom emerge under our canonical recipe.

Let us treat ghosts first. Consider a family of absorption and emission operators $c_a^\alpha(\mathbf{k})$ with $a = 1, 2$ and standard anticommutators

$$\Big[c_a^\alpha(\mathbf{k}), c_b^\beta(\mathbf{k}')\Big]_+ = \delta_{ab}\delta_{\alpha\beta}\delta(\mathbf{k} - \mathbf{k}')$$

defining a *bona fide* Fock space, with the definitions

$$\begin{aligned} u^\alpha(x) &= (2\pi)^{-3/2} \int d\mu(k) \left(e^{-ikx} c_2^\alpha(\mathbf{k}) - g^{\alpha\alpha} e^{ikx} c_1^\alpha(\mathbf{k})^\dagger\right), \\ \tilde{u}^\alpha(x) &= -(2\pi)^{-3/2} \int d\mu(k) \left(e^{-ikx} c_1^\alpha(\mathbf{k}) + g^{\alpha\alpha} e^{ikx} c_2^\alpha(\mathbf{k})^\dagger\right), \end{aligned} \qquad (30)$$

where $d\mu(k)$ is the usual Lorentz invariant volume over the lightcone. There is a Krein operator on the ghost Fock space that allows for u being self-conjugate and $\tilde{u}$ being skew-conjugate. This can be achieved by

$$c_1^i(\mathbf{k})^+ = c_2^i(\mathbf{k})^\dagger, \quad c_2^i(\mathbf{k})^+ = c_1^i(\mathbf{k})^\dagger, \quad c_1^0(\mathbf{k})^+ = -c_2^0(\mathbf{k})^\dagger, \quad c_2^0(\mathbf{k})^+ = -c_1^0(\mathbf{k})^\dagger,$$

with $i = 1, 2, 3$. Then formulae (30) are rewritten as

$$u^\alpha(x) = (2\pi)^{-3/2} \int d\mu(k)\, \Big(e^{-ikx} c_2^\alpha(\mathbf{k}) + e^{ikx} c_2^\alpha(\mathbf{k})^+\Big) = u^\alpha(x)^+,$$
$$\tilde{u}^\alpha(x) = (2\pi)^{-3/2} \int d\mu(k)\, \Big(-e^{-ikx} c_1^\alpha(\mathbf{k}) + e^{ikx} c_2^\alpha(\mathbf{k})^\dagger\Big) = -\tilde{u}^\alpha(x)^+. \qquad (31)$$

From (30) or (31) we obtain for $u, \tilde{u}$ the wave equations. *Covariant* anticommutation relations (13) also follow.

Note now

$$t^{\alpha\beta\mu\nu} := \tfrac{1}{2}\big(\eta^{\alpha\mu}\eta^{\beta\nu} + \eta^{\alpha\nu}\eta^{\beta\mu} - \tfrac{1}{2}\eta^{\alpha\beta}\eta^{\mu\nu}\big) = t^{\mu\nu\alpha\beta}.$$

That is,

$$\big(t^{\mu\nu\alpha\beta}\big) = \begin{pmatrix} 3/4 & 1/4 & 0 & 0 \\ 1/4 & \begin{pmatrix} 3/4 & -1/4 \\ -1/4 & 3/4 \end{pmatrix} & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix}$$

on a $(0,0), (j,j), (0,j), (j,l)$ block basis, with $j, l = 1, 2, 3$, $j \neq l$; and in particular

$$T \equiv \big(t^{\mu\mu\alpha\alpha}\big) = \begin{pmatrix} 0 & 1/4 & 1/4 & 1/4 \\ 1/4 & 3/4 & -1/4 & -1/4 \\ 1/4 & -1/4 & 3/4 & -1/4 \\ 1/4 & -1/4 & -1/4 & 3/4 \end{pmatrix}$$

on the $(0,0), (1,1), (2,2), (3,3)$ basis. Next we note that

$$T = MM^\dagger, \quad \text{with} \quad M = \begin{pmatrix} 0 & 1/2 & 1/2 & 1/2 \\ 0 & -1/2 & 1/2 & 1/2 \\ 0 & 1/2 & -1/2 & 1/2 \\ 0 & 1/2 & 1/2 & -/2 \end{pmatrix}.$$

Next we invoke operators defining a Fock space:

$$[b_{\alpha\beta}(\mathbf{k}), b_{\mu\nu}(\mathbf{k}')] = \tfrac{1}{2}(\delta_{\alpha\mu}\delta_{\beta\nu} + \delta_{\alpha\mu}\delta_{\beta\mu})\delta(\mathbf{k} - \mathbf{k}'),$$

with $b_{\alpha\beta} = b_{\beta\alpha}$. Define now operators $a_{\alpha\beta}$, with $a_{\alpha\beta} = a_{\beta\alpha}$ as well, by $a_{\alpha\beta} = b_{\alpha\beta}$ for $\alpha \neq \beta$ and

$$a_{\alpha\alpha} = \sum_\beta M_{\alpha\beta} b_{\beta\beta}.$$

The rule

$$\left[a_{\alpha\beta}(\mathbf{k}), a^{\dagger}_{\mu\nu}(\mathbf{k}')\right] = g^{\alpha\alpha} g^{\beta\beta} t^{\alpha\beta\mu\nu} \delta(\mathbf{k} - \mathbf{k}')$$

follows.

The scalar field is now constructed in a way close to the standard one:

$$\varphi(x) = (2\pi)^{-3/2} \int d\mu(k)\, (e^{-ikx} a(\mathbf{k}) - e^{ikx} a^{\dagger}(\mathbf{k})), \tag{32}$$

where the (not Lorentz covariant) operators $a^{\#}$ satisfy

$$[a(\mathbf{k}), a^{\dagger}(\mathbf{k})] = 4\delta(\mathbf{k} - \mathbf{k}').$$

The traceless sector is represented as

$$H^{\alpha\beta}(x) = (2\pi)^{-3/2} \int d\mu(k)\, \left(e^{-ikx} a_{\alpha\beta}(\mathbf{k}) + g^{\alpha\alpha} g^{\beta\beta} t^{\alpha\beta\mu\nu} e^{ikx} a^{\dagger}_{\alpha\beta}(\mathbf{k})\right).$$

Now one can verify (12) painstakingly.

The last task in this section is to identify finally the physical degrees of freedom. For that, let us choose and fix $k^{\mu} = (\omega, 0, 0, \omega)$. One can verify that the only states not present in Q (that is, belonging to the kernel of $[Q, Q^{\dagger}]_{+}$) are

$$(b_{11} - b_{22})^{\dagger}|0\rangle \quad \text{and} \quad b^{\dagger}_{12}|0\rangle = b^{\dagger}_{21}|0\rangle.$$

They correspond to linear polarization states. Their complex combinations (circular polarization states) may be represented by matrices

$$\varepsilon_{\pm} := \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & \pm i & 0 \\ 0 & \pm i & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

which transform like

$$\varepsilon'_{\pm} = e^{\pm 2i\phi} \varepsilon_{\pm}$$

under a rotation of angle ϕ about the direction of propagation. The reader can verify this by using the generator of rotations

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The two ± 2 helicity states have been thereby identified. These states satisfy

$$\varepsilon_{\pm}^{\mu\nu} k_{\nu} = 0. \tag{33}$$

These conditions are not Lorentz invariant. Notice the associated gauge freedom

$$\varepsilon_{\pm}^{\mu\nu} \to \varepsilon_{\pm}^{\mu\nu} + k^{\mu} f^{\nu} + f^{\mu} k^{\nu} - \eta^{\mu\nu}(k \cdot f).$$

We may add

$$\varepsilon_{\pm\nu}^{\ \ \nu} = 0. \tag{34}$$

This five conditions (33) and (34) are also possible for a massive graviton – say $k = (m, 0, 0, 0)$. Thus they characterize the spin two case in general, with up to five degrees of freedom. Now, for k lightlike as above, let e^1, e^2 denote two spacelike vectors orthogonal to k and mutually orthogonal, say $(0, 1, 0, 0)$, $(0, 0, 1, 0)$. The tensors

$$(k_{\mu}k_{\nu}) = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \left(k_{\mu}e^1_{\nu} + e^1_{\mu}k_{\nu}\right) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \left(k_{\mu}e^2_{\nu} + e^2_{\mu}k_{\nu}\right) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

verify (33) and (34) as well. They represent the three helicity states that disappear in the massless case.

1.3.9 Final Remarks

- The geometrical form of general relativity, due to Einstein, is supremely elegant for some. However, the accompanying *interpretation* clashes with the one advocated here, based in the identification of the quanta of the gravitational field and more or less standard quantum field theory procedures, not to speak of table top experiments. Since experiments probe gravity theory to very low orders in $G, \hbar$, one should keep an open mind and welcome any consistent quantum theory perturbatively compatible with general relativity. As string theory promises to be.
- Coupling to matter. The graviton naturally couples to another symmetric tensor field:

$$T_1^{\text{matter}} = i\lambda A_{\alpha\beta\mu\nu} h^{\alpha\beta} T^{\mu\nu} \quad \text{with} \quad sT = 0.$$

  Consideration that sT_1^{matter} must be a divergence leads at once to

$$\partial_{\mu} T^{\mu\nu} = 0;$$

just like it leads to charge conservation in quantum electrodynamics. Of course, the only conserved second-rank symmetric tensor in Poincaré-invariant field theory is the stress – energy tensor.

- Infrared freedom: in the Epstein – Glaser dispensation, vacuum diagrams, as any others, are ultraviolet-finite. Because of their high degree of singular order, however, we are assured that they are infrared finite. Therefore the vacuum is stable (no colour confinement or anything of the sort): a bonus for quantum gravity.
- The CGI formalism allows one to deal with massive gravity as well [47], although the shortcut in Sect. 1.3.7 apparently is not available. At the price of introducing Stückelberg-like vector Bose ghosts, the massless limit of massive gravity is relatively smooth. Suggestively, a cosmological constant $\Lambda = m^2/2$, with m the graviton mass, ensues; one is reminded of Mach's principle, as well.
  Note that the Fadeev – Popov approach to ghosts in quantum gravity is linked to existence of quasi-invariant measures on diffeomorphism groups [48].

1.3.10 Other Ways

- Path-integral quantization faces the stark difficulty (rather, the impossibility) of "counting" four-dimensional manifolds [49]. A way around it may be "dynamical triangulation" – see [50] and in the same vein the recent [51].
- We cannot close the section without mentioning the promise of "asymptotic safety" in quantum gravity, developed by Reuter and coworkers. Consult [52], and references therein. There are intriguing results within this approach, pointing out to effective 2-dimensionality of spacetime at the Planck scale – which has been used by Connes, somewhat dubiously, to justify that the finite noncommutative geometry part in his reinterpretation of the standard model Lagrangian be of *KO*-dimension 6 [53]. While, at the other end of the scale, exceptionally good infrared behaviour could mimic both "dark matter" and "dark energy" behaviour.
- In relation with the discussion at the end of Sect.1.3.6, support for the idea that UV divergences in gravity are not so intractable has come recently from work by Kreimer [54].

1.4 The Unimodular Theories

A recent edition of a standard text about cosmology by a well-respected author [55] ends with a chapter on "Twenty controversies in cosmology today". In the first one, about general relativity, he declares

> In fact it is theories without effective rivals that require the most vigilant testing.

Without contradicting this wisdom, let me point out that general relativity has some rivals which are too close for comfort. In order to grapple with them, let us go back to the fundamentals. We did omit the proof of that, for suitable variations of the metric $(g_{\alpha\beta})$, the Einstein *field equations* in vacuum

$$G^{\alpha\beta} + \Lambda g^{\alpha\beta} := R^{\alpha\beta} - \tfrac{1}{2} R g^{\alpha\beta} + \Lambda g^{\alpha\beta} = 0 \tag{35}$$

are equivalent to

$$\frac{\delta S_{\mathrm{EH}}}{\delta g_{\alpha\beta}} = 0.$$

It is worthwhile to go through that routine here. Now

$$S_{\mathrm{EH}} = -\frac{1}{16\pi G} \int d^4x \, \sqrt{-\det g}\,(R - 2\Lambda).$$

Clearly

$$\delta S_{\mathrm{EH}} = \frac{1}{16\pi G} \int d^4x \Big[-(R-2\Lambda) \frac{\delta\sqrt{-\det g}}{\delta g_{\alpha\beta}} \Big] + \sqrt{-\det g}\big[R^{\alpha\beta}\delta g_{\alpha\beta} + g^{\alpha\beta}\delta R_{\alpha\beta} \big],$$

where we take into account

$$R^{\alpha\beta}\delta g_{\alpha\beta} = -R_{\alpha\beta}\delta g^{\alpha\beta}, \quad \text{since} \quad \delta g^{\rho\sigma} g_{\sigma\varepsilon} + g^{\rho\sigma}\delta g_{\sigma\varepsilon} = 0.$$

Now,

$$\delta\sqrt{-\det g} = -\frac{1}{2\sqrt{-\det g}} \frac{\partial(-\det g)}{\partial g_{\alpha\beta}} \delta g_{\alpha\beta} = \tfrac{1}{2}\sqrt{-\det g}\, g^{\alpha\beta}\delta g_{\alpha\beta}.$$

It is easy to show that the last term in δS_{EH} does not contribute to the variation of the action. Therefore

$$\frac{\delta S_{\mathrm{EH}}}{\delta g_{\alpha\beta}} = \frac{\sqrt{-\det g}}{16\pi G} \big(R^{\alpha\beta} - \tfrac{1}{2} R g^{\alpha\beta} + \Lambda g^{\alpha\beta} \big),$$

hence (35).

It is apparent that life would be much simpler if $\sqrt{-g}$ were not a dynamical quantity. This is suggested by Weinberg in his well-known review [56], in relation with the discussion in Sect. 1.6; the idea basically goes back to Einstein. Let us see what happens. First of all Λ seems to vanish from the picture. Second, since now the action has to be stationary only with respect to variations keeping $\det g$ invariant, that is $g^{\alpha\beta}\delta g_{\alpha\beta} = 0$, one gathers the elegant

$$R^{\alpha\beta}_{\mathrm{trace-free}} = R^{\alpha\beta} - \tfrac{1}{4} g^{\alpha\beta} R = 0.$$

As it turns out, these are the Einstein equations again! The reason is that the contracted Bianchi identities

$$\nabla_\beta R^{\alpha\beta} = \tfrac{1}{2}\nabla^\alpha R, \quad \text{that is} \quad \nabla_\beta G^{\alpha\beta} = 0$$

are still valid. They can be derived from $R_{\mu\nu} = g^{\sigma\rho} R_{\sigma\mu\rho\nu}$ and the uncontracted Bianchi identities:

$$\partial_\tau R_{\mu\nu\rho\sigma} + \partial_\sigma R_{\mu\nu\tau\rho} + \partial_\rho R_{\mu\nu\sigma\tau} = 0.$$

Therefore, by integration,

$$-R = G^\alpha_\alpha = -4\kappa, \quad \text{and then} \quad G^{\alpha\beta} + \kappa g^{\alpha\beta} = 0,$$

which is but (35) with κ replacing Λ. However, the interpretation has changed. The term in Λ in the action does not contribute anything (so the Minkowski space is a solution of the field equations even in the presence of such a term), and κ arises as an initial condition.

Remark 1 The discussion in this section is mainly pertinent in the presence of matter. If we define here the matter stress – energy tensor $T \equiv (T^{\alpha\beta})$ by

$$\delta S_{\text{matter}} =: \tfrac{1}{2} \int d^4x \sqrt{-\det g}\, T^{\alpha\beta} \delta g_{\alpha\beta},$$

then varying $S_{\text{matter}} + S_{\text{EH}}$ while keeping the determinant fixed results in

$$R^{\alpha\beta}_{\text{trace-free}} = 8\pi G\, T^{\alpha\beta}_{\text{trace-free}}.$$

Since the conservation law $\nabla \cdot T = 0$ holds, we have now

$$R - 8\pi G T^\alpha_\alpha = 4\kappa,$$

and finally

$$R^{\alpha\beta} - \tfrac{1}{2} R g^{\alpha\beta} + \kappa g^{\alpha\beta} = 8\pi G\, T^{\alpha\beta},$$

exactly the usual Einstein equations in the presence of a cosmological constant term plus matter, with the mentioned replacement of Λ by κ, and the attending change of interpretation.

It should be remarked that we are not implying that the classical action for gravitational physics is invariant only under coordinate transformations ("transverse diffeomorphisms") that preserve the volume element. This is a stronger claim. Elegant justification for it is found in [57]. In accordance with the above, all known tests of general relativity probe equally the (several) unimodular theories. It has been argued that the matter graviton coupling gives rise to inconsistencies when "strong" unimodularity holds [58], but this objection we know not in relation with weak unimodularity. Only quantum effects would in principle allow tell it and general relativity apart [59] – after all the "measure" of the quantum functional integral for gravity is changed. Meanwhile, the interest of the unimodular theory is twofold: as indicated by Weinberg, it alleviates the cosmological constant problem (Sect. 1.6); moreover, it is natural from the current formulation of noncommutative manifold theory (Sect. 1.5.9.2). From the viewpoint of the preceding section, the key question is how the unimodular theory is arrived at the $\hbar \downarrow 0$ limit of a quantum theory of gravitons. We must leave the matter aside.

1.5 The Noncommutative Connection

1.5.1 Prolegomena

There is no general theory of noncommutative spaces. The practitioners' tactics has been that of multiplying the examples, whereas trying to anchor the generalizations on the more solid ground of ordinary (measurable/topological/differentiable/Riemannian, etc.) spaces. This is what we try to do here, within the limitations imposed by the knowledge of the speaker.

The first task is to learn to think of ordinary spaces in noncommutative terms. Arguably, this goes back to the Gelfand – Naĭmark theorem (1943), establishing that the information on any locally compact Hausdorff topological space X is fully stored in the commutative algebra $C(X)$ of continuous function over it, vanishing at ∞. This is a way to recognize the importance of C^*-algebras and to think of them as locally compact Hausdorff noncommutative spaces. If we had just asked for the functions to be measurable and bounded, we would have been led to von Neumann algebras. Vector bundles are identified through their spaces of sections, which algebraically are projective modules of finite type over the algebra of functions associated to the base space – this is the Serre – Swan theorem (1962). In this way, we come to think of noncommutative vector bundles.

Under the influence of quantum physics, the general idea is then to forget about sets of points and obtain all information from classes of functions, e.g. open sets in X are replaced by ideals. The rules of the game would then seem to be (1) find a way to express a mathematical category through algebraic conditions, and (2) relinquish commutativity. This works wonders in group theory, which is replaced by bialgebra theory, relinquishing (co)commutativity. However, that kind of generalization quickly runs into sands, for two reasons: (i) Some mathematical objects, like differentiable manifolds, and de Rham cohomology, are reluctant to direct noncommutative generalization. The same is true of Riemannian geometry; after all, all smooth manifolds are Riemann. (ii) Genuinely new "noncommutative phenomena" are missed.

For instance, in the second respect, in many geometrical situations the associated set is very pathological, and a direct examination yields no useful information. The set of orbits of a group action, such as the rotation of a circle by multiples of an irrational angle θ, is generally of this type. In such cases, when we examine the matter from the algebraic point of view, we are sometimes able to obtain a perfectly good operator algebra that holds the information we need; however, this algebra is generally not commutative.

One can situate the beginning of noncommutative geometry (NCG) in the 1980 paper by Connes, where the “noncommutative torus” T_θ was studied [60]. Not only is this algebra able to answer the question mentioned above, but one can decide what are the smooth functions on this noncommutative space, what vector bundles and connections on T_θ are, and, decisively, how to construct a Dirac operator on it.

Even now, the importance of this early example in the development of the theory can hardly be underestimated. The noncommutative torus provides a simple but nontrivial example of *spectral triple* (A, H, D) – see further on for the notation – or “noncommutative spin manifold”, the algebraic apparatus with which Connes eventually managed to push aside the obstacles to the definition of noncommutative Riemannian manifolds. The Dirac equation naturally lives on spin manifolds, and these constitute the crucial paradigm, too, for Connes program of research (and unification) of mathematics.

The more advanced rules of the game would now seem to be the following: (1) Escape the difficulties “from above” by finding the algebraic means of describing a richer structure. If we reformulate algebraically what a spin manifold is, we can describe its de Rham cohomology, its Riemannian distance and like geometrical concepts, algebraically as well. Choice of a Dirac operator D means imposing a metric. However, there is the risk that the link to the commutative world is obscured. (2) Therefore, make sure that the link is kept. In other words, prove that a noncommutative spin manifold is in fact a spin manifold in the everyday sense (!) when the underlying algebra is commutative. In point of fact, the second desideratum only received a definitive, satisfactory answer a few weeks ago.

1.5.2 Ironies of History

The following quotation of a popular book [61] provides a convenient rallying point:

> When physicists talk about the importance of beauty and elegance in their theories, the Dirac equation is often what they have in mind. Its combination of great simplicity and surprising new ideas, together with its ability both to explain previously mysterious phenomena and predict new ones [spin], make it a paradigm for any mathematically inclined theorist.

Thus the irony is in that, first and foremost [61],

> Mathematicians were much slower to appreciate the Dirac equation and it had little impact on mathematics at the time of its discovery. Unlike the case with the physicists, the equation did not immediately answer any questions that mathematicians had been thinking about.

The situation changed only *40 years* later, with the Atiyah – Singer theory of the index.

A second and minor irony is that, now that spin manifold theory is an established and respectable line of mathematical business, its community of practitioners seems mostly oblivious to the fact it underpins a whole new branch/paradigm/method of doing mathematics (although something is being done to fill up this gap).

Now come the *informal rules* for noncommutative geometers – rules which in any society insiders recognize as the most binding. These seem to be the following:

(1) Keep close to physics and in particular to quantum field theory. There is no doubt that Connes came to his "axioms" for noncommutative manifolds by thinking of the Standard Model of particle physics as a noncommutative space. (2) Try to interpret and solve most problems conceivably related to noncommutative geometry by use of spectral triple theory. This of course is not to everyone's taste, and a cynic could say "Whoever is good with the hammer, thinks everything is a nail"; moreover, it is of course literally impossible, as the mathematical world teems with virtual objects for which complete taxonomy is an impossible task. It has proved surprisingly rewarding, however.

A caveat about (2): there is an underlying layer of index theory and K-theory, which is a deep way of addressing quantization. But even there, when you need to compute K-theoretic invariants, you are led back to smoother structures where you have more tools, like (A, H, D).

1.5.2.1 A First Conceptual Star

Let us we imagine a star, with NCG in the centre, of subjects intimately related to it. This will include

- Operator algebra theory
- K-theory and index theory
- Hochschild and cyclic homology
- Bialgebras and Hopf algebras, including quantum groups
- Foliations, groupoids
- Singular spaces
- Deformation and quantization theory
- Topics in physics: quantum field theory, including noncommutative field theory and renormalization; gauge theories, including the Standard Model; condensed matter; gravity; strings

1.5.3 Spectral Triples

The root of the importance of spectral triples in NCG is found in *algebraic topology*. Noncommutative topology brings techniques of operator algebra to algebraic topology – and vice versa. As indicated earlier, the method of rephrasing concepts and results from topology using Gelfand – Naĭmark and Serre – Swan equivalence, and extending them to some category of noncommutative algebras, recurs for a while. Moreover, deeper proofs of some properties of objects in the commutative world are to be found in their noncommutative counterparts, with Bott periodicity providing an outstanding example.

Now to extend the standard (co)homology functors (not to speak of homotopy) is rather difficult. On the other hand, Atiyah's K-functor generalizes very smoothly. Given a unital algebra A, its algebraic K_0-group is defined as the Grothendieck group of the (direct sum) semigroup of isomorphism classes of finitely generated projective right (or left) modules over A. Then in view of the Serre – Swan theorem $K_0(C(X)) = K^0(X)$.

Given an ordinary space X, the real K-group $KO^0(X)$ – actually, it is a ring, with product given by pullback by the diagonal map of the tensor product – for X is obtained as the Grothendieck group for real vector bundles. Higher order groups are defined by suspension. If X is Hausdorff and compact, we have $KO^i(X) \simeq KO^{i+8}(X)$; this is real Bott periodicity. Recall that we have $KO^0(*) = \mathbb{Z}$, $KO^1(*) = KO^2(*) = \mathbb{Z}_2$, $KO^3(*) = 0$, $KO^4(*) = \mathbb{Z}$, $KO^5(*) = KO^6(*) = KO^7(*) = 0$. There is an isomorphism of the spin cobordism classes of a manifold X onto $KO^\bullet(X)$ [62].

The *K-homology* of topological spaces can be developed as a functorial theory whose cycles pair with vector bundles in the same way that currents pair with differential forms in the de Rham theory. Such cycles are given, interestingly enough, by spin^c structures. On the other hand, the index theorem shows that the right partners for vector bundles are elliptic pseudodifferential operators (with the pairing given by the index map). We can think of abstract K-cycles as of phases of Dirac operators. In NCG we want to generalize both this and the line element (entering the realm of Riemannian geometry). Note the result

Proposition 1 *On a spin manifold the geodesic distance between two points obeys the formula*

$$d(p,q) = \sup\{\,|f(x) - f(y)| : f \in C(X),\ \ \|[D,f]\| \leq 1\,\}. \tag{36}$$

This is actually trivial, since $\|[D,f]\|$ is the Lipschitz norm of f.

The foregoing motivates:

Definition 1 A noncommutative geometry (spectral triple) is a triple $(\mathcal{A}, H, D)$, where $\mathcal{A}$ is a $*$-algebra represented faithfully by bounded operators on the Hilbert space H and D is a self-adjoint operator $D : \mathrm{Dom}D \to \mathrm{H}$, with $\overline{\mathrm{Dom}D} = H$, such that $[D,a]$ extends to a bounded operator and $a(1+D^2)^{-1/2}$ is a compact operator, for any $a \in \mathcal{A}$; plus a postulate set of conditions given below.

We do not explicitly indicate the representation in the notation. A spectral triple is *even* when there exists on H a symmetry Γ such that $\mathcal{A}$ is even and D odd with respect to the associated grading. Otherwise, it is odd. A spectral triple is *compact* when $\mathcal{A}$ is unital; it is then enough to require that $(1+D^2)^{-1/2}$ be compact.

One should think of $\mathcal{A}$ as an algebra of "smooth", not "continuous" elements. Of course, it is important that $K(\mathcal{A}) = K(\bar{\mathcal{A}})$, with $\bar{\mathcal{A}}$ the C^*-algebra completion of $\mathcal{A}$. Sufficient conditions are known for this.

In the compact case the maximal set of postulates includes

1. *Summability or Dimension*: for a fixed *positive integer p*, we have

$$(1+D^2)^{-1/2} \in L^{p,+}(H), \quad \text{implying} \quad \mathrm{Tr}_\omega((1+D^2)^{-p/2}) \geq 0,$$

for all generalized limits ω, and moreover, $\mathrm{Tr}_\omega((1+D^2)^{-p/2}) \neq 0$.

If we have regularity (see directly below), then the functional on $\mathcal{A}$

$$a \mapsto \mathrm{Tr}_\omega(a(1+D^2)^{-p/2})$$

is a hypertrace.

2. *Regularity*: with $\delta a := [|D|, a]$, one has

$$\mathcal{A} \cup [D, \mathcal{A}] \subseteq \bigcap_{m=1}^{\infty} \mathrm{Dom}\, \delta^m.$$

3. *Finiteness:* the dense subspace of H which is the smooth domain of D,

$$H_\infty := \bigcap_{m\geq 1} \mathrm{Dom}\, \mathrm{D}^m,$$

is a finitely generated projective (left) $\mathcal{A}$-module, which carries an $\mathcal{A}$-valued Hermitian pairing $(\cdot \mid \cdot)_{\mathcal{A}}$ satisfying

$$\langle \xi \mid a\eta \rangle = \mathrm{Tr}_\omega\big(\mathrm{a}\,(\xi \mid \eta)_{\mathcal{A}}(1+D^2)^{-p/2}\big)$$

when $\xi, \eta \in H_\infty$, and $a \in \mathcal{A}$. This also implies the *absolute continuity* property of the hypertrace:

$$\mathrm{Tr}_\omega(\mathrm{a}(1+D^2)^{-p/2}) > 0, \quad \text{whenever } a > 0 \text{ in } \mathcal{A}.$$

4. *First-order condition*: as well as the defining representation we require a commuting representation of the opposite algebra $\mathcal{A}^\circ$. Now H_∞ can be regarded as a right $\mathcal{A}$-module. Then we furthermore ask for $[[D, a], b] = 0$ for $a \in \mathcal{A}$, $b \in \mathcal{A}^\circ$. (When $\mathcal{A}$ is commutative, we could still have different left and right actions on H. If they are equal, the postulate entails that the subalgebra $\mathcal{C}_D\mathcal{A}$ of $\mathcal{B}(H)$ generated by $\mathcal{A}$ and $[D, \mathcal{A}]$ belongs in $End_{\mathcal{A}}(H_\infty)$.)
5. *Orientation*: let p be the metric dimension of $(\mathcal{A}, H, D)$. We require that the spectral triple be even if and only if p is even. For convenience, we take $\Gamma = 1$ when p is odd. We say the spectral triple $(\mathcal{A}, H, D)$ is *orientable* if there exists a Hochschild p-cycle

$$\mathbf{c} = \sum_{\alpha=1}^{n} \left(a_\alpha^0 \otimes b_\alpha\right) \otimes a_\alpha^1 \otimes \cdots \otimes a_\alpha^p \in Z_p(\mathcal{A}, \mathcal{A} \otimes \mathcal{A}^\circ)$$

whose Hochschild class may be called the "orientation" of $(\mathcal{A}, H, D)$, such that

$$\pi_D(\mathbf{c}) := \sum_\alpha a_\alpha^0 b_\alpha \left[D, a_\alpha^1\right] \ldots \left[D, a_\alpha^p\right] = \Gamma. \tag{37}$$

6. *Reality*: there is an antiunitary operator $C : \mathcal{H} \to \mathcal{H}$ such that $Ca^*C^{-1} = a$ for all $a \in \mathcal{A}$, and moreover, $C^2 = \pm 1$, $CDC^{-1} = \pm D$, and also $C\Gamma C^{-1} = \pm\Gamma$ in the even case, according to the following table of signs depending only on p mod 8:

p mod 8	0 2 4 6
$C^2 = \pm 1$	$+ - - +$
$CDC^{-1} = \pm D$	$+ + + +$
$C\Gamma C^{-1} = \pm\Gamma$	$+ - + -$

p mod 8	1 3 5 7
$C^2 = \pm 1$	$+ - - +$
$CDC^{-1} = \pm D$	$- + - +$

For the origin of this sign table in KR-homology, we refer to [63]. (This postulate is optional, but important in practice. It makes the difference between spin^c and spin manifolds.)
7. *Poincaré duality*: the C^*-module completion of H_∞ is a Morita equivalence bimodule between $\bar{\mathcal{A}}$ and the norm completion of $\mathcal{C}_D\mathcal{A}$.

With the exception of the last, they are essentially in the form given to them by Connes.

What good are these terms? We have the following:

Proposition 2 *Let M be a compact Riemannian manifold without boundary with Riemannian volume form ν_g and assume there exists a spinor bundle S over it, with conjugation C. Define the* ***Dirac spectral triple*** *associated with it as*

$$(C^\infty(M), L^2(M, S), D\!\!\!/),$$

where $L^2(M, S)$ is the spinor space obtained by completing the spinor module $\Gamma^\infty(M, S)$ with respect to the natural scalar product (using $|\nu_g|$) and $D\!\!\!/ := -i(\hat{c} \circ \nabla^S)$ is the Dirac operator (for the notation: if c is the action of the Clifford algebra bundle over M, then $\hat{c}(\alpha, s) = c(\alpha)s$, for α in that bundle and s a spinor). Also $\Gamma = c(\gamma)$, where γ is the chirality element of the Clifford bundle, either the identity operator or the standard grading operator on $L^2(M, S)$, according as dim *M is odd or even.*

Then the Dirac spectral triple is a commutative noncommutative spin geometry. (Sorry for the bad joke!)

The proof is routine. We can relax postulate 6 and obtain just a spin^c geometry. The most important thing is to think of the spinor bundle as an algebraic object: this comes from Plymen's characterization [64], suggested by Connes, of spin^c structures as Morita equivalence bimodules for the Clifford action induced by the metric. The existence of that equivalence is tantamount to the vanishing of the usual topological obstruction to the existence of spin^c structures. A precedent for this algebraization is Karrer's [65]. A recent article by Trautman [66] contains interesting historical asides.

1.5.4 On the Reconstruction Theorem

So far, so good, but there will be a point to the precedent exercise only if we can prove that the algebraic terms of the previous section lead in an essentially unique way to a spin manifold. That is, assuming conditions 1–7, excluding 6 for the time being, and furthermore that $\mathcal{A}$ is commutative (this of course entails some simplification in the orientation axiom), is there a spin^c manifold M – with $\dim M = p$ – such that $A \simeq C^\infty(M)$ and similarly all of the original spectral triple is reproduced by its Dirac geometry?

Proof of this on the assumption that $A \simeq C^\infty(M)$ for some M is found already in [63]. An attempt to prove it without that strong assumption was announced in October 2006 by A. Rennie and J. C. Várilly [67]. However, this work had some flaws, recently corrected by Connes [68, 69].

Some extra technical assumptions are needed for the proof. Rennie and Várilly assume that the spectral triple $(\mathcal{A}, H, D)$ is *irreducible*, that is, the only operators in $\mathcal{B}(\mathcal{H})$ commuting (strongly) with D and with all $a \in \mathcal{A}$ are the scalars in $\mathbb{C}\,1$. (This ensures the connectedness of the underlying topological space M.) Moreover, they postulate the following *closedness* condition: for any p-tuple of elements $(a_1, .., a_p)$ in $\mathcal{A}$, the operator $\Gamma\,[D, a_1]\dots[D, a_p](1 + D^2)^{-p/2}$ has vanishing Dixmier trace; thus, for any ω,

$$\mathrm{Tr}_\omega\big(\Gamma\,[D, a_1]\dots[D, a_p]\,(1 + D^2)^{-p/2}\big) = 0.$$

This is an algebraic analogue of Stokes' theorem.

Their argument to show that the Gelfand – Naĭmark spectrum M of $\mathcal{A}$ is a differential manifold may be conceptually broken into two stages. The first is to construct a vector bundle over the spectrum which will play the role of the cotangent bundle. For that, one identifies local trivializations and bases of this bundle in terms of the "1-forms" $\left[D, a_\alpha^j\right]$ given by the orientability condition. The aim is then to show that the maps $a_\alpha = \left(a_\alpha^1, \dots, a_\alpha^p\right) : M \to \mathbb{R}^p$ provide coordinates on suitable open subsets of M; for that, one must prove that the maps a_α are open and locally one to one.

At this stage one needs to deploy, besides the technical conditions, postulates 1–5 on our spectral triple. A basic tool is a multivariate $\mathbb{C}^\infty$ functional calculus for regular spectral triples that enables to construct partitions of unity and local inverses within the algebra $\mathcal{A}$.

However, the strategy of [67] failed to ensure that the maps a_α are local homeomorphisms. Instead, Connes [69] resorted to the inverse function theorem [70] by showing that regularity and finiteness provide enough smooth derivations of $\mathcal{A}$ to build nonvanishing Jacobians where needed. This requires delicate arguments with unbounded derivations of C^*-algebras, and two other technical assumptions, replacing those of [67]:

- *Skewsymmetry* of the Hochschild cycle $\mathbf{c}$ under permutations of $a_\alpha^1, \dots, a_\alpha^p$. This enables one to bypass the cotangent bundle construction and omit the closedness property, but is arguably a stronger assumption.
- *Strong regularity*: all elements of $\mathrm{End}_{\mathcal{A}}(H_\infty)$, not merely those in $\mathcal{C}_D\mathcal{A}$, lie in $\bigcap_{m=1}^\infty \mathrm{Dom}\,\delta^{\mathrm{m}}$.

The local injectivity of the maps a_α is established by first showing that their multiplicity (as maps into $\mathbb{R}^p$) is bounded: this needs delicate estimates in order to invoke the measure theoretic results of Voiculescu [71, 72]. The smooth functional calculus can then be used to construct local charts at all points of M by small shifts of the original maps a_α.

Poincaré duality in K-theory plays no role in the reconstruction of a manifold as a compact space M with charts and smooth transition functions. However, once that has been achieved, it is needed to show that M carries a spin^c structure and to identify the class of $(\mathcal{A}, H, D)$ as the fundamental class of the spin^c manifold. This is done by showing that in this case $\mathrm{End}_{\mathcal{A}}(H_\infty)$ coincides with $\mathcal{C}_D\mathcal{A}$ – see [67, 69] – and in particular strong regularity is moot. The Dirac operator is shown to differ from D by at most an endomorphism of the corresponding spinor bundle. When M is spin, the latter can be eliminated by a variational argument – as shown by Kastler and by Kalau and Walze, the Wodzicki residue of $(1 + D^2)^{-p/2+1}$ gives the EH action; see [63, Sect. 11.4].

Once one has at one's disposal a spin^c structure, axiom 6 (Reality) allows to refine it to a spin structure. For that, we refer to [64] – or consult [63] – wherein it is shown that the spinor module for a spin structure is just the spinor module for a spin^c structure equipped with compatible change conjugation, which is none other than the real structure operator C (acting on H_∞); the spin structure is extracted, using C, from a representation of the real Clifford algebra of T^*M.

It is unlikely [H. Moscovici, private communication] that the reconstruction theorem holds under the more stringent conditions set out originally by Connes [73]. Possible redundancy of the system of postulates has not been much investigated, but certainly there are indications that the ones related with dimension are independent.

1.5.5 The Noncommutative Torus

This was the early paradigm for nc manifolds, where everything works smoothly. For a fixed irrational real number θ, let A_θ be the unital C^*-algebra generated by two elements u, v subject only to the relations $uu^* = u^*u = 1$, $vv^* = v^*v = 1$, and

$$vu = \lambda\, uv, \quad \text{where} \quad \lambda := e^{2\pi i\theta}. \tag{38}$$

Let $\mathcal{S}(\mathbb{Z}^2)$ denote the double sequences $\underline{a} = \{a_{rs}\}$ that are *rapidly decreasing* in the sense that

$$\sup_{r,s\in\mathbb{Z}} (1 + r^2 + s^2)^k\, |a_{rs}|^2 < \infty \quad \text{for all} \quad k \in \mathbb{N}.$$

The irrational rotation algebra or *noncommutative torus algebra* T_θ is defined as

$$\mathrm{T}_\theta := \left\{ a = \sum_{r,s} a_{rs}\, u^r v^s : \underline{a} \in \mathcal{S}(\mathbb{Z}^2) \right\}.$$

It is a *pre-C^*-algebra* that is dense in A_θ. The product and involution in T_θ are computable from (38):

$$ab = \sum_{r,s} a_{r-n,m}\, \lambda^{mn}\, b_{n,s-m}\, u^r v^s, \qquad a^* = \sum_{r,s} \lambda^{rs}\, \bar{a}_{-r,-s}\, u^r v^s.$$

The irrational rotation algebra gets its name from another representation, on $L^2(\mathrm{T})$: the multiplication operator U and the rotation operator V given by $(U\psi)(z) := z\psi(z)$ and $(V\psi)(z) := \psi(\lambda z)$ satisfy (38). In the C^*-algebraic framework, U generates the C^*-algebra $C(\mathrm{T})$ and conjugation by V gives an automorphism α of $C(\mathrm{T})$. Under such circumstances, the C^*-algebra generated by $C(\mathrm{T})$ and the unitary operator V is called the *crossed product* of $C(\mathrm{T})$ by the automorphism group $\{\alpha^n : n \in \mathbb{Z}\}$. In symbols

$$A_\theta \simeq C(\mathrm{T}) \times_\alpha \mathbb{Z}.$$

The corresponding action by the rotation angle $2\pi\theta$ on the circle is ergodic and minimal (all orbits are dense); it is known that the C^*-algebra A_θ is therefore simple.

Using the abstract presentation by (38), certain *isomorphisms* become evident. First of all, $\mathrm{T}_\theta \simeq \mathrm{T}_{\theta+n}$ for any $n \in \mathbb{Z}$, since λ is the same for both. Next, $\mathrm{T}_\theta \simeq \mathrm{T}_{-\theta}$ via the isomorphism determined by $u \mapsto v$, $v \mapsto u$. There are no more isomorphisms among the T_θ.

The linear functional $\tau_0 : \mathrm{T}_\theta \to \mathbb{C}$ given by $\tau_0(a) := a_{00}$ is positive definite since $\tau_0(a^*a) = \sum_{r,s} |a_{rs}|^2 > 0$ for $a \neq 0$; it satisfies $\tau_0(1) = 1$ and is a trace, since $\tau_0(ab) = \tau_0(ba)$. Also, it can be shown that τ_0 extends to a faithful continuous trace on the C^*-algebra A_θ, and, in fact, this normalized trace on A_θ is unique. The GNS representation space $\mathcal{H}_0 = L^2(\mathrm{T}_\theta, \tau_0)$ may be described as the completion of the vector space T_θ in the Hilbert norm $\|a\|_2 := \sqrt{\tau_0(a^*a)}$. Since τ_0 is faithful, the obvious map $\mathrm{T}_\theta \to \mathcal{H}_0$ is injective; to keep the bookkeeping straight, in this section we shall denote by $\underline{a}$ the image in $\mathcal{H}_0$ of $a \in \mathrm{T}_\theta$. The GNS representation of T_θ is just $\underline{b} \mapsto \underline{ab}$. The vector $\underline{1}$ is obviously cyclic and separating, and the Tomita involution is given by $J(\underline{a}) := \underline{a^*}$, thus $J = J^\dagger$. The commuting representation is then given by

$$b \mapsto J\pi(a^*)J^\dagger\, \underline{b} = J\, \underline{a^*b^*} = \underline{ba}.$$

To build a two-dimensional geometry, we need to have a $\mathbb{Z}_2$-graded Hilbert space on which there is an antilinear involution C that anticommutes with the grading and satisfies $C^2 = -1$. There is a simple device that solves all of these requirements: we simply double the GNS Hilbert space by taking $H := H_0 \oplus H_0$ and define

$$C := \begin{pmatrix} 0 & -J \\ J & 0 \end{pmatrix}.$$

In order to have a spectral triple, it remains to introduce the operator D. For D to be self-adjoint and anticommute with Γ, it must be of the form

$$D = -i \begin{pmatrix} 0 & \underline{\partial}^\dagger \\ \underline{\partial} & 0 \end{pmatrix},$$

for a suitable closed operator $\underline{\partial}$ on $L^2(\mathrm{T}_\theta, \tau_0)$. The order-one axiom, together with the regularity axiom and the finiteness property lead to $\partial, \partial^\dagger$ being derivations of T_θ. The reality condition $CDC^\dagger = D$ is equivalent to the condition that $J\,\underline{\partial}\,J = -\underline{\partial}^\dagger$ on $L^2(\mathrm{T}_\theta, \tau_0)$. Consider the derivations

$$\delta_1(a_{rs}\, u^r v^s) := 2\pi i r\, a_{rs}\, u^r v^s, \quad \delta_2(a_{rs}\, u^r v^s) := 2\pi i s\, a_{rs}\, u^r v^s.$$

For concreteness, take ∂ to be a linear combination of the basic derivations δ_1, δ_2. Apart from a scale factor, the most general such derivation is $\partial = \partial_\tau := \delta_1 + \tau\delta_2$ with $\tau \in \mathbb{C}$. In fact, real values of τ must be excluded. Now, D_τ^{-2} has discrete spectrum of eigenvalues $(4\pi^2)^{-1}|m + n\tau|^{-2}$, each with multiplicity 2. The Eisenstein series $\sum_{m,n\neq 0,0} \frac{1}{(m+n\tau)^2}$ diverges logarithmically, thereby establishing the two-dimensionality of the geometry. The orientation cycle is given by

$$\frac{1}{4\pi^2(\tau - \bar{\tau})}(v^{-1}u^{-1} \otimes u \otimes v - u^{-1}v^{-1} \otimes v \otimes u).$$

This makes sense only if $\tau - \bar{\tau} \neq 0$, i.e. $\tau \notin \mathbb{R}$. Thus $(\Im\tau)^{-1}$ is a scale factor in the metric determined by D_τ. (Note a difference with the commutative volume form: since $v^{-1}u^{-1} = \lambda\, u^{-1}v^{-1}$, there is also a phase factor $\lambda = e^{2\pi i\theta}$ in the orientation cycle.)

We conclude by indicating that the noncommutative torus can be regarded as well as a deformation, as it corresponds to the Moyal product of periodic functions. There are of course nc tori of all dimensions greater than 2.

1.5.6 The Noncompact Case

Real *noncompact spectral triples* (also called nonunital spectral triples) have implicitly been already defined. In practice the data are of the form

$$(\mathcal{A}, \widetilde{\mathcal{A}}, H, D; C, \Gamma),$$

where now $\mathcal{A}$ is a nonunital algebra and the new element $\widetilde{\mathcal{A}}$ is a preferred unitization of $\mathcal{A}$, acting on the same Hilbert space.

To get an idea of the difficulties involved in the choice of $\mathcal{A}$, consider the simplest commutative case, say of the manifold $\mathbb{R}^p$. Depending on the fall-off conditions deemed suitable, the smooth nonunital algebras that can represent the manifold are numerous as the stars in the sky. The problem is compounded in the noncommutative case, say when $\mathcal{A}$ is a deformation of an algebra of functions. To be on the safe side, one should take a relatively small algebra at the start of any investigation of examples.

Postulates 2, 4, and 6 need no changes with respect to the compact case formulation.

Now, we ponder:

- Dimension of the geometry: for p a positive integer $a(1 + D^2)^{-1/2}$ belongs to the generalized Schatten class $\mathcal{L}^{p,+}$ for each $a \in \mathcal{A}$, and, moreover, $\mathrm{Tr}_\omega(a(1 + |D|)^{-p})$ is finite and not identically zero.
- Finiteness: the algebra $\mathcal{A}$ and its preferred unitization $\widetilde{\mathcal{A}}$ are pre-C^*-algebras. There exists an ideal $\mathcal{A}_1$ of $\widetilde{\mathcal{A}}$, including $\mathcal{A}$, which is also a pre-C^*-algebra with the same C^*-completion as $\mathcal{A}$, such that the space of smooth vectors is an $\mathcal{A}_1$-pullback of a finitely generated projective $\widetilde{\mathcal{A}}$-module. Moreover, an $\mathcal{A}_1$-valued hermitian structure is defined on H_∞ with the noncommutative integral; this is an absolute continuity condition.
- Orientation: there is a *Hochschild p-cycle* $\mathbf{c}$ on $\widetilde{\mathcal{A}}$, with values in $\widetilde{\mathcal{A}} \otimes \widetilde{\mathcal{A}}^\circ$. Such a p-cycle is a finite sum of terms like $(a^0 \otimes b) \otimes a^1 \otimes \cdots \otimes a^p$, whose natural representative by operators on $\mathcal{H}$ is given by $\pi_D(\mathbf{c})$ in formula (37); the volume form $\pi_D(\mathbf{c})$ must solve the equation

$$\pi_D(\mathbf{c}) = \Gamma \quad \text{(even case)} \quad \text{or} \quad \pi_D(\mathbf{c}) = 1 \quad \text{(odd case)}.$$

The need for some preferred unitization is plain, as finiteness requires the presence both of a nonunital and a unital algebra. Then examples show the need for a further subtlety, to wit, the nonunital algebra for which summability works is *smaller* than the nonunital algebra required for finiteness. Also, orientation is defined directly on the preferred unitization.

The commutative examples were worked out in [74, 75]; there summability works in view of asymptotic spectral analysis for the Dirac operator. In [76] – to some surprise of Alain Connes – it was shown that Moyal algebras are noncompact spectral triples.

It is worthwhile to point out that the NCG versions of the Standard Model are noncompact spectral triples, too; while there is no end of algebraic intricacies for the finite dimensional representation [77] required to reproduce the quirks of particle physics, analytically the problem is to be tackled by the methods of the mentioned papers [74–76].

1.5.7 Nc Toric Manifolds (Compact and Noncompact)

How does one recover the metric geometry of the Riemann sphere $\mathbb{S}^2$ from spectral triple data? If $\mathcal{A}$ is a dense subalgebra of some C^*-algebra containing elements x, y, z and if the matrix

$$p = \tfrac{1}{2}\begin{pmatrix} 1+z & x+iy \\ x-iy & 1-z \end{pmatrix}$$

is a projector, it is easy to see from the projector relations that x, y, z commute and that $x^2 + y^2 + z^2 = 1$. Thus $A = C(X)$ where $X \subset \mathbb{S}^2$ is closed. The condition

$$\pi_D\Big(\operatorname{tr}\big(\big(p - \tfrac{1}{2}\big) \otimes p \otimes p\big)\Big) = \Gamma$$

can only hold if $X = \mathbb{S}^2$. In the same way, Connes sought to obtain the sphere $\mathbb{S}^4$ with its round metric by starting with an analogous projector in $M_4(\mathcal{A})$:

$$p = \begin{pmatrix} (1+z)1_2 & q \\ q & (1-z)1_2 \end{pmatrix},$$

with q the quaternion

$$q = \begin{pmatrix} a & b \\ -b^* & a \end{pmatrix},$$

imposing conditions so that

$$\pi_D\Big(\operatorname{tr}\big((p - \tfrac{1}{2}) \otimes p \otimes p \otimes p \otimes p\big)\Big) = \Gamma.$$

Again $\mathcal{A}$ is commutative and the 4-sphere relation holds. But then Landi surprised everyone by pointing out that one could substitute $-\lambda b^*$ for the entry $-b^*$. With $\lambda = e^{2\pi i\theta}$, this works into a spectral triple. It was called an *isospectral deformation* because the Dirac operator remains untouched [78].

Again, this generalizes into a θ-deformation of any Riemannian manifold M that admits T^2 as a subgroup of its group of isometries. And again, this is essentially a Moyal deformation: if $M = G/K$, with G compact of rank at least two, then $C^\infty(G)$ can be deformed in such a way that $C^\infty(M_\theta)$ is a homogenous space of the compact quantum group $C^\infty(G_\theta)$ [79].

The procedure can be generalized to a large family of noncompact Riemannian spin manifolds (with "bounded geometry") that admit an action of T^l, for $l \geq 2$, or a free action of $\mathbb{R}^l$, for $l \geq 2$ [80]. The upshot is more noncommutative spin geometries.

(Even lowly $\mathbb{S}^2$ hids surprises, too, if one allows for relaxing the notion of what a Dirac operator is [81].)

1.5.8 Closing Points

1.5.8.1 Fabricating nc Spaces: A Second Conceptual Star and Catalogue

So far, we have played it very safe, and we have said little on how to handle wilder examples of nc manifolds. Connes himself recommends the following steps [82]:

1. Given an algebra $\mathcal{A}$ (putative "of smooth functions on a nc manifold"), try first of finding a resolution of it as an $\mathcal{A}$-bimodule, with a view to compute its Hochschild cohomology and eventually its cyclic homology and cohomology. This is not an easy task in general; it has been performed in the commutative case and for foliations.
2. Many nc spaces arise as "bad quotients". Consider $Y := X/\sim$. If one tries to study

$$C(Y) = \{ f \in C(X) : f(a) = f(b),\ \forall a \sim b \},$$

one often ends up with only constant functions. (It is true that, for proper actions of Lie groups, even if M/G is not a manifold, there is, however, an interesting functional structure [83, 84] that can be usefully studied by a mixture of "commutative" and "noncommutative" methods.) It beckons to drop the commutativity requirement by considering complex functions of two variables defined on the graph of the equivalence relation. They will act as bounded operators on the Hilbert space of the equivalence class, and they multiply with the convolution product:

$$(fg)_{ab} = \sum_{a\sim c\sim b} f_{ac} g_{cb}. \tag{39}$$

Of course, when the quotient space is "nice", one can do that, too; as a rule in this case, the commutative and noncommutative algebras are Morita equivalent. But in a case as simple as $X = [0, 1] \times \mathbb{Z}_2$ with $\sim$ given by $(x, +) \sim (x, -)$ for $x \in]0, 1[$, we obtain for the convolution algebra the "dumbbell" algebra:

$$\{ f \in C([0, 1]) \otimes M_2\mathbb{C} : f(0), f(1) \text{ diagonal} \},$$

and there is no such equivalence. The idea is then to compute the K-theory, in order to learn as much as possible on the space. Ideally, one would also like to have "vector bundles", Chern character (using connections and curvature), and even moduli spaces for Yang-Mills connections – this works wonderfully for nc tori, which after all are quotient spaces.

Incidentally, families of maps that are semigroups in the commutative word naturally become C^*-bialgebras in the noncommutative context. We may refer to the recent beautiful paper by Soltan [85], where the quantum family of maps from $\mathbb{C}^2$ to $\mathbb{C}^2$ is identified to the dumbbell algebra.

Let us add as well that Connes contends that the foundational step of Quantum Mechanics (by Heisenberg in 1925) amounts to replacing an abelian group law by a groupoid law like (39), in order to make sense of the combination principles of spectral lines.

3. Then come the spectral triples. They respond for K-homology classes, smooth structure, and metric. There is a surprisingly vast class of spaces that can be described in this way, under conditions in general less strict than the ones required for the reconstruction theorem.
4. The time evolution and thermodynamic aspects.

That said, we can prepare our catalogue (leaving aside subjects related to physics, for a moment):

- Spaces of leaves of foliations. This was an early, successful application of nc geometry. By elaborating on the construction of point 2 above, Connes was able to apply methods of operator theory to foliation theory.
- Tilings (periodic and aperiodic). Also under point 2.
- Dynamical systems. Also point 2.
- Cantor sets and fractals. One can associate spectral triples (Dirac operators) to them! The algebra of continuous functions on a Cantor set is AF commutative. We omit the details on the construction of (H, D). It is then very interesting to investigate the dimension spectrum of the spectral triple. For the classical middle-third Cantor set:

$$\mathrm{Tr}(|D|^{-s}) = 2\sum_k l_k^s = \sum_{k\geq 1} 2^k 3^{-sk} = \frac{2\,3^{-s}}{1 - 2\,3^{-s}},$$

given that $l_k = 3^{-k}$ with multiplicities 2^{k-1}. This yields as dimension spectrum

$$\frac{\log 2}{\log 3} + \frac{2\pi i n}{\log 3},$$

for $n \in \mathbb{Z}$. For compact fractal subsets of $\mathbb{R}^n$. Christensen and Ivan recently have constructed spectral triples not satisfying Weyl's asymptotic formula – there is no constant c so that the number of eigenvalues $N(\Lambda)$ bounded by Λ fulfils

$$N(\Lambda) - c\Lambda \sim \text{ lower order in } \Lambda.$$

- Algebraic deformations. Of this the Moyal-like spaces are the outstanding example. More on that below.
- Spherical manifolds which are not isospectral deformations. I refer to [86] and subsequent papers by Connes and Dubois-Violette.
- Nc spaces related to arithmetic problems (including some that have been used by Connes to try to prove the Riemann hypothesis). On this I claim zero expertise.

1.5.8.2 What About Physics?

- Quantum Hall effect, related to nc tori. This is due to Bellissard.
- Nc spaces from axiomatic QFT. For instance, the local algebras in a supersymmetric model, together with the supercharge as a Dirac operator, constitute a spectral triple.
- Nc spaces from renormalization, via dimensional regularization. This has been only hinted at.
- The mentioned Standard Model reconstruction from NCG.
- Nc spaces from strings. If one goes to the physics archives and asks for "noncommutative geometry" or "noncommutative field theory", what one finds is something as puzzling as particular, that is, perturbative quantum field theory over Moyal hyperplanes. This was popularized by Seiberg and Witten [87] as a certain limit of string theory, but has acquired a life of its own. Nevertheless [76] and subsequent papers [88, 89] tried to make a bridge between this and Connes' paradigm.

1.5.8.3 Some Neglected Tools

- Lie algebroids, Lie – Rinehart algebras, and the like. It is a little mystery why, while groupoids play a central role in NCG, their infinitesimal version does not seem to play any role. All the more so because the algebraic version of Lie algebroids, the theory of Lie – Rinehart(– Gerstenhaber) algebras, which seems to be the good framework for BRS theory, has very much the flavour of NCG and is quite able to deal with many singular spaces [90].
  Lie – Rinehart algebras are usually commutative, but some of the results pertaining to them can be extended to "softly noncommutative" cases. Most importantly, the theory of Adams operations, that plays such an important role in the Hochschild and cyclic cohomology of commutative algebras, can be extended to the realm of noncommutative spaces [91]. This connects the local index formula by Connes and Moscovici [92] with combinatorial aspects (the Dynkin operator of free Lie algebra theory and noncommutative symmetric functions) that have not been fully explored.
- Rota – Baxter operators and skewderivations. A poor man's path to the nc world (akin to the one taken by some quantum group theorists) is to try to generalize the usual derivative/integral pair. This is elementary stuff with many ramifications. A skewderivation of weight $\theta \in \mathbb{R}$ is a linear map $\delta : A \to A$ fulfilling the condition

$$\delta(ab) = a\delta(b) + \delta(a)b - \theta\delta(a)\delta(b). \tag{40}$$

We may call skewdifferential algebra a double $(A, \delta; \theta)$ consisting of an algebra A and a skewderivation δ of weight θ. A *Rota – Baxter map* R of weight $\theta \in \mathbb{R}$ on a not necessarily associative algebra A, commutative or not, is a linear map

$R : A \to A$ fulfilling the condition

$$R(a)R(b) = R(R(a)b) + R(aR(b)) - \theta R(ab), \qquad a, b \in A. \tag{41}$$

When $\theta = 0$ we obtain the integration-by-parts rule. The triple $(A, \delta, R; \theta)$ will denote an algebra A endowed with a skewderivation δ and a corresponding Rota – Baxter map R, both of weight θ, such that $R\delta a = a$ for any $a \in A$ such that $\delta a \neq 0$, as well as $\delta R a = a$ for any $a \in A$, $Ra \neq 0$. We can check consistency of conditions (40) and (41) imposed on δ, R:

$$\begin{aligned}
\theta\delta R(ab) &= R(a)b + aR(b) - \delta(R(a)R(b)) \\
&= R(a)b + aR(b) - R(a)b - aR(b) + \theta ab = \theta ab, \\
R\delta(ab) &= R(a\delta(b)) + R(\delta(a)b) - \theta R(\delta(a)\delta(b)) = R(a\delta(b)) + R(\delta(a)b) \\
&\quad - R(a\delta(b)) - R(\delta(a)b) + ab = ab.
\end{aligned}$$

Rota – Baxter operators have proved their worth in probability theory and combinatorics, and in the Connes – Kreimer approach to renormalization, but their range of applications is much wider.

- What is the natural noncommutative algebra structure that one should impose on ordinary, well-behaved manifolds? The author has long contended that the answer, at least in the equivariant case, is general Moyal theory. Given the naturalness of ordinary Moyal quantization on hyperplanes, the high number of nc spaces that turn out to be related to Moyal quantization, plus the usefulness of Moyal quantization in proofs (for instance of Bott periodicity in the algebraic context), it is surprising that few nc geometers seem interested in general Moyal theory.

  But how to define the latter? It would run as follows. Let X be a phase space, μ a Liouville measure on X, and H the Hilbert space associated to (X, μ). A Moyal or Stratonovich – Weyl quantizer for (X, μ, H) is a mapping Ω of X into the space of self-adjoint operators on H, such that $\Omega(X)$ is weakly dense in $B(H)$, and verifying

$$\begin{aligned}
\mathrm{Tr}\Omega(u) &= 1, \\
\mathrm{Tr}\big[\Omega(u)\Omega(v)\big] &= \delta(u - v),
\end{aligned}$$

  in the distributional sense. (Here $\delta(u - v)$ denotes the reproducing kernel for the measure μ.) Moyal quantizers, if they exist, are unique, and ownership of a Moyal quantizer solves in principle all quantization problems: *quantization* of a (sufficiently regular) function or “symbol” a on X is effected by

$$a \mapsto \int_X a(u)\Omega(u)\, d\mu(u) =: Q(a),$$

and *dequantization* of an operator $A \in B(H)$ is achieved by

$$A \mapsto \mathrm{TrA}\Omega(\cdot) =: W_A(\cdot).$$

Indeed, it follows that $1_H \mapsto 1$ by dequantization, and also

$$\mathrm{TrQ}(a) = \int_{\mathrm{X}} a(u)\, d\mu(u).$$

Moreover, using the weak density of $\Omega(X)$, it is clear that

$$W_{Q(a)}(u) = \mathrm{Tr}\Big[\Big(\int_{\mathrm{X}} a(\mathrm{v})\Omega(v)\, d\mu(v)\Big)\Omega(u)\Big] = a(u),$$

so Q and W are inverses. In particular, $W_{Q(1)} = 1$ says that $1 \mapsto 1_H$ by quantization, and this amounts to the reproducing property $\int_X \Omega(u)\, d\mu(u) = 1_H$. Finally, we also have

$$\mathrm{Tr}[Q(a)Q(b)] = \int_{\mathrm{X}} a(u)b(u)\, d\mu(u).$$

This is the key property. Most interesting cases occur in an equivariant context; that is to say, there is a (Lie) group G for which X is a symplectic homogeneous G-space, with μ then being a G invariant measure on X, and G acts by a projective unitary irreducible representation U on the Hilbert space H. A Moyal quantizer for the combo (X, μ, H, G, U) is a map Ω taking X to self-adjoint operators on $\mathcal{H}$ that satisfies the previous defining equations and the equivariance property

$$U(g)\Omega(u)U(g)^{-1} = \Omega(g \cdot u), \quad \text{for all} \quad g \in G,\ u \in X.$$

The question is How to find the quantizers? The fact that the solution in flat spaces leads to (bounded) parity operators points out to the framework of *symmetric spaces* as the natural one to find Moyal quantizers by interpolation. This heuristic parity rule was found to work for orbits of the Poincaré group [93]. Noncompact symmetric spaces should provide a wealth of noncompact spectral triples (the compact case is somewhat pathological). Recently the author, together with V. Gayral and J.C. Várilly, has given the Moyal quantization of the surface of constant negative curvature [94]; a new special function plays there the main role in framing a subtler version of the parity rule.

- Algebraic K-theory, noncommutative geometry and field theory. The role of the two first functors of algebraic K-theory in QFT with external fields is "well known"; Connes has dabbled on this, but he has not pursued the subject. To this writer, also in relation with [92], it seems extremely promising.

1.5.9 Some Interfaces with Quantum Gravity

This section is intended as a taunt. We just lift a corner of the veil.

1.5.9.1 Noncommutative Field Theory and Quantum Gravity

Direct connection between noncommutative field theory and quantum gravity has been sought in several papers. The basic idea is due to Rivelles [95]. In noncommutative *gauge* theories, translations are equivalent to gauge transformations. This at once reminds one of gravitation (the case can be made that translations necessarily involve gauge transformations in Yang – Mills theories, too [96], but this is a weaker statement). In general, the distinction between internal and geometrical degrees of freedom fades in noncommutative geometry [97]. Indeed in [95] it is shown, using Seiberg – Witten maps [87], how the field action can be regarded as a coupling to a gravitational background. The idea has been further developed in [98]. In some other papers suggesting a noncommutative geometry formalism for pure classical gravity, the apparatus is so heavy as to make it difficult to see the forest for the trees [99]. A different approach is to look for noncommutative corrections to particular classes of spacetimes. This is found in [100]. The “barriers to entry” in this field being relatively modest, we cut our remarks short.

1.5.9.2 Isospectral Deformations and Unimodularity

There seems to be no good reason to exclude noncommutative manifolds in the sense of Connes from the approaches to quantum gravity based on “sum over geometries”. Already, in an important paper [101], Yang has showed that the Eguchi and Hanson gravitational instantons [102] give rise by isospectral deformation to noncommutative noncompact manifolds in the sense of [76]. Now, isospectral deformation leaves the orientation condition unchanged. The general paradigm is as follows: *any* Dirac operator, describing a K-homology class, corresponding to a commutative manifold (thus, for any Riemannian geometry over it) or noncommutative one, solves equally well, and on the same footing, the “topological equation” that defines the manifold itself. With the proviso that the volume form remains the same. The advantages indicated in [57] should apply in this context, too.

The punch line: in its present form at least, noncommutative geometry favours the unimodular theory.

1.6 More on the “Cosmological Constant Problem” and the Astroparticle Interface

Notice that both terminologies “cosmological constant” and “dark energy” betray theoretical prejudices.

The first name that we can deal with the observations pointing to an acceleration of the expansion rate of the universe by just including the so-called cosmological term in the Einstein equations. In fact, we do not know the equation of state, not to speak of the evolution laws, of whatever exotic "substance" that might be involved [103].

The second is related to the belief that the acceleration be caused by fluctuations, or "zero-point energies" of the quantum vacuum, somehow. Alas, this notion here was entertained by nobody less than Weinberg, whose already mentioned [56] threw both light and obscurity on the subject.

The whole review hangs on the thread that there must be a problem, since

> the energy density of the vacuum acts just like a cosmological constant.

However, the effective cosmological constant is quite small (we would not be here otherwise). On the face of it, zero-point energies are infinite (well, this is not the case in Epstein – Glaser renormalization, but let us go with the argument). If we take as a sensible cut-off the Planck scale, the amount of "fine-tuning" necessary to cancel their contribution is mind boggling. Thus,

> Perhaps surprisingly, it was a long time before particle physicists began seriously to worry about this problem, despite the demonstration in the Casimir effect of the reality of zero-point energies.

The trouble is, that "demonstration" is another urban legend. The negative weight of zero-point fluctuations is *unobserved* in any laboratory experiment, *including the Casimir effect*. The latter is measured nowadays well enough. However, the usual derivation in terms of differences of zero-point energies, and its neat result, where only c, $\hbar$, and the geometry of the plates enter, inviting us to think of it as a "property of the vacuum", is misleading. The point has been made recently by Jaffe [104]. In truth, the Casimir effect distinguishes itself from other quantum electrodynamics only in that (for some geometrical configurations, not for all) it reaches a finite limit as the fine structure constant $\alpha \uparrow \infty$; this limit is the usually quoted result. In that derivation, the plates are treated as perfect conductors. However, a perfect conductor at all frequencies is a physical impossibility. The plasma frequency

$$\omega_{\mathrm{pl}} = 2e\sqrt{\frac{\pi n}{m}}$$

indicates the frequency above which the conductivity goes to zero; here n is the density of conduction band electrons and m their effective mass. So the perfect conductor approximation is good if $c/d \ll \omega_{\mathrm{pl}}$, with d the distance between plates, that is, for materials and plate distances such that

$$\frac{1}{137} \sim \alpha \gg \frac{mc}{4\pi \hbar n d^2}.$$

Still, it remains an approximation. Casimir forces can be and have been calculated without reference to the vacuum. Whether there can be experimental evidence for zero-point energies, apart from gravity, is an open question, which may be answered in the negative for all we know. The lesson is that their putative contribution to the cosmological constant must be in doubt. As Jaffe puts it [104]

> Caution is in order when an effect, for which there is no direct experimental evidence, is the source of a huge discrepancy between theory and experiment.

Indeed.

We might add nowadays there is a "vacuum fluctuations" branch of mathematics, conductors which are always perfectly so and plates of vanishing thickness *etsi daretur*. This is to the good, and may be helpful, provided we keep the origins in mind and do not start to draw unwarranted physical inferences! We are reminded of Manin's dicta. A mathematically rigorous and physically sound account of the Casimir effect without invoking "zero-point energies", particularly unveiling the unphysical nature of Dirichlet boundary conditions, has been given by Herdegen [105].

Parenthetically, one finds in the work of Vachapasti and coworkers on "black stars" mentioned in the first section [7] a commendable retreat to consideration of *physical* black holes – collapsing bodies suspended above their Schwarzschild radius forever from a remote observer viewpoint – rather than mathematical black holes – vacuum solutions of the general relativity equations. While the mathematical study of black holes remains a useful and fascinating subject, the former is required to explain astrophysical observations.

On the other hand, it is hard to dispute that the energy density of the vacuum itself should act like a cosmological constant. Thus it is rather less clear why the flavourdynamics scale – whereby we are talking not of phantom fluctuations, but of the vacuum expected value of the energy itself – does not play a role. Even if one (as this writer) does not trust the Higgs mechanism, there is reason to worry about the contribution of chiral symmetry breaking in the quark condensate, still 12 orders of magnitude above the "observed" range for the cosmological constant. For this reason unimodularity as discussed in Sect. 1.4 should be taken seriously.

A recommended review on the cosmological constant is [106]. Its author dismisses "fine-tuning" out of hand. Suggestive thinking on the dark energy problem is found in [107].

We cannot conclude without mentioning the "LHC connection". After all, fundamental scalar fields, hitherto unseen, are assumedly involved in inflation, dark energy, and other cosmological scenarios. It is widely believed that the Higgs particle will be observed after the few first stages of the LHC's proper operation.

Some scepticism is also warranted on that. The reason is that "minimality" of the scalar sector of the Standard Model of particle physics is just a theoretical prejudice. This has been particularly emphasized by Strassler [108]. (Yes, *entes non sunt multiplicanda praeter necessitatem*. But Nature does not care for Ockham's razor: who ordered the muon?)

There is the distinct possibility that something was overlooked at LEP and that the Higgs sector be considerably more complicated that in standard lore. Tension has been growing for a while between precision results and direct Higgs searches. The basic trouble was laid down by Chanowitz a few years ago [109]: if one eliminates from the precision electroweak data the (outlier) value of the forward – backward asymmetry into b-quarks, then the expected value for the Higgs mass drops to less than 50 GeV or so, with the mentioned outlier attributable to new physics. Otherwise, the overall fit is poor. This leads us to take *cum grano salis* the exclusion results at LEP. For instance, mixing with "hidden world" scalars leads to reduction to the standard Higgs couplings (consult [110] and references therein), in particular the ZZh coupling, and this could not be, and was not, ruled out by LEP for those relatively low energies. Other Higgs sector scenarios shielding the Higgs particle from detection have been discussed in [111, 112].

Recent experiment has made the situation even murkier: on Halloween night of 2008, ghostly (albeit rather abundant) multi-muon events at Fermilab were reported by (a majority segment of) the CDF collaboration [113]. A possible explanation for them invokes "new" light Higgs-like particles coupling relatively strongly to the "old" ones and much less so to the SM fermions and MVB [114, 115]. There is also the possibility that the visible Higgs boson be rather *heavier* than expected, the discrepancy with the precision results being (somewhat brazenly) attributed to new physics [116, 117]. Then the *inert* Higgs boson would be a prime candidate for dark matter.

Acknowledgements Help of J.C. Várilly in preparing these notes is most gratefully acknowledged. Thanks are due as well to Michael Dütsch, who helped me with fine points of the argument in Sect. 1.3.7. I am also indebted to the referee, whose detailed comments and remarks helped to improve the final version.

References

1. S.L. Glashow, *Interactions*, Warner Books, New York, (1988).
2. C. Kiefer, *Quantum Gravity*, Oxford University Press, Oxford, (2007).
3. D.L. Wiltshire, *Exact solution to the averaging problem in cosmology*. Phys. Rev. Lett. **99** (2007) 251101.
4. D.L. Wiltshire, Dark energy without dark energy, astro-ph/0712.3984.
5. B.M. Leith, S.C. Cindy Ng, D.L. Wiltshire, *Gravitational energy as dark energy: concordance of cosmological tests*. Astrophys. J. **672** (2008) L91.
6. D.L. Wiltshire, *Cosmological equivalence principle and the weak field limit*. Phys. Rev. D **78** (2008) 084032.
7. T. Vachaspati, D. Stojkovic, L.M. Krauss, *Observation of incipient black holes and the information loss problem*. Phys. Rev. D **76** (2007) 024005.
8. J.D. Anderson, P.A. Laing, E.L. Lau, A.S. Liu, M.M. Nieto, S.G. Turyshev, *Study of the anomalous acceleration of Pioneer 10 and 11*. Phys. Rev. D **65** (2002) 082004.
9. M.M. Nieto, J.D. Anderson, *Search for a solution of the Pioneer anomaly*. Contemp. Phys. **48** (2007) 41.
10. M.M. Nieto, *New horizons and the onset of the Pioneer anomaly*. Phys. Lett. B **659** (2008) 483.
11. Q. Exirifard, *Constraints on $f\left(R_{ijkl}^{ijkl}\right)$ gravity: an evidence against the covariant resolution of the Pioneer anomaly*. Class. Quant. Grav. **26** (2009) 025001.

12. A.F. Rañada, A. Tiemblo, *Time, clocks and parametric invariance*. Found. Phys. **38** (2008) 458.
13. A.F. Rañada, A. Tiemblo, The Pioneer anomaly as a quantum cosmological effect, gr-qc/0804.1904.
14. J.D. Anderson, J.K. Campbell, J.E. Ekelund, J. Ellis, J.F. Jordan, *Anomalous orbital-energy changes observed during spacecraft flybys of Earth*. Phys. Rev. Lett. **100** (2008) 091102.
15. M. Milgrom, *A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis*. Astrophys. J. **270** (1983) 365.
16. L. Smolin, *The Trouble with Physics*, Houghton Mifflin, New York, (2006).
17. T. Van Flandern, *The speed of gravity —what the experiments say*. Phys. Lett. A **250** (1998) 1.
18. R. Colella, A.W. Overhauser, S.A. Werner, *Observation of gravitationally induced quantum interference*. Phys. Rev. Lett. **34** (1975) 1472.
19. J.L. Anderson, *Principles of Relativity Physics*, Academic Press, New York, (1967).
20. D.M. Greenberger, *The neutron interferometer as a device for illustrating the strange behavior of quantum systems*. Rev. Mod. Phys. **55** (1983) 875.
21. D.M. Greenberger, A.W. Overhauser, *Coherence effects in neutron diffraction and gravity experiments*. Rev. Mod. Phys. **51** (1979) 43.
22. U. Bonse, T. Wroblewski, *Measurement of neutron quantum interference in noninertial frames*. Phys. Rev. Lett. **51** (1983) 1401.
23. M. Springborg, J.P. Dahl, *Wigner's phase-space function and atomic structure. II. Ground states for closed shell atoms*. Phys. Rev. A **36** (1987) 1050.
24. Ph. Blanchard, José M. Gracia-Bondía, *Density functional theory on phase space*. 2010, in preparation.
25. S. Doplicher, K. Fredenhagen, J.E. Roberts, *The quantum structure of spacetime at the Planck scale and quantum fields*. Commun. Math. Phys. **172** (1995) 187.
26. V.V. Nesvizhevsky, A.K. Petukhov, H.G. Börner, T.A. Baranova, A.M. Gagarski, G.A. Petrov, K.V. Protasov, A. Yu. Voronin, S. Baeßler, H. Abele, A. Westphal and L. Lucovac, *Study of the neutron quantum states in the gravity field*. Eur. Phys. J. C **40** (2005) 479.
27. A. Herdegen, J. Wawrzycki, *Is Einstein equivalence principle valid for a quantum particle?*. Phys. Rev. D **66** (2002) 044007.
28. M. Kreuz, V.V. Nesvizhevsky, P. Schmidt-Wellenburg, T. Soldner, M. Thomas, H.G. Börner, F. Naraghi, G. Pignol, K.V. Protasov, D. Rebreyend, F. Vezzu, R. Flaminio, C. Michel, L. Pinard, A. Remillieux, S. Baessler, A.M. Gagarski, L.A. Grigorieva, T.M. Kuzmina, A.E. Meyerovich, L.P. Mezhov-Deglin, G.A. Petrov, A.V. Strelkov, A. Yu. Voronin, A method to measure the resonance transitions between the gravitationally bound quantum states of neutrons in the GRANIT spectrometer. physics.ins-det/0902.0156.
29. C. Rovelli, *Quantum Gravity*, Cambridge University Press, Cambridge, (2004).
30. E.P. Wigner, *Relativistic invariance and quantum phenomena*, Rev. Mod. Phys. **29** (1957) 255.
31. B. Booß-Bavnbek, G. Esposito, M. Lesch, Quantum gravity: unification of principles and interactions, and promises of spectral geometry, hep-th/0708.1705.
32. R.P. Feynman, F.B. Morinigo, W.G. Wagner, *Feynman Lectures on Gravitation*, Addison-Wesley, Reading, MA, (1995).
33. T. Ortín, *Gravity and strings*, Cambridge University Press, Cambridge, (2004).
34. S. Weinberg, *Photons and gravitons in $\mathbb{S}$-matrix theory: derivation of charge conservation and equality of gravitational and inertial mass*. Phys. Rev. **135 B** (1964) 1049.
35. G. Scharf, *Quantum Gauge Theories: A True Ghost Story*, Wiley, New York, (2001).
36. R.P. Feynman, *Quantum theory of gravity*. Acta Phys. Polon. **24** (1963) 697.
37. G. Scharf, M. Wellmann, *Spin-2 gauge theories and perturbative gauge invariance*. Gen. Rel. Grav. **33** (2001) 553.
38. M. Dütsch, *Proof of perturbative gauge invariance for tree diagramas to all orders*. Ann. Phys. (Leipzig) **14** (2005) 438.

39. T. Kugo, I. Ojima, *Subsidiary conditions and physical $\mathbb{S}$-matrix unitarity in indefinite-metric quantum gravitation theory*. Nucl. Phys. B **144** (1978) 234.
40. J. M. Gracia-Bondía, "BRS invariance for massive boson fields", to appear in the proceedings of the Summer School "Geometrical and topological methods for quantum field theory", Cambridge University Press, 2009; hep-th/0808.2853.
41. M. Fierz, W. Pauli, *On relativistic wave equations for particles of arbitrary spin in an electromagnetic field*, Proc. Roy. Soc. (London) A **173** (1939) 211.
42. S. Gómez-Avila, M. Napsuciale, J.A. Nieto, M. Kirchbach, "High integer spins beyond the Fierz – Pauli framework", Guanajuato preprint, (2005).
43. N. Berline, E. Getzler, M. Vergne, *Heat kernels and Dirac operators*, Springer, Berlin, (1996).
44. L.D. Landau, E.M. Lifshitz, *The classical theory of fields*, Pergamon Press, Oxford, (1971).
45. N. Grillo, *Scalar matter coupled to quantum gravity in the causal approach: finite one-loop calculations and perturbative gauge invariance*. Ann. Phys. (New York) **287** (2001) 153.
46. E. Alvarez, *Quantum gravity: an introduction to some recent results*. Rev. Mod. Phys. **61** (1989) 561.
47. D.R. Grigore, G. Scharf, *Massive gravity as a quantum gauge theory*. Gen. Rel. Grav. **37** (2005) 1075.
48. M.U. Khafizov, *A quasi-invariant smooth measure on the diffeomorphism group of a domain*, Math. Notes (Matematicheskie Zametkie) **48** (1990) 134.
49. E. Alvarez, "Some general problems of quantum gravity", in: J. Pérez-Mercader, J. Solà, E. Verdaguer (eds.) *Quantum Gravity and Cosmology*, World Scientific, Singapore, (1992).
50. J. Ambjørn, R. Loll, *Non-perturbative Lorentzian quantum gravity, causality and topology change*. Nucl. Phys. B **536** (1998) 407.
51. J. Ambjørn, R. Janik, W. Westra, S. Zohren, The emergence of background geometry from quantum fluctuations, gr-qc/0607013.
52. O. Lauscher, M. Reuter, Asymptotic safety in quantum Einstein gravity: nonperturbative renormalizability and fractal spacetime structure, hep-th/0511260.
53. A. Connes, *Noncommutative Geometry and the Standard Model with neutrino mixing*. J. High Energy Phys. **0611** (2006) 081.
54. D. Kreimer, "Not so non-renormalizable gravity", in: B. Fauser J. Tolksdorf, E. Zeidler, (eds.) *Quantum Field Theory. Competitive Models*, Birkäuser, Basel, (2009).
55. M. Rowan-Robinson, *Cosmology*, Oxford University Press, Oxford, (2004).
56. S. Weinberg, *The cosmological constant problem*. Rev. Mod. Phys. **61** (1989) 1.
57. J.J. van der Bij, H. van Dam, Y.J. Ng, *The exchange of massless spin-two particles*. Physica A **116** (1982) 307.
58. E. Alvarez, A.F. Faedo, *A comment on the matter-graviton coupling*. Phys. Rev. D **76** (2007) 124016.
59. E. Alvarez, *Can one tell Einstein's unimodular theory from Einstein's general relativity?* J. High Energy Phys. **0503** (2005) 002.
60. A. Connes, C^*-algebras and differential geometry. C.R. Acad Sci. Paris A **290** (1980) 599.
61. P. Woit, *Not even wrong*, Jonathan Cape, London<??> (2006).
62. H.B. Lawson, jr., M.-L. Michelsohn, *Spin geometry*, Princeton University Press, Princeton University, New Jersey, (1989).
63. J.M. Gracia-Bondía, J.C. Várilly, H. Figueroa, *Elements of Noncommutative Geometry*, Birkhäuser, Boston, (2001).
64. R.J. Plymen, *Strong Morita equivalence, spinors and symplectic spinors*. J. Oper. Theory **16** (1986) 305.
65. G. Karrer, *Einführung von Spinoren auf Riemannschen Mannigfaltigkeiten*. Ann. Acad. Sci. Fennicae Ser. A I Math. **336/5** (1963) 3.
66. A. Trautman, *Connections and the Dirac operator on spinor bundles*. J. Geom. Phys. **58** (2008) 238.

67. A. Rennie, J.C. Várilly, *Reconstruction of manifolds in noncommutative geometry*. math.OA/0610418.
68. A. Connes, course given at the Collège de France, January–March 2008.
69. A. Connes, On the spectral characterization of manifolds, math.OA/0810.2088.
70. R.S. Hamilton, *The inverse function theorem of Nash and Moser*, Bull. Amer. Math. Soc. **7** (1982), 65.
71. D. Voiculescu, *Some results on norm-ideal perturbations of Hilbert-space operators, I & II*, J. Oper. Theory **2** (1979) 3.
72. D. Voiculescu J. Oper. Theory **5** (1981) 77.
73. A. Connes, *Gravity coupled with matter and foundation of non-commutative geometry*. Commun. Math. Phys. **182** (1996) 155.
74. A. Rennie, *Smoothness and locality for nonunital spectral triples*. K-Theory **28** (2003) 127.
75. A. Rennie, *Summability for nonunital spectral triples*. K-Theory **31** (2004) 71.
76. V. Gayral, J.M. Gracia-Bondía, B. Iochum, J. C. Várilly, *Moyal planes are spectral triples*. Commun. Math. Phys. **246** (2004) 569.
77. A. Chamseddine, A. Connes, M. Marcolli, *Gravity and the Standard Model with neutrino mixing*. Adv. Theor. Math. Phys. **11** (2007) 991.
78. A. Connes, G. Landi, *Noncommutative manifolds, the instanton algebra and isospectral deformations*. Commun. Math. Phys. **221** (2001) 141.
79. J.C. Várilly, *Quantum symmetry groups of noncommutative spheres*. Commun. Math. Phys. **221** (2001) 511.
80. V. Gayral, B. Iochum, J.C. Várilly, *Dixmier traces on noncompact isospectral deformations*, J. Funct. Anal. **237** (2006) 507.
81. A. Sitarz, *Quasi-Dirac operators and quasi-fermions*, math-ph/0602030.
82. A. Connes, M. Marcolli, A walk in the noncommutative garden, math.OA/060154.
83. G. W. Schwarz, *Smooth functions invariant under the action of a compact Lie group*, Topology **14** (1975) 63.
84. P.W. Michor, Isometric actions of Lie groups and invariants, lecture notes, Vienna, 1997.
85. P.M. Soltan, Quantum families of maps and quantum semigroups on finite quantum spaces, math.OA/0610922.
86. A. Connes, M. Dubois-Violette, *Noncommutative finite-dimensional manifolds. I. Spherical manifolds and related examples*. Commun. Math. Phys. **230** (2002) 539.
87. N. Seiberg, E. Witten, *String theory and Noncommutative Geometry*, J. High Energy Phys. **9909** (1999) 032.
88. V. Gayral, *Heat-Kernel approach to UV/IR mixing on isospectral deformation manifolds*. Ann. Henri Poincaré **6** (2005) 991.
89. V. Gayral, J.M. Gracia-Bondía, F. Ruiz Ruiz, *Position-dependent noncommutative products: classical construction and field theory*. Nucl. Phys. **B727** (2005) 513.
90. J. Huebschmann, *Poisson cohomology and quantization*. J. reine angew. Math. **408** (1990) 57.
91. F. Patras, "Lambda-rings", in: M. Hazewinkel, (ed.), *Handbook of Algebra*, vol. 3, Elsevier, Dordrecht, (2003).
92. N. Higson, "The residue index theorem of Connes and Moscovici", in: N. Higson and J. Roe, (eds.), *Surveys in Noncommutative Geometry*, vol. 6, American Mathematical Society, Providence, RI, 2006.
93. J.F. Cariñena, J.M. Gracia-Bondía, J.C. Várilly, *Relativistic quantum kinematics in the Moyal representation*. J. Phys. A **23** (1990) 901.
94. V. Gayral, J.M. Gracia-Bondía, J. C. Várilly, *Fourier analysis on the affine group, quantization and noncompact Connes geometries*. J. Noncommutative Geom. **2** (2008) 215.
95. V.O. Rivelles, *Noncommutative field theories and gravity*. Phys. Lett. B **558** (2003) 191.
96. R. Jackiw, N. S. Manton, *Symmetries and conservation laws in gauge theories*. Ann. Phys. **127** (1980) 257.
97. F. Lizzi, R. Szabo, A. Zampini, *Geometry of the gauge algebra in noncommutative Yang – Mills theory*. J. High Energy Phys. **0108** (2001) 032.

98. H. Steinacker, *Emergent gravity from noncommutative gauge theory*. J. High Energy Phys. **0712** (2007) 049.
99. P. Aschieri, M. Dimitrijevic, F. Meyer, J. Wess, Noncommutative Geometry and gravity. Class. Quant. Grav. **23** (2006) 1883.
100. S. Marculescu, F. Ruiz Ruiz, *Noncommutative Einstein-Maxwell pp-waves*. Phys. Rev. D **74** (2006) 105004.
101. C. Yang, Isospectral deformations of Eguchi – Hanson spaces as nonunital spectral triples, math.OA/0804.2114.
102. T. Eguchi, A.J. Hanson, *Asymptotically at self-dual solutions to euclidean gravity*, Phys. Lett. B **74** (1978) 249.
103. R. Miquel, *Dark energy: an observational primer*. Acta Phys. Polon. B **39** (2008) 2765.
104. R.L. Jaffe, *The Casimir effect and the quantum vacuum*. Phys. Rev. D **72** (2005) 021301.
105. A. Herdegen, *No-nonsense Casimir force*. Acta Phys. Polon. B **32** (2001) 55.
106. S. Nobbenhuis, *Categorizing different approaches to the cosmological constant problem*. Found. Phys. **36** (2006) 613.
107. T. Padmanabhan, Emergent gravity and dark energy, gr-qc/0802.1798.
108. M.J. Strassler, New signatures and challenges for the LHC, in Proceedings of the 38th International Symposium on Multiparticle Dynamics, hep-ph/0902.0377.
109. M.S. Chanowitz, *$Z \to b\bar{b}$ decay asymmetry: lose-lose for the Standard Model*. Phys. Rev. Lett. **87** (2001) 231802.
110. S. Gopalakrishna, S. Jung, J.D. Wells, *Higgs boson decays to four fermions through an abelian hidden sector*. Phys. Rev. D **78** (2008) 055002.
111. J.J. Van der Bij, No Higgs at the LHC, hep-ph/0804.3534.
112. R. Jora, S. Moussa, S. Nasri, J. Schechter, M. Naeem Shahid, *Simple two Higgs doublet model*. Int. J. Mod. Phys. A **23** (2008) 5159.
113. T. Aaltonen et al. (of the CDF collaboration), Study of multi-muon events produced in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, hep-ex/0810.5357.
114. P. Giromini et al., Phenomenological interpretation of the multi-muon events reported by the CDF collaboration, hep-ph/0810.5730.
115. M.J. Strassler, Flesh and blood, or merely ghosts? Some comments on the multi-muon study at CDF, hep-ph/0811.1560.
116. R. Barbieri, L.J. Hall, V. S. Rychkov, *Improved naturalness with a heavy Higgs boson: an alternative road to CERN LHC physics*. Phys. Rev. D **74** (2006) 015007.
117. P. Binétruy, *The LHC and the universe at large*, Int. J. Mod. Phys. A **24** (2009) 657.

Chapter 2
Quantum Gravity as Sum over Spacetimes

J. Ambjørn, J. Jurkiewicz and R. Loll

2.1 Introduction

A major unsolved problem in theoretical physics is to reconcile the classical theory of general relativity with quantum mechanics. These lectures will deal with an attempt to describe quantum gravity as a path integral over geometries. Such an approach has to be non-perturbative since gravity is a non-renormalizable quantum field theory when the dimension of spacetime is four. In that case the dimension of the gravitational coupling constant G is -2 in units where $\hbar = 1$ and $c = 1$ and the dimension of mass is 1. Thus conventional, perturbative quantum field theory is only expected to be good for energies

$$E^2 \ll 1/G. \tag{1}$$

That is still perfectly good in all experimental situations we can imagine in the laboratory, but an indication that something "new" has to happen at sufficiently large energy, or equivalently, at sufficiently short distances. It is possible, or maybe even likely, that a breakdown of perturbation theory when (1) is not satisfied indicates that new degrees of freedom should be present in a theory valid at higher energies. Indeed, we have a well-known example in the electroweak theory. Originally the electroweak theory was described by a four-fermion interaction. Such a theory is not renormalizable and perturbation theory breaks down at sufficiently high energy. In fact it breaks down unless the energy satisfies (1) with the gravitational coupling constant G replacing the coupling constant G_{F} in front of the four-Fermi interaction (since G_F also has mass dimension -2). The breakdown reflects the appearance of new degrees of freedom, the W and the Z particles, and the four-Fermi interaction

J. Ambjørn (✉)
The Niels Bohr Institute, Copenhagen, Denmark, ambjorn@nbi.dk

J. Jurkiewicz
Institute of Physics, Jagellonian University, Krakow, Poland, jurkiewicz@th.if.uj.edu.pl

R. Loll
Institute for Theoretical Physics, Utrecht University, The Netherlands, r.loll@uu.nl

Ambjørn, J. et al.: *Quantum Gravity as Sum over Spacetimes*. Lect. Notes Phys. **807**, 59–124 (2010)
DOI 10.1007/978-3-642-11897-5_2

is now just an approximation to the process where a fermion interacts via W and Z particles with another fermion. The corresponding electroweak theory is renormalizable.

When it comes to gravity there seems to be no "simple" fix like the one just described. However, string theory is an example of a theory which tries to solve the problem by adding (infinitely many) new degrees of freedom. Loop quantum gravity is another approach to quantum gravity which tries to circumvent the problem of non-renormalizability by introducing rules of quantization which are unconventional from a perturbative point of view. The point of view taken here in these lectures is much more mundane. In a sum-over-histories approach we will attempt to define a non-perturbative quantum field theory which has as its infrared limit ordinary classical general relativity and at the same time has a nontrivial ultraviolet limit. From this point of view it is in the spirit of the renormalization group approach, first advocated long ago by Weinberg [1] and more recently substantiated by several groups of researchers [2–7].

To understand the possibility of a nontrivial ultraviolet fixed point let us first apply ordinary perturbation theory to quantum gravity in a regime where (1) is satisfied. One can in a reliable way calculate the lowest-order quantum correction to the gravitational potential of a point particle:

$$\frac{G}{r} \to \frac{G(r)}{r}, \quad G(r) = G\left(1 - \omega \frac{G}{r^2} + \cdots\right), \quad \omega = \frac{167}{30\pi}. \tag{2}$$

Thus the gravitational coupling constant becomes scale-dependent and transferring from distance to energy we have

$$G(E) = G(1 - \omega G E^2 + \cdots) \approx \frac{G}{1 + \omega G E^2}. \tag{3}$$

It should be stressed that the scenario described in (2) and (3) is completely standard in quantum field theory. Let us take the simplest quantum field theory relevant in nature: quantum electrodynamics. The electron as we observe it in low-energy scattering experiments is screened by vacuum polarization: virtual electron – positron pairs, created out of the vacuum and annihilated again so fast that one has consistency with the energy – time uncertainty relations, act like dipoles and the observed charge becomes less than the "bare" charge. To lowest order (one loop) (2) and (3) are replaced by

$$\frac{e^2}{r} \to \frac{e^2(r)}{r}, \quad e(r) = e\left(1 - \frac{e^2}{6\pi^2}\ln(m\,r)\right) + \cdots, \quad m\,r \ll 1. \tag{4}$$

$$e^2(E) = e^2\left(1 + \frac{e^2}{6\pi^2}\ln(E/m)\right) + \cdots \approx \frac{e^2}{1 - \frac{e^2}{6\pi^2}\ln(E/m)}. \tag{5}$$

The last expression in (5) is precisely the renormalization group-improved formula for the running coupling constant in QED. In fact, it is easy to calculate the so-called (one-loop) β-function for QED from the first equation in (5) and use this β-function to obtain the expression on the left-hand side of (5). Contrary to (3) it breaks down for sufficiently high energy: it has a so-called Landau pole and it reflects that we expect the interactions of QED to be infinitely strong at short distances. We do not believe that such quantum field theories really exist as "stand-alone" theories. They either require an explicit cut-off, which will then enter in the observables at high energies, or have to be embedded in a larger theory without a Landau pole.

Assume that the last expression in (3) was exact for all E (which it is not). One can argue that one should use $G(E)$ in (1) rather than G, in which case one obtains

$$G(E)E^2 < \frac{1}{\omega} = \frac{30\pi}{167} \quad (< 1). \tag{6}$$

Thus, assuming (3) it suddenly seems as if quantum gravity had become a reliable quantum theory at all energy scales, the reason being that the effective coupling constant $G(E)$ becomes weaker at high energies. The behaviour (3) can be described in terms of a β-function for quantum gravity. Introduce the dimensionless coupling constant $\tilde{G}(E)$:

$$\tilde{G}(E) = G(E)E^2. \tag{7}$$

From (3) it follows that $\tilde{G}(E)$ satisfies the following equation:

$$E\frac{\mathrm{d}\tilde{G}}{\mathrm{d}E} = \beta(\tilde{G}), \quad \beta(\tilde{G}) = 2\tilde{G} - 2\omega\tilde{G}^2. \tag{8}$$

The zeros of $\beta(\tilde{G})$ determine the fixed points of the running coupling constant $\tilde{G}(E)$. The zero at $\tilde{G} = 0$ is an infrared fixed point: for $E \to 0$ the coupling constant $\tilde{G}(E) \to 0$ and correspondingly the coupling constant $G(E) \to G$. The zero at $\tilde{G} = 1/\omega$ is an ultraviolet fixed point: for $E \to \infty$ the coupling constant $\tilde{G}(E) \to 1/\omega$ (and $G(E) \to 0$ as $1/(\omega E^2)$).

While there is no compelling reason to take the above arguments very seriously, since they are based on a one-loop calculation, the claim in [2–7] is that a more careful analysis using the so-called exact renormalization group equations or a systematic $2 + \varepsilon$ expansion confirms the picture of a nontrivial UV fixed point. A typical result of the renormalization group calculation is shown in Fig. 2.1 in the case where the effective action has been "truncated" to just include the Einstein term and a cosmological term. The figure has an UV fixed point from where the flow to low energies starts and an infrared fixed point (the origin in the coupling constant coordinate system).

Where do we meet such scenarios (IR fixed points at zero coupling and a nontrivial UV fixed point for a suitably defined dimensionless coupling constant)? Assume one has an asymptotically free field theory in d dimensions, i.e. $g = 0$ is an UV fixed point. The strong interactions in four-dimensional flat spacetime, quantum

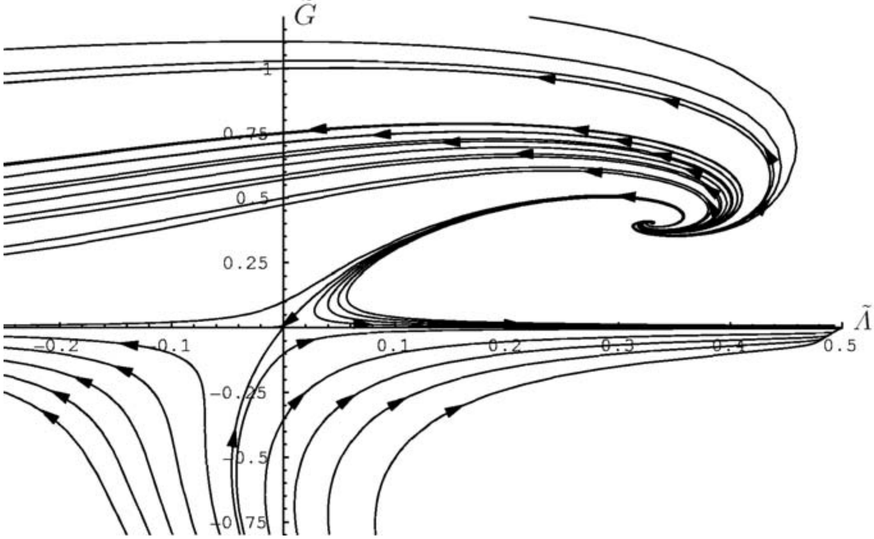


Fig. 2.1 The flow of dimensionless gravitational and cosmological coupling constants $\tilde{G}$ and $\tilde{\Lambda}$ from an UV fixed point. One line flows to the infrared Gaussian fixed point where $\tilde{G}$ and $\tilde{\Lambda}$ are zero. The figure is reprinted with permission from M. Reuter and F. Saueressig, Phys. Rev. D Vol. 65, 065016 (2002). Copyright 2002 by the American Physical Society, (see http://published.aps.org/copyrightFAQ.html)

chromodynamics (QCD), are such a theory. In high-energy scattering experiments the effective, running coupling constant goes to zero. This has been beautifully verified in high-energy experiments. The non-linear sigma model in two-dimensional spacetime is another model. It plays a very important role in string theory, but even before that was extensively studied as a toy model of QCD since it is asymptotically free (i.e. the running, effective coupling constant goes to zero at high energies). The asymptotically free theories have a negative β-function. This is what makes the running coupling constant go to zero at high energies. Change now (artificially) the dimension of spacetime infinitesimally from d to $d + \varepsilon$. Then the β-function to lowest order in ε will change as follows:

$$\beta_d(g) \to \beta_{d+\varepsilon} = \varepsilon g + \beta_g(g), \tag{9}$$

and the situation is as shown in Fig. 2.2: $g = 0$ changes from an UV fixed point to an infrared fixed point while the new UV fixed point will be displaced to finite positive value $g_c(\varepsilon)$ of g, a value which goes to zero when ε goes to zero. In the case of gravity we have formally a renormalizable theory when $d = 2$, the dimension where the gravitational coupling constant G is dimensionless. One can show that the two-dimensional theory can be viewed as asymptotically free (see (113) below). To apply (9) to four-dimensional quantum gravity starting with a renormalizable theory of gravity means that ε has to be two, which is not very small. Thus the considerations above make little sense at a quantitative level, but the use of the

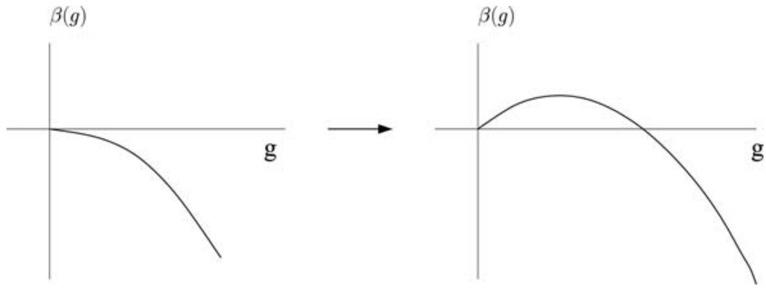


Fig. 2.2 The change in the beta function $\beta(g)$ in an asymptotically free theory when the dimension changes from the critical dimension d, where the coupling constant g is dimensionless, to $d + \varepsilon$

exact renormalization group indicates, as mentioned, that the qualitative picture is correct.

The discussion above shows that there might be a chance that quantum gravity can be defined as an ordinary quantum field theory at a nontrivial ultraviolet fixed point. It clearly requires non-perturbative tools to address the question of the existence of such a fixed point and to analyse the properties of the field theory defined by approaching the fixed point. One way to proceed is by using a lattice regularization of the quantum field theory in question. The lattice provides an UV regularization of the quantum field theory, namely, the inverse lattice spacing $1/a$. The task is then to define a suitable "continuum" limit of this lattice theory. The procedure used is typically as follows: let $\mathcal{O}(x_n)$ be an observable, x_n denoting a lattice point. We write $x_n = a\,n$, n measuring the position in integer lattice spacings. One can then obtain, either by computer simulations or by analytical calculations, the correlation length $\xi(g_0)$ in lattice units, from

$$-\log\langle\mathcal{O}(x_n)\mathcal{O}(y_m)\rangle \sim |n-m|/\xi(g_0) + o(|n-m|). \tag{10}$$

A continuum limit of the lattice theory may then exist if it is possible to fine-tune the bare coupling constant g_0 of the theory to a critical value g_0^c such that the correlation length goes to infinity, $\xi(g_0) \to \infty$. Knowing how $\xi(g_0)$ diverges for $g_0 \to g_0^c$ determines how the lattice spacing a should be taken to zero as a function of the coupling constants, namely,

$$\xi(g_0) \propto \frac{1}{|g_0 - g_0^c|^\nu}, \qquad a(g_0) \propto |g_0 - g_0^c|^\nu. \tag{11}$$

This particular scaling of the lattice spacing ensures that one can define a physical mass m_{ph} by

$$m_{ph} a(g_0) = 1/\xi(g_0)$$

such that the correlator $\langle \mathcal{O}(x_n)\mathcal{O}(y_m)\rangle$ falls off exponentially like $\mathrm{e}^{-m_{ph}|x_n-y_m|}$ for $g_0 \to g_0^c$ when $|x_n - y_m|$, but not $|n-m|$, is kept fixed in the limit $g_0 \to g_0^c$. Thus we have created a picture where the underlying lattice spacing goes to zero while the physical mass (or the correlation length measured in physical length units, not in lattice spacings) is kept fixed. This is the standard Wilsonian scenario for obtaining the continuum (Euclidean) quantum field theory associated with the critical point g_0^c of a *second-order* phase transition (for second-order phase transitions there exists a correlation length which diverges, usually associated with the order parameter characterizing the transition).

We would like to apply a similar approach to quantum gravity, and thus obtain a new way to investigate if quantum gravity can be defined non-perturbatively as a quantum field theory. The predictions from such a theory could then be compared with the renormalization group predictions related to the asymptotic safety picture described above. It should be mentioned that the asymptotic safety picture is not the only suggestion for a *continuum* quantum theory of gravity using only "conventional" ideas of quantum field theory. Very recently two other scenarios have been suggested. One is called Lifshitz gravity [8, 9] and is a theory where the non-renormalizability of the Einstein – Hilbert theory is cural derivatives in a way somewhat similar to what Lifshitz did many years ago in statistical models. In fact, the set-up of the theory has some resemblance with the lattice-theory set-up of "causal dynamical triangulations (CDT)", to be described below, since a time foliation is assumed and the infrared limit is that of GR. However, contrary to Lifshitz gravity, we do not attempt to put in higher spatial derivatives in the lattice theory. However, when a continuum limit in the lattice theory is taken in a specific way which is not entirely symmetric in space and time one cannot rule out that higher spatial derivatives can play a role. The other model goes by the name of "scale-invariant gravity" [10–12]. It modifies gravity into a renormalizable theory by introducing a scalar degree of freedom in addition to the transverse gravitational degrees of freedom. Also this model has interesting features not incompatible with the results of computer simulations using the CDT lattice model.

As already mentioned, we will use a lattice approach known as *causal dynamical triangulations* (CDT) as a regularization. In Sect. 2.2 we give a short description of the formalism, providing the definitions which are needed later to describe the measurements. CDT establishes a non-perturbative way of performing the sum over four-geometries (for more extensive definitions, see [13, 14]). It sums over the class of piecewise linear four-geometries which can be assembled from four-dimensional simplicial building blocks of link length a, such that only *causal* spacetime histories are included. The challenge when searching for a *field theory* of quantum gravity is to find a theory which behaves as described above, i.e. as in (11). The challenge is threefold: (i) to find a suitable non-perturbative formulation of such a theory which satisfies a minimum of reasonable requirements, (ii) to find observables which can be used to test relations like (10), and (iii) to show that one can adjust the coupling constants of the theory such that (11) is satisfied. Although we will focus on (i) in what follows, let us immediately mention that (ii) is notoriously difficult in a theory of quantum gravity, where one is faced with a number of questions originating in the dynamical nature of geometry. What is the meaning of distance when integrating over all geometries? How do we attach a meaning to local spacetime points like x_n and y_n? How can we define at all local, diffeomorphism-invariant quantities in the continuum which can then be translated to the regularized (lattice) theory? What we want to point out here is that although (i)–(iii) are standard requirements when relating critical phenomena and (Euclidean) quantum field theory, gravity *is* special and may require a reformulation of (part of) the standard scenario sketched above. We will return to this issue later.

Our proposed non-perturbative formulation of four-dimensional quantum gravity has a number of nice properties.

First, it sums over a class of piecewise linear geometries. The characteristic feature of piecewise linear geometries is that they admit a description without the use of coordinate systems. In this way we perform the sum over geometries directly, avoiding the cumbersome procedure of first introducing a coordinate system and then getting rid of the ensuing gauge redundancy, as one has to do in a continuum calculation. Our underlying assumptions are that (1) the class of piecewise linear geometries is in a suitable sense dense in the set of all geometries relevant for the path integral (probably a fairly mild assumption) and (2) that we are using a correct measure on the set of geometries. This is a more questionable assumption since we do not even know whether such a measure exists. Here one has to take a pragmatic attitude in order to make progress. We will simply examine the outcome of our construction and try to judge whether it is promising.

Second, our scheme is background-independent. No distinguished geometry, accompanied by quantum fluctuations, is put in by hand. If the CDT-regularized theory is to be taken seriously as a potential theory of quantum gravity, there has to be a region in the space spanned by the bare coupling constants where the geometry of spacetime bears some resemblance with the kind of universe we observe around us. That is, the theory should create dynamically an effective background geometry around which there are (small) quantum fluctuations. This is a very non-trivial property of the theory and one we are going to investigate in some detail. Computer simulations presented in these lectures confirm in a much more direct way the indirect evidence for such a scenario which we have known for some time and first reported in [15, 16]. They establish the de Sitter nature of the background spacetime, quantify the fluctuations around it, and set a physical scale for the universes we are dealing with. The main results of these investigations, without the numerical details, were announced in [17] and a detailed account of the results was presented in [18].

The remainder of these lecture notes is organized as follows: in Sect. 2.2 we describe the lattice formulation of four-dimensional quantum gravity. In Sect. 2.3 the numerical results in four dimensions are summarized. We view these results as very important, but they also serve as a motivation for moving to two dimensions. While there is no propagating graviton in two-dimensional quantum gravity, it is a diffeomorphism-invariant theory and almost all of the conceptional problems mentioned above are present there. Thus it is an important exercise to solve two-dimensional quantum gravity. Surprisingly this can be done in the lattice regularization known as "dynamical triangulation". An important corollary is that (1) one can explicitly construct the continuum limit of the lattice theory and (2) show that it agrees with the so-called Liouville two-dimensional quantum gravity theory. This latter theory is a continuum conformal field theory, explicit solvable, and, when viewed in the correct way, a diffeomorphism-invariant theory. Thus there is indeed no problem having a lattice regularization of a diffeomorphism-invariant theory. In Sect. 2.4 we solve what is known as two-dimensional Euclidean quantum gravity. In Sect. 2.5 we show how one can interpolate from Euclidean two-dimensional quantum gravity to "Lorentzian" two-dimensional quantum gravity which is a two-dimensional version of the four-dimensional gravity theory we have discussed in Sect. 2.3. Finally Sect. 2.7 discusses the results obtained and outlines perspectives.

2.2 CDT

The use of so-called causal dynamical triangulations (CDT) stands in the tradition of [19, 20], which advocated that in a gravitational path integral with the correct, Lorentzian signature of spacetime one should sum over causal geometries only. More specifically, we adopted this idea when it became clear that attempts to formulate a *Euclidean* non-perturbative quantum gravity theory run into trouble in spacetime dimension d larger than two as will be described below.

This implies that we start from Lorentzian simplicial spacetimes with $d = 4$ and insist that only causally well-behaved geometries appear in the (regularized) Lorentzian path integral. A crucial property of our explicit construction is that each of the configurations allows for a rotation to Euclidean signature. We rotate to a Euclidean regime in order to perform the sum over geometries (and rotate back again afterward if needed). We stress here that although the sum is performed over geometries with Euclidean signature, it is different from what one would obtain in a theory of quantum gravity based ab initio on Euclidean spacetimes. The reason is that not all Euclidean geometries with a given topology are included in the "causal" sum since in general they have no correspondence to a causal Lorentzian geometry.

We refer to [13] for a detailed description of how to construct the class of piecewise linear geometries used in the Lorentzian path integral. The most important assumption is the existence of a global proper-time foliation. This is symbolically illustrated in Fig. 2.3 where we compare the construction to the one of ordinary quantum mechanics: the path integral of ordinary quantum mechanics is regularized as a sum over piecewise linear paths from point x_i to point x_f in time t_f–t_i. The time steps have length a and the continuum limit is obtained when the length a of these "building blocks" goes to zero. Similarly, in the quantum gravity case we have a sum over four-geometries, "stretching" between two three-geometries separated by a proper time t and constructed from four-dimensional building blocks, as described below. On the figure we show for illustrational simplicity only a single spacetime history and replace the three-dimensional spatial geometries with a one-dimensional one of S^1-topology. Moving from left to right we have a time foliation where at

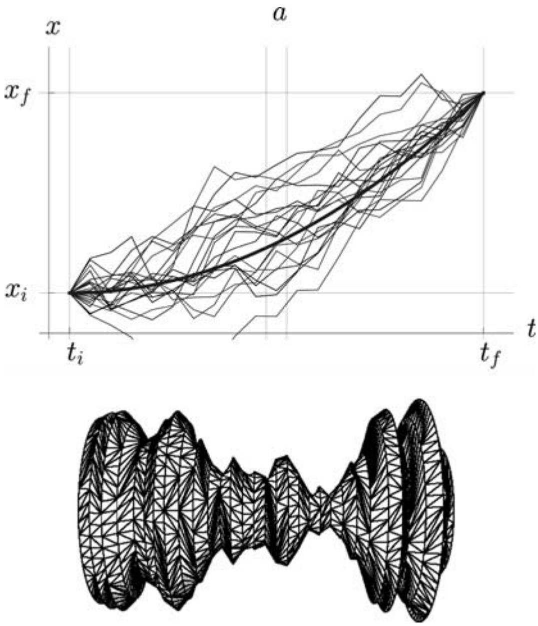


Fig. 2.3 Piecewise linear spacetime histories in quantum mechanics and in (1+1)-dimensional quantum gravity. In the gravity case we show only a single spacetime history, while in the quantum particle case we show many such histories as well as the average path (*thick line*)

each discrete time step space is represented by a circle. Neighbouring circles are then connected by piecewise flat building blocks, usually triangles, as illustrated in Fig. 2.17 in Sect. 2.5. In the "real" four-dimensional case, the spatial slices of topology S^1 will be replaced by spatial slices of topology S^3, and neighbouring spatial S^3 slices are then connected by four-simplices as illustrated in Fig. 2.4 and described in detail below. There is an important difference between the quantum mechanical sum over paths and our sum over geometries with a time foliation: the time t in

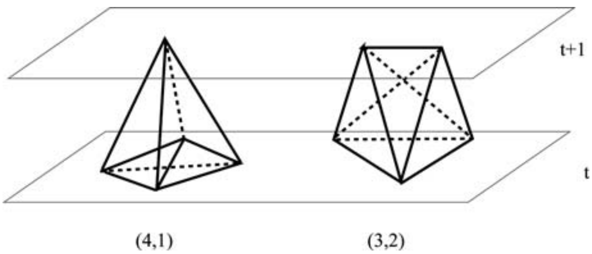


Fig. 2.4 (4, 1) and a (3, 2) simplices connecting two neighbouring spatial slices. We also have symmetric (1, 4) and (2, 3) simplices with a vertex and a line, respectively, at time t and a tetrahedron and a triangle, respectively, at time $t + 1$. For simplicity we denote the total number of (4, 1) and (1, 4) simplices by $N_4^{(4,1)}$ and similarly the total number of (3, 2) and (2, 3) simplices by $N_4^{3,2}$

the quantum mechanical example is an external parameter, while the time t in the case of quantum gravity is intrinsic. Also, again for the purpose of illustration, the two-dimensional geometry has been drawn embedded in three-dimensional space, but in the path integral implementing the summation over geometries there is no such embedding present. Finally, we cannot refrain from mentioning that the paths shown for the quantum mechanical particle are in fact typical paths which appear for the (Euclideanized) path integral of a particle placed in an external potential. They are picked out from an actual Monte Carlo simulation of such a physical system. Similarly, the two-dimensional surface is a surface picked out from a Monte Carlo simulation of two-dimensional quantum gravity and thus corresponds to a typical two-dimensional surface which appears in the path integral. This is the reason for the somewhat poor graphic representation: there is no natural length-preserving representation of the surface in three-dimensions such that the surface is not self-intersecting. For better graphic illustrations (animations) of the two-dimensional surfaces which appear in the two-dimensional quantum-gravitational path integral we refer to the link [21].

As mentioned above, we assume that the spacetime topology is that of $S^3 \times R$, the spatial topology being that of S^3 merely for convenience. The spatial geometry at each discrete proper time step t_n is represented by a triangulation of S^3, made up of equilateral spatial tetrahedra with squared side length $\ell_s^2 \equiv a^2 > 0$. In general, the number $N_3(t_n)$ of tetrahedra and how they are glued together to form a piecewise flat three-dimensional manifold will vary with each time step t_n. In order to obtain a four-dimensional triangulation, the individual three-dimensional slices must still be connected in a causal way, preserving the S^3-topology at all intermediate times t between t_n and t_{n+1}.[1] This is done as illustrated in Fig. 2.4, introducing what we call (4, 1)-simplices and (3, 2)-simplices. More precisely, a (4, 1)-simplex is a four-simplex with four of its vertices (i.e. a boundary tetrahedron) belonging to the triangulation of $S^3(t_n)$, the time slice corresponding to time t_n, and the fifth vertex belonging to the triangulation of $S^3(t_{n+1})$, the time slice corresponding to time t_{n+1}. Similarly, a (3, 2) simplex has three vertices, i.e. a triangle, belonging to the triangulation of $S^3(t_n)$ and two vertices, i.e. a link, belonging to the triangulation of $S^3(t_{n+1})$. We have also simplices of type (1, 4) and (2, 3), which are defined in an obvious way, interchanging the role of $S^3(t_n)$ and $S^3(t_{n+1})$. One can show that two triangulations of $S^3(t_n)$ and $S^3(t_{n+1})$ can be "connected" by these four building blocks glued together in a suitable way such that we have a four-dimensional triangulation of $S^3 \times [0, 1]$. Also, two given triangulations of $S^3(t_n)$ and $S^3(t_{n+1})$ can be connected in many ways compatible with the topology $S^3 \times [0, 1]$. In the path integral we will be summing over all possible ways to connect a given triangulation $S^3(t_n)$ to a given triangulation of $S^3(t_{n+1})$ compatible with the topology

[1] This implies the absence of branching of the spatial universe into several disconnected pieces, so-called *baby universes*, which (in Lorentzian signature) would inevitably be associated with causality violations in the form of degeneracies in the light-cone structure, as has been discussed elsewhere (see, for example, [22–24]).

$S^3 \times [0, 1]$. In addition we will sum over all three-dimensional triangulations of S^3 at all times t_n.

We allow for an asymmetry between temporal and spatial lattice length assignments. Denote by ℓ_t and ℓ_s the length of the time-like links and the space-like links, respectively. Then $\ell_t^2 = -\alpha \ell_s^2$, $\alpha > 0$. The explicit rotation to Euclidean signature is done by performing the rotation $\alpha \to -\alpha$ in the complex lower half-plane, $|\alpha| > 7/12$, such that we have $\ell_t^2 = |\alpha| \ell_s^2$ (see [13] for a discussion).

The Einstein – Hilbert action S^{EH} has a natural geometric implementation on piecewise linear geometries in the form of the Regge action. This is given by the sum of the so-called deficit angles around the two-dimensional "hinges" (subsimplices in the form of triangles), each multiplied with the volume of the corresponding hinge. In view of the fact that we are dealing with piecewise linear, and not smooth metrics, there is no unique "approximation" to the usual Einstein – Hilbert action, and one could in principle work with a different form of the gravitational action. We will stick with the Regge action, which takes on a very simple form in our case, where the piecewise linear manifold is constructed from just two different types of building blocks. After rotation to Euclidean signature one obtains for the action (see [14] for details)

$$
\begin{aligned}
S_E^{\mathrm{EH}} &= \frac{1}{16\pi^2 G} \int d^4x \sqrt{g}(-R + 2\Lambda) \quad \longrightarrow \qquad (12) \\
S_E^{\mathrm{Regge}} &= -(\kappa_0 + 6\Delta) N_0 + \kappa_4 \left(N_4^{(4,1)} + N_4^{(3,2)} \right) + \Delta \left(2N_4^{(4,1)} + N_4^{(3,2)} \right),
\end{aligned}
$$

where N_0 denotes the total number of vertices in the four-dimensional triangulation and $N_4^{(4,1)}$ and $N_4^{(3,2)}$ denote the total number of the four-simplices described above, so that the total number N_4 of four-simplices is $N_4 = N_4^{(4,1)} + N_4^{(3,2)}$. The dimensionless coupling constants κ_0 and κ_4 are related to the bare gravitational and bare cosmological coupling constants, with appropriate powers of the lattice spacing a already absorbed into κ_0 and κ_4. The *asymmetry parameter* Δ is related to the parameter α introduced above, which describes the relative scale between the (squared) lengths of space- and time-like links. It is both convenient and natural to keep track of this parameter in our set-up, which from the outset is not isotropic in time and space directions, see again [14] for a detailed discussion. Since we will in the following work with the path integral after Wick rotation, let us redefine $\tilde{\alpha} := -\alpha$ [14], which is positive in the Euclidean domain.[2] For future reference, the Euclidean four-volume of our universe for a given choice of $\tilde{\alpha}$ is given by

$$
V_4 = \tilde{C}_4(\xi)\, a^4\, N_4^{(4,1)} = \tilde{C}_4(\xi)\, a^4\, N_4/(1 + \xi), \qquad (13)
$$

where ξ is the ratio

[2] The most symmetric choice is $\tilde{\alpha} = 1$, corresponding to vanishing asymmetry, $\Delta = 0$.

$$\xi = N_4^{(3,2)}/N_4^{(4,1)}, \tag{14}$$

and $\tilde{C}_4(\xi)\,a^4$ is a measure of the "effective four-volume" of an "average" four-simplex. ξ will depend on the choice of coupling constants in a rather complicated way (for a detailed discussion we refer to [13, 18]).

The path integral or partition function for the CDT version of quantum gravity is now

$$Z(G,\Lambda) = \int \mathcal{D}[g]\, \mathrm{e}^{-S_E^{\mathrm{EH}}[g]} \quad \rightarrow \quad Z(\kappa_0,\kappa_4,\Delta) = \sum_{\mathcal{T}} \frac{1}{C_{\mathcal{T}}}\, \mathrm{e}^{-S_E(\mathcal{T})}, \tag{15}$$

where the summation is over all causal triangulations $\mathcal{T}$ of the kind described above, and we have dropped the superscript "Regge" on the discretized action. The factor $1/C_{\mathcal{T}}$ is a symmetry factor, given by the order of the automorphism group of the triangulation $\mathcal{T}$. The actual set-up for the simulations is as follows. We choose a fixed number N of spatial slices at proper times t_1, $t_2 = t_1 + a_t$, up to $t_N = t_1 + (N-1)a_t$, where $\Delta t \equiv a_t$ is the discrete lattice spacing in temporal direction and $T = Na_t$ the total extension of the universe in proper time. For convenience we identify t_{N+1} with t_1, in this way imposing the topology $S^1 \times S^3$ rather than $I \times S^3$. This choice does not affect physical results, as will become clear in due course.

Our next task is to *evaluate* the non-perturbative sum in (15), if possible, analytically. This can be done in spacetime dimension $d = 2$ ([25–31] (and we discuss this in detail below) and at least partially in $d = 3$ [32–35], but presently an analytic solution in four dimensions is out of reach. However, we are in the fortunate situation that $Z(\kappa_0,\kappa_4,\Delta)$ can be studied quantitatively with the help of Monte Carlo simulations. The type of algorithm needed to update the piecewise linear geometries has been around for a while, starting from the use of dynamical triangulations in bosonic string theory (two-dimensional Euclidean triangulations) [36–39] and later extended to their application in Euclidean four-dimensional quantum gravity [40–42]. In [13] the algorithm was modified to accommodate the geometries of the CDT set-up. The algorithm is such that it takes the symmetry factor $C_{\mathcal{T}}$ into account automatically.

We have performed extensive Monte Carlo simulations of the partition function Z for a number of values of the bare coupling constants. As reported in [14], there are regions of the coupling constant space which do not appear relevant for continuum physics in that they seem to suffer from problems similar to the ones found earlier in *Euclidean* quantum gravity constructed in terms of dynamical triangulations, which essentially led to its abandonment in $d > 2$. What is observed in Euclidean four-dimensional quantum gravity is the following: when the (inverse, bare) gravitational coupling κ_0 is sufficiently large one sees so-called branched polymers, i.e. not really a four-dimensional universe, but a universe which branches out like a tree with so many branches that it becomes truly fractal when the number of four-simplices becomes infinite, and its Hausdorff dimension is 2. Such triangulations represent the most extended triangulations one can construct unless one explicitly forbids branching. When the (inverse, bare) gravitational coupling κ_0 is sufficiently small one observes a totally crumpled universe with almost no extension. In this phase there exist vertices of very high order and the connectivity of the triangulation is such that it is possible to move from any four-simplex to any other crossing only a few neighbouring four-simplices. The Hausdorff dimension of such a triangulation is infinite in the limit where the number of four-simplices goes to infinity. These two phases, the crumpled and the branched-polymer phase, are separated by a phase transition line along which there is a first-order transition. It was originally hoped that one could find a point on the critical line where the first-order transition becomes second order and which could then be used as a fixed point where a continuum theory of quantum gravity could be defined along the lines suggested by (10) and (11). However, such a second-order transition was not found, and eventually the idea of a theory of four-dimensional Euclidean quantum gravity was abandoned. A new principle for selecting the class of geometries one should use in the path integral was needed and this led to the suggestion to include only *causal* triangulations in the sum over spacetime histories.

When we include only the causal triangulations in the path integral, we still see a remnant of the Euclidean structure just described, namely, when the (inverse, bare) gravitational coupling κ_0 is sufficiently large, the Monte Carlo simulations exhibit a sequence in time direction of small, disconnected universes, none of them showing any sign of the scaling one would expect from a macroscopic universe. We denote this phase by A. We believe that this phase of the system is a Lorentzian version of the branched-polymer phase of Euclidean quantum gravity. By contrast, when Δ is sufficiently small, the simulations reveal a universe with a vanishing temporal extension of only a few lattice spacings, ending both in past and future in a vertex of very high order, connected to a large fraction of all vertices. This phase is most likely related to the so-called crumpled phase of Euclidean quantum gravity. We denote this phase by B. The crucial and new feature of the quantum superposition in terms of *causal* dynamical triangulations is the appearance of a region in coupling constant space which is different and interesting and where continuum physics may emerge. It is in this region that we have performed the simulations discussed here and where work up to now has already uncovered a number of intriguing physical results [14–16, 43]. In Fig. 2.6 we have shown how different configurations look in the three phases discussed above, and in Fig. 2.5 we have shown the tentative phase diagram in the coupling constant space of κ_0, κ_4 and Δ. A “critical” surface is shown in the figure. Keeping κ_0 and Δ fixed, κ_4 acts as a chemical potential for N_4; the smaller the κ_4, the larger the $\langle N_4 \rangle$. At some critical value $\kappa_4(\kappa_0, \Delta)$, depending on the choice of κ_0 and Δ, $\langle N_4 \rangle \to \infty$. For $\kappa_4 < \kappa_4(\kappa_0, \Delta)$ the partition function is plainly divergent and not defined. When we talk about phase transitions we are always at the “critical” surface

$$\kappa_4 = \kappa_4(\kappa_0, \Delta), \tag{16}$$

simply because we cannot have a phase transition unless $N_4 = \infty$. We put “critical” into quotation marks since it only means that we probe infinite four-volume.

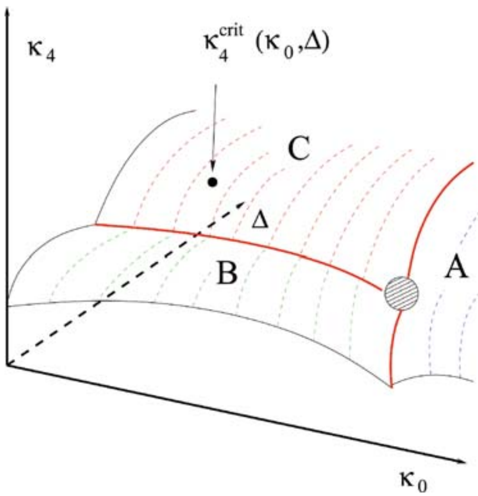


Fig. 2.5 The phases A, B and C in the coupling constant space $(\kappa_0, \Delta, \kappa_4)$. Phase C is the one where extended four-dimensional geometries emerge

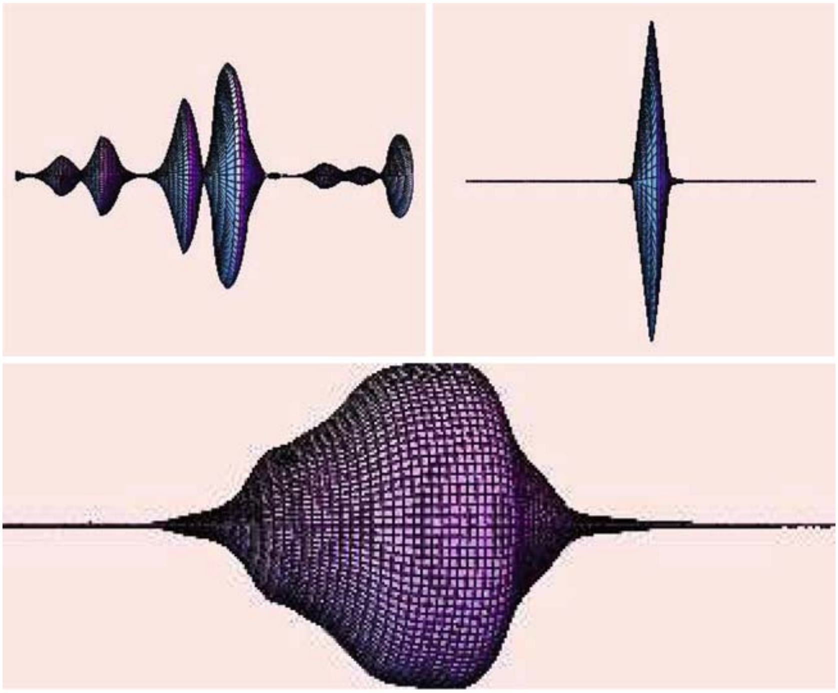


Fig. 2.6 Typical configurations in the phases A, B and C (*lowest figure*). Phase C is the one where extended four-dimensional geometries emerge

No continuum limit is necessarily associated with a point on this surface. To decide this issue requires additional investigation. A good analogy is the Ising model on a finite lattice. To have a genuine phase transition for the Ising model we have to take the lattice volume to infinity since there are no genuine phase transitions for finite systems. However, just taking the lattice volume to infinity is not sufficient to ensure critical behaviour of the Ising model. We also have to tune the coupling constant to its critical value. Being on the "critical" surface, or rather "infinite-volume" surface (16), we can discuss various phases, and these are the ones indicated in the figure. The different phases are separated by phase transitions, which might be first-order. However, we have not yet conducted a systematic investigation of the order of the transitions. Looking at Fig. 2.5, we have two lines of phase transitions, separating phase A and phase C and separating phase B and phase C respectively. They meet in the point indicated on the figure. It is tempting to speculate that this point might be associated with a higher-order transition, as is common for statistical systems in such a situation. We will return to this point later.

In the Euclideanized setting the value of the cosmological constant determines the spacetime volume V_4 since the two appear in the action as conjugate variables. We therefore have $\langle V_4 \rangle \sim G/\Lambda$ in a continuum notation, where G is the gravitational coupling constant and Λ the cosmological constant. In the computer simulations it is more convenient to keep the four-volume fixed or partially fixed. We will implement this by fixing the total number of four-simplices of type $N_4^{(4,1)}$ or, equivalently, the total number N_3 of tetrahedra making up the spatial S^3 triangulations at times t_i, $i = 1, \ldots, N$,

$$N_3 = \sum_{i=1}^{N} N_3(t_i) = \frac{1}{2} N_4^{(4,1)}. \tag{17}$$

We know from the simulations that in the phase of interest $\left\langle N_4^{(4,1)} \right\rangle \propto \left\langle N_4^{(3,2)} \right\rangle$ as the total volume is varied [14]. This effectively implies that we only have two bare coupling constants κ_0, Δ in (15), while we compensate by hand for the coupling constant κ_4 by studying the partition function $Z\left(\kappa_0, \Delta; N_4^{(4,1)}\right)$ for various $N_4^{(4,1)}$. To keep track of the ratio $\xi(\kappa_0, \Delta)$ between the expectation value $\left\langle N_4^{(3,2)} \right\rangle$ and $N_4^{(4,1)}$, which depends weakly on the coupling constants, we write (c.f. (14))

$$\langle N_4 \rangle = N_4^{(4,1)} + \left\langle N_4^{(3,2)} \right\rangle = N_4^{(4,1)}(1 + \xi(\kappa_0, \Delta)). \tag{18}$$

For all practical purposes we can regard N_4 in a Monte Carlo simulation as fixed. The relation between the partition function we use and the partition function with variable four-volume is given by the Laplace transformation

$$Z(\kappa_0, \kappa_4, \varDelta) = \int_0^\infty \mathrm{d}N_4 \, \mathrm{e}^{-\kappa_4 N_4} \, Z(\kappa_0, N_4, \Delta), \tag{19}$$

where strictly speaking the integration over N_4 should be replaced by a summation over the discrete values N_4 can take. Returning to Fig. 2.5, keeping N_4 fixed rather than fine-tuning κ_4 to the critical value κ_4^c implies that one is already on the "critical" surface drawn in Fig. 2.5, assuming that N_4 is sufficiently large (in principle infinite). Whether N_4 *is* sufficiently large to qualify as "infinite" can be investigated by performing the computer simulations for different N_4's and comparing the results. This is a technique we will use over and over again in the following.

2.3 Numerical Results

The Monte Carlo simulations referred to above will generate a sequence of spacetime histories. An individual spacetime history is not an observable, in the same way as a path $x(t)$ of a particle in the quantum mechanical path integral is not. However, it is perfectly legitimate to talk about the *expectation value* $\langle x(t)\rangle$ as well as the *fluctuations around* $\langle x(t)\rangle$. Both of these quantities are in principle calculable in quantum mechanics. Let us make a slight digression and discuss this in some detail since it illustrates well the picture we also hope emerges in a theory of quantum gravity. Consider the particle example shown in Fig. 2.3. We have a particle moving from x_i at t_i to x_f at t_f. In general there will be a classical motion of the particle satisfying these boundary conditions (we will assume that for simplicity). If $\hbar$ can be considered small compared to the other parameters entering into the description of the system, the classical path will be a good approximation to $\langle x(t)\rangle$ according to Ehrenfest's theorem. In Fig. 2.3 the smooth curve represents $\langle x(t)\rangle$. In the path integral we sum over all continuous paths from (x_i, t_i) to (x_f, t_f) as illustrated in Fig. 2.3. However, when all other parameters in the problem are large compared to $\hbar$ we expect a "typical" path to be close to $\langle x(t)\rangle$ which also will be close to the classical path. Let us make this explicit in the simple case of the harmonic oscillator. Let $x_{cl}(t)$ denote the solution to the classical equations of motion such that $x_{cl}(t_i) = x_i$ and $x_{cl}(t_f) = x_f$. For the harmonic oscillator the decomposition

$$x(t) = x_{cl}(t) + y(t), \quad y(t_i) = y(t_f) = 0$$

leads to an exact factorization of the path integral thanks to the quadratic nature of the action. The part involving $x_{cl}(t)$ gives precisely the classical action and the part involving $y(t)$ the contributions from the fluctuations, independent of the classical part. Taking the classical path to be macroscopic gives a picture of a macroscopic path dressed with small quantum fluctuations, small because they are independent of the classical motion. Explicitly we have for the fluctuations (Euclidean calculation)

$$\left\langle \int_{t_i}^{t_f} \mathrm{d}t\; y^2(t) \right\rangle = \frac{\hbar}{2m\omega^2}\,\left(\frac{\omega(t_f - t_i)}{\tanh(\omega(t_f - t_i))} - 1\right).$$

Thus the harmonic oscillator is a simple example of what we hope for in quantum gravity: Let the size of the system be macroscopic, i.e. $x_{cl}(t)$ is macroscopic (put in by hand), then the quantum fluctuations around this path are small and of the order

$$\langle |y| \rangle \propto \sqrt{\frac{\hbar}{m\omega^2(t_f - t_i)}}.$$

We hope this translates into the description of our universe: the macroscopic size of the universe dictated by the (inverse) cosmological constant in any Euclidean description (trivial to show in the model by simply differentiating the partition function with respect to the cosmological constant and in the simulations thus put in by hand) and the small quantum fluctuations dictated by the other coupling constant, namely, the gravitational coupling constant.

2.3.1 The Emergent de Sitter Background

Obviously, there are many more dynamical variables in quantum gravity than there are in the particle case. We can still imitate the quantum mechanical situation by picking out a particular one, for example, the spatial three-volume $V_3(t)$ at proper time t. We can measure both its expectation value $\langle V_3(t)\rangle$ and fluctuations around it. The former gives us information about the large-scale "shape" of the universe we have created in the computer. First we will describe the measurements of $\langle V_3(t)\rangle$, keeping a more detailed discussion of the fluctuations to Sect. 2.3.2 below.

A "measurement" of $V_3(t)$ consists of a table $N_3(i)$, where $i = 1, \ldots, N$ denotes the number of time slices. Recall from Sect. 2.2 that the sum over slices $\sum_{i=1}^{N} N_3(i)$ is kept constant. The time axis has a total length of N time steps, where $N = 80$ in the actual simulations, and we have cyclically identified time slice $N + 1$ with time slice 1.

What we observe in the simulations is that for the range of discrete volumes N_4 under study the universe does *not* extend (i.e. has appreciable three-volume) over the entire time axis, but rather is localized in a region much shorter than 80 time-slices. Outside this region the spatial extension $N_3(i)$ will be minimal, consisting of the minimal number (five) of tetrahedra needed to form a three-sphere S^3, plus occasionally a few more tetrahedra.[3] This thin "stalk" therefore carries little four-volume and in a given simulation we can for most practical purposes consider the total four-volume of the remainder, the extended universe, as fixed.

In order to perform a meaningful average over geometries which explicitly refers to the extended part of the universe, we have to remove the translational zero mode which is present. We refer to [18] for a discussion of the procedure. Having defined the centre of volume along the time direction of our spacetime configurations we can now perform superpositions of configurations and define the average $\langle N_3(i)\rangle$ as a function of the discrete time i. The results of measuring the average discrete

[3] This kinematic constraint ensures that the triangulation remains a *simplicial manifold* in which, for example, two d-simplices are not allowed to have more than one $(d-1)$-simplex in common.

spatial size of the universe at various discrete times i are illustrated in Fig. 2.7 and can be succinctly summarized by the formula

$$N_3^{cl}(i) := \langle N_3(i)\rangle = \frac{N_4}{2(1+\xi)}\,\frac{3}{4}\,\frac{1}{s_0 N_4^{1/4}}\cos^3\left(\frac{i}{s_0 N_4^{1/4}}\right), \quad s_0 \approx 0.59, \tag{20}$$

where $N_3(i)$ denotes the number of three-simplices in the spatial slice at discretized time i and N_4 the total number of four-simplices in the entire universe. Since we are keeping $N_4^{(4,1)}$ fixed in the simulations and since ξ changes with the choice of bare coupling constants, it is sometimes convenient to rewrite (20) as

$$N_3^{cl}(i) = \frac{1}{2}N_4^{(4,1)}\,\frac{3}{4}\,\frac{1}{\tilde{s}_0\left(N_4^{(4,1)}\right)^{1/4}}\cos^3\left(\frac{i}{\tilde{s}_0\left(N_4^{(4,1)}\right)^{1/4}}\right), \tag{21}$$

where $\tilde{s}_0$ is defined by $\tilde{s}_0\left(N_4^{(4,1)}\right)^{1/4} = s_0 N_4^{1/4}$. Of course, formula (20) is only valid in the extended part of the universe where the spatial three-volumes are larger than the minimal cut-off size.

The data shown in Fig. 2.7 have been collected at the particular values $(\kappa_0, \Delta) = (2.2, 0.6)$ of the bare coupling constants and for $N_4 = 362,000$ (corresponding to $N_4^{(4,1)} = 160,000$). For this value of (κ_0, Δ) we have verified relation (20) for N_4 ranging from 45,500 to 362,000 building blocks (45,500, 91,000, 181,000 and 362,000). After rescaling the time and volume variables by suitable

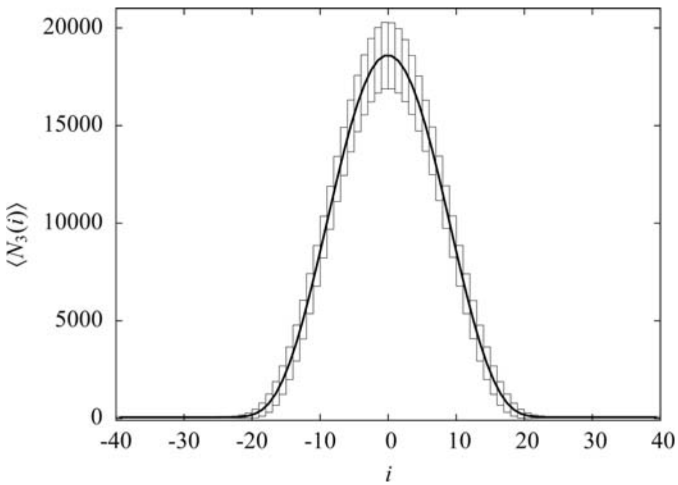


Fig. 2.7 Background geometry $\langle N_3(i)\rangle$: MC measurements for fixed $N_4^{(4,1)} = 160,000$ ($N_4 = 362,000$) and best fit (20) yield indistinguishable curves at given plot resolution. The *bars* indicate the average size of quantum fluctuations

powers of N_4 according to relation (20), and plotting them in the same way as in Fig. 2.7, one finds almost total agreement between the curves for different spacetime volumes. This is illustrated in Fig. 2.8. Thus we have here a beautiful example of finite-size scaling, and at least when we discuss the average three-volume $V_3(t)$ all our discretized volumes N_4 are large enough that we can treat them as infinite, in the sense that no further change will occur for larger N_4.

By contrast, the quantum fluctuations indicated in Fig. 2.7 as vertical bars *are* volume-dependent and will be larger the smaller the total four-volume, see Sect. 2.3.2 for details. Equation (20) shows that spatial volumes scale according to $N_4^{3/4}$ and time intervals according to $N_4^{1/4}$, as one would expect for a genuinely *four*-dimensional spacetime and this is exactly the scaling we have used in Fig. 2.8. This strongly suggests a translation of (20) to a continuum notation. The most natural identification is given by

$$\sqrt{g_{tt}}\; V_3^{cl}(t) = V_4\, \frac{3}{4B} \cos^3\left(\frac{t}{B}\right), \tag{22}$$

where we have made the identifications

$$\frac{t_i}{B} = \frac{i}{s_0 N_4^{1/4}}, \qquad \Delta t_i \sqrt{g_{tt}}\; V_3(t_i) = 2\tilde{C}_4 N_3(i) a^4, \tag{23}$$

such that we have

$$\int dt \sqrt{g_{tt}}\; V_3(t) = V_4. \tag{24}$$

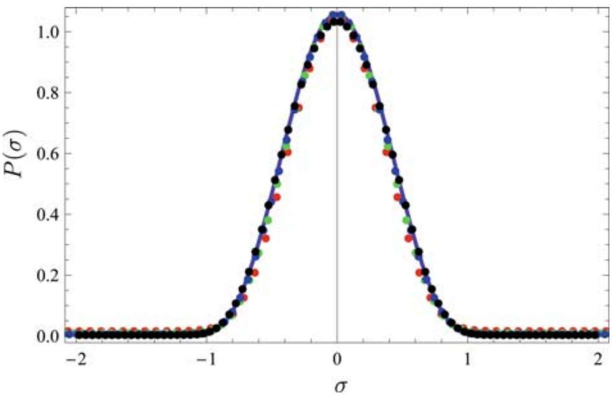


Fig. 2.8 Rescaling of time and volume variables according to relation (20) for $N_4 = 45{,}500$, 91,000, 181,000 and 362,000. The plot also includes the curve (20). More precisely: $\sigma \propto i/N_4^{1/4}$ and $P(\sigma) \propto N_3(i)/N_4^{3/4}$

In (23), $\sqrt{g_{tt}}$ is the constant proportionality factor between the time t and genuine continuum proper time τ, $\tau = \sqrt{g_{tt}}\ t$. (The combination $\Delta t_i \sqrt{g_{tt}} V_3$ contains $\tilde{C}_4$, related to the four-volume of a four-simplex rather than the three-volume corresponding to a tetrahedron, because its time integral must equal V_4). Writing $V_4 = 8\pi^2 R^4/3$, and $\sqrt{g_{tt}} = R/B$, (22) is seen to describe a Euclidean *de Sitter universe* (a four-sphere, the maximally symmetric space for positive cosmological constant) as our searched-for, dynamically generated background geometry! In the parametrization of (22) this is the classical solution to the action

$$S = \frac{1}{24\pi G} \int dt \sqrt{g_{tt}} \left(\frac{g^{tt} \dot{V_3}^2(t)}{V_3(t)} + k_2 V_3^{1/3}(t) - \lambda V_3(t) \right), \tag{25}$$

where $k_2 = 9(2\pi^2)^{2/3}$ and λ is a Lagrange multiplier, fixed by requiring that the total four-volume be V_4, $\int dt \sqrt{g_{tt}}\ V_3(t) = V_4$. Up to an overall sign, this is precisely the Einstein – Hilbert action for the scale factor $a(t)$ of a homogeneous, isotropic universe (rewritten in terms of the spatial three-volume $V_3(t) = 2\pi^2 a(t)^3$), although we of course never put any such simplifying symmetry assumptions into the CDT model.

A discretized, dimensionless version of (25) is

$$S_{\text{discr}} = k_1 \sum_i \left(\frac{(N_3(i+1) - N_3(i))^2}{N_3(i)} + \tilde{k}_2 N_3^{1/3}(i) \right), \tag{26}$$

where $\tilde{k}_2 \propto k_2$. This can be seen by applying the scaling (20), namely, $N_3(i) = N_4^{3/4} n_3(s_i)$ and $s_i = i/N_4^{1/4}$. This enables us to finally conclude that the identifications (23) when used in the action (26) lead naïvely to the continuum expression (25) under the identification

$$G = \frac{a^2}{k_1} \frac{\sqrt{\tilde{C}_4}\, \tilde{s}_0^2}{3\sqrt{6}}. \tag{27}$$

Next, let us comment on the universality of these results. First, we have checked that they are not dependent on the particular definition of time-slicing we have been using, in the following sense. By construction of the piecewise linear CDT geometries we have at each integer time step $t_i = i\, a_t$ a spatial surface consisting of $N_3(i)$ tetrahedra. Alternatively, one can choose as reference slices for the measurements of the spatial volume non-integer values of time, for example, all time slices at discrete times $i - 1/2$, $i = 1, 2, \ldots$. In this case the "triangulation" of the spatial three-spheres consists of tetrahedra – from cutting a (4,1)- or a (1,4)-simplex half-way – and "boxes", obtained by cutting a (2,3)- or (3,2)-simplex (the geometry of this is worked out in [44]). We again find a relation like (20) if we use the total number of spatial building blocks in the intermediate slices (tetrahedra+boxes) instead of just the tetrahedra.

Second, we have repeated the measurements for other values of the bare coupling constants. As long as we stay in the phase where an extended universe is observed, the phase C in Fig. 2.5, a relation like (20) remains valid. In addition, the value of s_0, defined in (20), is almost unchanged until we get close to the phase transition lines beyond which the extended universe disappears. Only for the values of κ_0 around 3.6 and larger will the measured $\langle N_3(t)\rangle$ differ significantly from the value at 2.2. For values larger than 3.8 (at $\Delta = 0.6$), the universe will disintegrate into a number of small and disconnected components distributed randomly along the time axis, and one can no longer fit the distribution $\langle N_3(t)\rangle$ to the formula (20). Later we will show that while s_0 is almost unchanged, the constant k_1 in (26), which governs the quantum fluctuations around the mean value $\langle N_3(t)\rangle$, is more sensitive to a change of the bare coupling constants, in particular, in the case where we change κ_0 (while leaving Δ fixed).

2.3.2 Fluctuations Around de Sitter Space

In the following we will test in more detail how well the actions (25) and (26) describe the computer data. A crucial test is how well it describes the quantum fluctuations around the emergent de Sitter background.

The correlation function (the covariance matrix $\hat{C}$) is defined by

$$C_{N_4}(i,i') = \langle \delta N_3(i)\delta N_3(i')\rangle, \quad \delta N_3(i) \equiv N_3(i) - \bar{N}_3(i), \tag{28}$$

where we have included an additional subscript N_4 to emphasize that N_4 is kept constant in a given simulation.

The first observation extracted from the Monte Carlo simulations is that under a change in the four-volume $C_{N_4}(i,i')$ scales as[4]

$$C_{N_4}(i,i') = N_4\ F\Big(i/N_4^{1/4}, i'/N_4^{1/4}\Big), \tag{29}$$

where F is a universal scaling function. This is illustrated by Fig. 2.9 for the rescaled version of the diagonal part $C_{N_4}^{1/2}(i,i)$, corresponding precisely to the quantum fluctuations $\langle(\delta N_3(i))^2\rangle^{1/2}$ of Fig. 2.7. While the height of the curve in Fig. 2.7 will grow as $N_4^{3/4}$, the superimposed fluctuations will only grow as $N_4^{1/2}$. We conclude that *for fixed bare coupling constants* the relative fluctuations will go to zero in the infinite-volume limit.

Let us rewrite the minisuperspace action (25) for a fixed, finite four-volume V_4 in terms of dimensionless variables by introducing $s = t/V_4^{1/4}$ and $V_3(t) = V_4^{3/4} v_3(s)$:

[4] We stress again that the form (29) is only valid in that part of the universe whose spatial extension is considerably larger than the minimal S^3 constructed from 5 tetrahedra. (The spatial volume of the stalk typically fluctuates between 5 and 15 tetrahedra.)

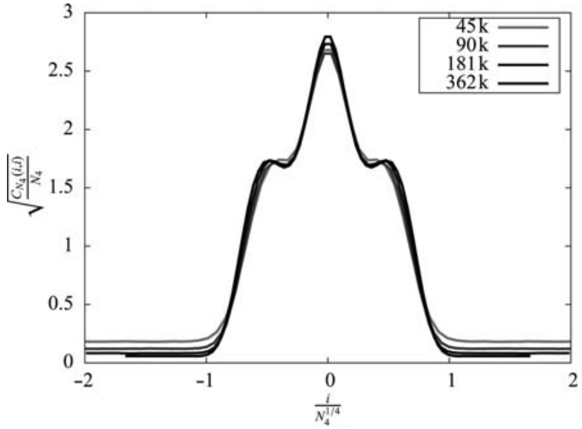


Fig. 2.9 Analyzing the quantum fluctuations of Fig. 2.7: diagonal entries $F(t,t)^{1/2}$ of the universal scaling function F from (29), for $N_4^{(4,1)} = 20{,}000$, 40,000, 80,000 and 160,000

$$S = \frac{1}{24\pi}\,\frac{\sqrt{V_4}}{G}\int ds\sqrt{g_{ss}}\left(\frac{g^{ss}\dot{v_3}^2(s)}{v_3(s)} + k_2 v_3^{1/3}(s)\right), \tag{30}$$

now assuming that $\int ds\sqrt{g_{ss}}\; v_3(s) = 1$, and with $g_{ss} \equiv g_{tt}$. The same rewriting can be done to (26) which becomes

$$S_{discr} = k_1\sqrt{N_4}\sum_i \Delta s\left(\frac{1}{n_3(s_i)}\left(\frac{n_3(s_{i+1}) - n_3(s_i)}{\Delta s}\right)^2 + \tilde{k}_2 n_3^{1/3}(s_i)\right), \tag{31}$$

where $N_3(i) = N_4^{3/4} n_3(s_i)$ and $s_i = i/N_4^{1/4}$.

From the way the factor $\sqrt{N_4}$ appears as an overall scale in (31) it is clear that to the extent a quadratic expansion around the effective background geometry is valid one will have a scaling

$$\langle \delta N_3(i)\delta N_3(i')\rangle = N_4^{3/2}\langle \delta n_3(t_i)\delta n_3(t_{i'})\rangle = N_4 F(t_i, t_{i'}), \tag{32}$$

where $t_i = i/N_4^{1/4}$. This implies that (29) provides additional evidence for the validity of the quadratic approximation and the fact that our choice of action (26) with k_1 independent of N_4 is indeed consistent.

To demonstrate in detail that the full function $F(t,t')$ and not only its diagonal part is described by the effective actions (25), (26), let us for convenience adopt a continuum language and compute its expected behaviour. Expanding (25) around the classical solution according to $V_3(t) = V_3^{cl}(t) + x(t)$, the quadratic fluctuations are given by

$$\langle x(t)x(t')\rangle = \int \mathcal{D}x(s)\; x(t)x(t')\; e^{-\frac{1}{2}\iint dsds' x(s)M(s,s')x(s')}$$
$$= M^{-1}(t,t'), \tag{33}$$

where $\mathcal{D}x(s)$ is the normalized measure and the quadratic form $M(t,t')$ is determined by expanding the effective action S to second order in $x(t)$,

$$S(V_3) = S\left(V_3^{cl}\right) + \frac{1}{18\pi G}\frac{B}{V_4}\int \mathrm{d}t\; x(t)\hat{H}x(t). \tag{34}$$

In expression (34), $\hat{H}$ denotes the Hermitian operator

$$\hat{H} = -\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\cos^3(t/B)}\frac{\mathrm{d}}{\mathrm{d}t} - \frac{4}{B^2\cos^5(t/B)}, \tag{35}$$

which must be diagonalized under the constraint that $\int dt\sqrt{g_{tt}}\; x(t) = 0$, since V_4 is kept constant.

Let $e^{(n)}(t)$ be the eigenfunctions of the quadratic form given by (34) with the volume constraint enforced, ordered according to increasing eigenvalues λ_n. As we will discuss shortly, the lowest eigenvalue is $\lambda_1 = 0$, associated with translational invariance in time direction, and should be left out when we invert $M(t,t')$, because we precisely fix the centre of volume when making our measurements. Its dynamics is therefore not accounted for in the correlator $C(t,t')$.

If this cosmological continuum model were to give the correct description of the computer-generated universe, the matrix

$$M^{-1}(t,t') = \sum_{n=2}^{\infty}\frac{e^{(n)}(t)e^{(n)}(t')}{\lambda_n} \tag{36}$$

should be proportional to the measured correlator $C(t,t')$. Figure 2.10 shows the eigenfunctions $e^{(2)}(t)$ and $e^{(4)}(t)$ (with two and four zeros respectively), calculated from $\hat{H}$ with the constraint $\int dt\sqrt{g_{tt}}\; x(t) = 0$ imposed. Simultaneously we show the corresponding eigenfunctions calculated from the data, i.e. from the matrix $C(t,t')$, which correspond to the (normalizable) eigenfunctions with the highest and third-highest eigenvalues. The agreement is very good, in particular, when taking into consideration that no parameter has been adjusted in the action (we simply take $B = s_0 N_4^{1/4}\Delta t$ in (22) and (34), which gives $B = 14.47a_t$ for $N_4 = 362,000$).

The reader may wonder why the first eigenfunction exhibited has two zeros. As one would expect, the ground state eigenfunction $e^{(0)}(t)$ of the Hamiltonian (35), corresponding to the lowest eigenvalue, has no zeros, but it does not satisfy the volume constraint $\int dt\sqrt{g_{tt}}\; x(t) = 0$. The eigenfunction $e^{(1)}(t)$ of $\hat{H}$ with next-lowest eigenvalue has one zero and is given by the simple analytic function

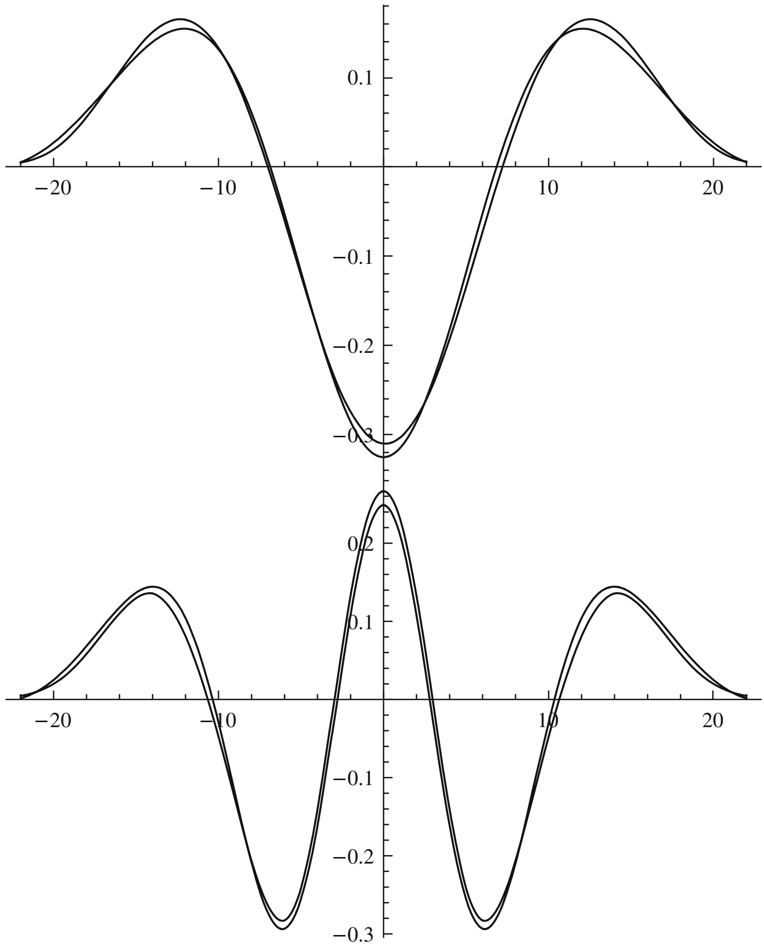


Fig. 2.10 Comparing the two highest even eigenvector of the covariance matrix $C(t, t')$ measured directly (*grey curves*) with the two lowest even eigenvectors of $M^{-1}(t, t')$, calculated semiclassically (*black curves*)

$$e^{(1)}(t) = \frac{4}{\sqrt{\pi B}} \sin\left(\frac{t}{B}\right) \cos^2\left(\frac{t}{B}\right) = c^{-1} \frac{dV_3^{cl}(t)}{dt}, \tag{37}$$

where c is a constant. One realizes immediately that $e^{(1)}$ is the translational zero mode of the classical solution $V_3^{cl}(t)$ ($\propto \cos^3 t/B$). Since the action is invariant under time translations, we have

$$S\left(V_3^{cl}(t + \Delta t)\right) = S\left(V_3^{cl}(t)\right), \tag{38}$$

and since $V_3^{cl}(t)$ is a solution to the classical equations of motion we find to second order (using the definition (37))

$$S\left(V_3^{cl}(t+\Delta t)\right) = S\left(V_3^{cl}(t)\right) + \frac{c^2(\Delta t)^2}{18\pi G}\frac{B}{V_4}\int dt\, e^{(1)}(t)\hat{H}e^{(1)}(t), \tag{39}$$

consistent with $e^{(1)}(t)$ having eigenvalue zero.

It is clear from Fig. 2.10 that some of the eigenfunctions of $\hat{H}$ (with the volume constraint imposed) agree very well with the measured eigenfunctions. All even eigenfunctions (those symmetric with respect to reflection about the symmetry axis located at the centre of volume) turn out to agree very well. The odd eigenfunctions of $\hat{H}$ agree less well with the eigenfunctions calculated from the measured $C(t,t')$. The reason seems to be that we have not managed to eliminate the motion of the centre of volume completely from our measurements. There is an inherent ambiguity in fixing the centre of volume of one lattice spacing, which turns out to be sufficient to reintroduce the zero mode in the data. Suppose we had by mistake misplaced the centre of volume by a small distance Δt. This would introduce a modification

$$\Delta V_3 = \frac{dV_3^{cl}(t)}{dt}\,\Delta t \tag{40}$$

proportional to the zero mode of the potential $V_3^{cl}(t)$. It follows that the zero mode can re-enter whenever we have an ambiguity in the position of the centre of volume. In fact, we have found that the first odd eigenfunction extracted from the data can be perfectly described by a linear combination of $e^{(1)}(t)$ and $e^{(3)}(t)$. It may be surprising at first that an ambiguity of one lattice spacing can introduce a significant mixing. However, if we translate ΔV_3 from (40) to "discretized" dimensionless units using $V_3(i) \sim N_4^{3/4}\cos\left(i/N_4^{1/4}\right)$, we find that $\Delta V_3 \sim \sqrt{N_4}$, which because of $\langle(\delta N_3(i))^2\rangle \sim N_4$ is of the same order of magnitude as the fluctuations themselves. In our case, this apparently does affect the odd eigenfunctions.

One can also compare the data and the matrix $M^{-1}(t,t')$ calculated from (36) directly. This is illustrated in Fig. 2.11, where we have restricted ourselves to data from inside the extended part of the universe. We imitate the construction (36) for M^{-1}, using the data to calculate the eigenfunctions, rather than $\hat{H}$. One could also have used $C(t,t')$ directly, but the use of the eigenfunctions makes it somewhat easier to perform the restriction to the bulk. The agreement is again good (better than 15% at any point on the plot), although less spectacular than in Fig. 2.10 because of the contribution of the odd eigenfunctions to the data.

2.3.3 The Size of the Universe and the Flow of G

It is natural to view the coupling constant G in front of the effective action for the scale factor as the gravitational coupling constant G. The effective action which

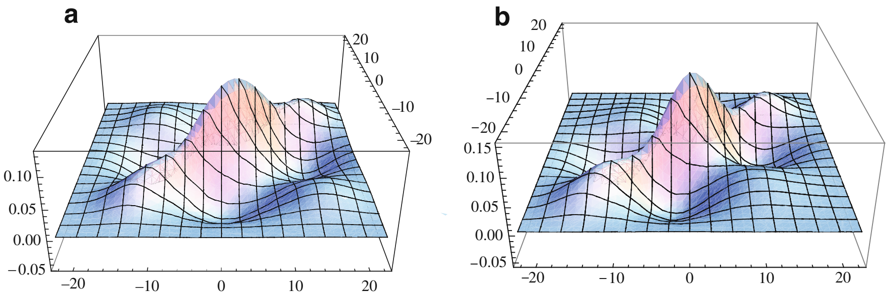


Fig. 2.11 Comparing data for the extended part of the universe: measured $C(t,t')$ (*above*) versus $M^{-1}(t,t')$ obtained from analytical calculation (*below*). The agreement is good, and would have been even better had we included only the even modes

described our computer-generated data was given by (25) and its dimensionless lattice version by (26). The computer data allows us to extract $k_1 \propto a^2/G$, a being the spatial lattice spacing, the precise constant of proportionality being given by (27):

$$G = \frac{a^2}{k_1} \frac{\sqrt{\tilde{C}_4}\,\tilde{s}_0^2}{3\sqrt{6}}. \tag{41}$$

For the bare coupling constants $(\kappa_0, \Delta) = (2.2, 0.6)$ we have high-statistics measurements for N_4 ranging from 45,500 to 362,000 four-simplices (equivalently, $N_4^{(4,1)}$ ranging from 20,000 to 160,000 four-simplices). The choice of Δ determines the asymmetry parameter α, and the choice of (κ_0, Δ) determines the ratio ξ between $N_4^{(3,2)}$ and $N_4^{(4,1)}$. This in turn determines the "effective" four-volume $\tilde{C}_4$ of an average four-simplex, which also appears in (41). The number $\tilde{s}_0$ in (41) is determined directly from the time extension T_{univ} of the extended universe according to

$$T_{\text{univ}} = \pi\ \tilde{s}_0 \left(N_4^{(4,1)}\right)^{1/4}. \tag{42}$$

Finally, from our measurements we have determined $k_1 = 0.038$. Taking everything together according to (41), we obtain $G \approx 0.23a^2$, or $\ell_{Pl} \approx 0.48a$, where $\ell_{Pl} = \sqrt{G}$ is the Planck length.

From the identification of the volume of the four-sphere, $V_4 = 8\pi^2 R^4/3 = \tilde{C}_4 N_4^{(4,1)} a^4$, we obtain that $R = 3.1a$. In other words, *the linear size πR of the quantum de Sitter universes studied here lies in the range of 12–21 Planck lengths for N_4 in the range mentioned above and for the bare coupling constants chosen as* $(\kappa_0, \Delta) = (2.2, 0.6)$.

Our dynamically generated universes are therefore not very big, and the quantum fluctuations around their average shape are large as is apparent from Fig. 2.7. It is rather surprising that the semi-classical minisuperspace formulation is applicable for universes of such a small size, a fact that should be welcome news to anyone performing semi-classical calculations to describe the behaviour of the early universe. However, in a certain sense our lattices are still coarse compared to the Planck scale ℓ_{Pl} because the Planck length is roughly half a lattice spacing. If we are after a theory of quantum gravity valid on all scales, we are in particular interested in uncovering phenomena associated with Planck-scale physics. In order to collect data free from unphysical short-distance lattice artefacts at this scale, we would ideally like to work with a lattice spacing much smaller than the Planck length, while still being able to set by hand the physical volume of the universe studied on the computer.

The way to achieve this, under the assumption that the coupling constant G of formula (41) is indeed a true measure of the gravitational coupling constant, is as follows. We are free to vary the discrete four-volume N_4 and the bare coupling constants (κ_0, Δ) of the Regge action (see [14] for further details on the latter). Assuming for the moment that the semi-classical minisuperspace action is valid, the effective coupling constant k_1 in front of it will be a function of the bare coupling constants (κ_0, Δ), and can in principle be determined as described above for the case $(\kappa_0, \Delta) = (2.2, 0.6)$. If we adjusted the bare coupling constants such that in the limit as $N_4 \to \infty$ both

$$V_4 \sim N_4 a^4 \quad \text{and} \quad G \sim a^2 / k_1(\kappa_0, \Delta) \tag{43}$$

remained constant (i.e. $k_1(\kappa_0, \Delta) \sim 1/\sqrt{N_4}$), we would eventually reach a region where the Planck length was significantly smaller than the lattice spacing a, in which event the lattice could be used to approximate spacetime structures of Planckian size and we could initiate a genuine study of the sub-Planckian regime. Since we have no control over the effective coupling constant k_1, the first obvious question which arises is whether we can at all adjust the bare coupling constants in such a way that at large scales we still see a four-dimensional universe, with k_1 going to zero at the same time. The answer seems to be in the affirmative, as we will go on to explain.

Figure 2.12 shows the results of extracting k_1 for a range of bare coupling constants for which we still observe an extended universe. In the top figure $\Delta = 0.6$ is kept constant while κ_0 is varied. For κ_0 sufficiently large we eventually reach a point where a phase transition takes place (the point in the square in the bottom right-hand corner is the measurement closest to the transition we have looked at). For even larger values of κ_0, beyond this transition, the universe disintegrates into a number of small universes, in a CDT analogue of the branched-polymer phase of Euclidean quantum gravity. The plot shows that the effective coupling constant k_1 becomes smaller and possibly goes to zero as the phase transition point is approached, although our current data do not yet allow us to conclude that k_1 does indeed vanish at the transition point.

Conversely, the bottom figure of Fig. 2.12 shows the effect of varying Δ, while keeping $\kappa_0 = 2.2$ fixed. As Δ is decreased towards 0, we eventually hit another phase transition, separating the physical phase of extended universes from the CDT equivalent of the crumpled phase of Euclidean quantum gravity, where the entire universe will be concentrated within a few time steps, as already mentioned above.

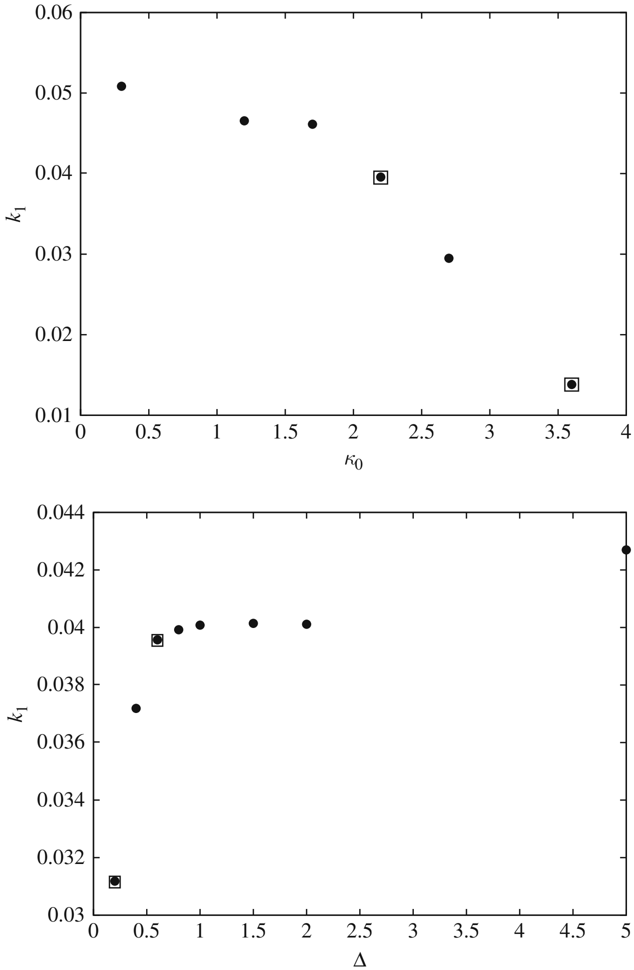


Fig. 2.12 The measured effective coupling constant k_1 as function of the bare κ_0 (*top*, $\Delta = 0.6$ fixed) and the asymmetry Δ (*bottom*, $\kappa_0 = 2.2$ fixed). The marked point near the middle of the data points sampled is the point $(\kappa_0, \Delta) = (2.2, 0.6)$ where most measurements in the remainder of the paper were taken. The other marked points are those closest to the two phase transitions, to the "branched-polymer phase" (*top*) and the "crumpled phase" (*bottom*)

(The point closest to the transition where we have taken measurements is the one in the bottom left-hand corner.) Also when approaching this phase transition the effective coupling constant k_1 goes to 0, leading to the tentative conclusion that $k_1 \to 0$ along the entire phase boundary.

However, to extract the coupling constant G from (41) we have to take into account not only the change in k_1, but also that in $\tilde{s}_0$ (the width of the distribution $N_3(i)$) and in the effective four-volume $\tilde{C}_4$ as a function of the bare coupling constants. Combining these changes, we arrive at a slightly different picture. Approaching the boundary where spacetime collapses in time direction (by lowering Δ), the gravitational coupling constant G *decreases*, despite the fact that $1/k_1$ increases. This is a consequence of $\tilde{s}_0$ decreasing considerably. On the other hand, when (by increasing κ_0) we approach the region where the universe breaks up into several independent components, the effective gravitational coupling constant G increases, more or less like $1/k_1$, where the behaviour of k_1 is shown in Fig. 2.12 (top). This implies that the Planck length $\ell_{Pl} = \sqrt{G}$ increases from approximately $0.48a$ to $0.83a$ when κ_0 changes from 2.2 to 3.6. Most likely we can make it even bigger in terms of Planck units by moving closer to the phase boundary.

On the basis of these arguments, it seems likely that the non-perturbative CDT formulation of quantum gravity does allow us to penetrate into the sub-Planckian regime and probe the physics there explicitly. Work in this direction is currently ongoing. One interesting issue under investigation is whether and to what extent the simple minisuperspace description remains valid as we go to shorter scales. We have already seen deviations from classicality at short scales when measuring the spectral dimension [14, 43], and one would expect them to be related to additional terms in the effective action (25) and/or a nontrivial scaling behaviour of k_1. This raises the interesting possibility of being able to test explicitly the scaling violations of G predicted by renormalization group methods in the context of asymptotic safety [2–7].

2.4 Two-Dimensional Euclidean Quantum Gravity

The results described above are of course interesting and suggest that there might exist a field theory of quantum gravity in four dimensions (three space and one time dimension). However, the results are based on numerical simulations. As already mentioned it is of great conceptional interest that we have a toy model, two-dimensional quantum gravity, where both the lattice theory and the continuum quantum gravity theory can be solved analytically and agree. Of course we can still be in the situation that there exists no description of quantum gravity as a field theory in four dimensions (although we have presented some evidence in favour of such a scenario above), but we can then not blame the underlying formalism for being inadequate.

2.4.1 Continuum Formulation

Let M^h denote a closed, compact, connected and orientable surface of genus h and Euler characteristic $\chi(h) = 2 - 2h$. The partition function of two-dimensional Euclidean quantum gravity is formally given by

$$Z(\Lambda, G) = \sum_{h=0}^{\infty} \int \mathcal{D}[g]\, e^{-S(g;\Lambda,G)}, \tag{44}$$

where Λ denotes the cosmological constant, G is the gravitational coupling constant and S is the continuum Einstein – Hilbert action defined by

$$S(g;\,\Lambda, G) = \Lambda \int_{M^h} d^2\xi \sqrt{g} - \frac{1}{2\pi G} \int_{M^h} d^2\xi \sqrt{g}\, R. \tag{45}$$

In (44), we take the sum to include all possible topologies of two-dimensional manifolds (i.e. over all genera h), and in (45) R denotes the scalar curvature of the metric g on the manifold M^h. The functional integration is over all *diffeomorphism equivalence classes* $[g]$ of metrics on M^h.

In two dimensions the curvature part of the Einstein – Hilbert action is a topological invariant according to the Gauss – Bonnet theorem, which allows us to write

$$Z(\Lambda, G) = \sum_{h=0}^{\infty} e^{\chi(h)/G} Z_h(\Lambda), \tag{46}$$

where

$$Z_h(\Lambda) = \int \mathcal{D}[g]\, e^{-S(g;\Lambda)} \tag{47}$$

and

$$S(g;\,\Lambda) = \Lambda\, V_g, \tag{48}$$

where $V_g = \int d^2\xi \sqrt{g}$ is the volume of the universe for a given diffeomorphism class of metrics. In the remainder of this section we will, for simplicity, restrict our attention to manifolds homeomorphic to S^2 or S^2 with a fixed number of holes unless explicitly stated otherwise. In this case we disregard the topological term in the action since it is a constant. The sphere S^2 with b boundary components will be denoted S_b^2 and we denote the partition function for the sphere, $Z_0(\Lambda)$ in (47), by $Z(\Lambda)$.

In the presence of a boundary it is natural to add to the action a boundary term

$$S(g;\,\Lambda, Z_1, ..., Z_b) = \Lambda V_g + \sum_{i=1}^{b} Z_i L_{i,g}, \tag{49}$$

where $L_{i,g}$ denotes the length of the ith boundary component with respect to the metric g. We refer to the Z_i's as the cosmological constants of the boundary components. The partition function is in this case given by

$$W(\Lambda; Z_1, ..., Z_b) = \int \mathcal{D}[g]\, e^{-S(g;\Lambda,Z_1,...,Z_b)}. \tag{50}$$

Since the lengths of the boundary components are invariant under diffeomorphisms, it makes sense to fix them to values $L_1, \ldots, L_b$ and define the *Hartle – Hawking wave functionals* by

$$W(\Lambda; L_1, ..., L_b) = \int \mathcal{D}[g]\, e^{-S(g;\Lambda)} \prod_{i=1}^{b} \delta(L_i - L_{i,g}), \tag{51}$$

where $S(g; \Lambda)$ is given by (48). Since (50) is the Laplace transform of (51), i.e.

$$W(\Lambda; Z_1, ..., Z_b) = \int_0^\infty \prod_{i=1}^{b} dL_i\, e^{-Z_i L_i}\, W(\Lambda; L_1, ..., L_b), \tag{52}$$

we denote them by the same symbol. We distinguish between the two by the names of the arguments.

2.4.2 The Lattice Regularization

At the outset we restrict the topology of surfaces to be that of S^2 with a fixed number of holes. We view abstract triangulations of S^2_b as defining a grid in the space of diffeomorphism equivalence classes of metrics on S^2_b. Each triangle is a "building block" with side lengths a. This a will be an UV cut-off which we will relate to the bare coupling constants on the lattice. However, presently it is convenient to view a as being 1 (length unit).

Let T denote a triangulation of S^2_b. The regularized theory of gravity will be defined by replacing the action $S_g(\Lambda, Z_1, \ldots, Z_b)$ in (49) by

$$S_T(\mu, \lambda_1, \ldots, \lambda_b) = \mu N_t + \sum_{i=1}^{b} \lambda_i l_i, \tag{53}$$

where N_t denotes the number of triangles in T and l_i is the number of links in the ith boundary component. The parameter μ is the bare cosmological constant and the λ_i's are the bare cosmological constants of the boundary components. The integration over diffeomorphism equivalence classes of metrics in (50) becomes a summation over non-isomorphic triangulations. We define the loop functions (discretized versions of $W(\Lambda, Z_1, \ldots, Z_b)$) by summing over all triangulations of S^2_b:

$$w(\mu, \lambda_1, \ldots, \lambda_b) = \sum_{l_1,\ldots,l_b} \sum_{T \in \mathcal{T}(l_1,\ldots,l_b)} e^{-S_T(\mu,\lambda_1,\ldots,\lambda_b)}. \tag{54}$$

Analogously, we define the partition function for closed surfaces by

$$Z(\mu) = \sum_{T \in \mathcal{T}} \frac{1}{C_T} \, e^{-S_T(\mu)}, \tag{55}$$

where C_T is the symmetry factor of T and $S_T(\mu) = \mu N_t$. Since we consider surfaces of a fixed topology we have left out the curvature term in the action. It will be introduced later, when the restriction on topology is lifted.

Next we write down the regularized version of the Hartle – Hawking wave functionals $W(\Lambda, L_1, \ldots, L_b)$:

$$w(\mu, l_1, \ldots, l_b) = \sum_{T \in \mathcal{T}(l_1, \ldots, l_b)} e^{-S_T(\mu)} \tag{56}$$

with an abuse of notation similar to the one in the previous section. This can also be written in the form

$$w(\mu, l_1, \ldots, l_b) = \sum_k e^{-\mu k} w_{k, l_1, \ldots, l_b} \ , \tag{57}$$

where we have introduced the notation

$$w_{k, l_1, \ldots, l_b}$$

for the number of triangulations in $\mathcal{T}(l_1, \ldots, l_b)$ with k triangles.

The discretized analogues of the Laplace transformations which relate $W(\Lambda, Z_1, \ldots, Z_b)$ and $W(\Lambda, L_1, \ldots, L_b)$ are

$$w(\mu, \lambda_1, \ldots, \lambda_b) = \sum_{l_1, \ldots, l_b} e^{-\sum_i \lambda_i l_i} w(\mu, l_1, \ldots, l_b) \tag{58}$$

and (57). Similarly, we have for the partition functions

$$Z(\mu) = \sum_k e^{-\mu k} Z(k), \tag{59}$$

$$Z(k) = \sum_{T \in \mathcal{T}, N_t = k} \frac{1}{C_T}. \tag{60}$$

It follows from the definitions (57), (58) that $w(\mu, \lambda_1, \ldots, \lambda_b)$ is the generating function for the numbers $w_{k, l_1, \ldots, l_b}$, the arguments of the generating function being $e^{-\mu}$ and $e^{-\lambda_i}$. *In this way the evaluation of the loop functions of two-dimensional quantum gravity is reduced to the purely combinatorial problem of finding the number of non-isomorphic triangulations of S^2 or S^2_b with a given number of triangles and boundary components of given lengths.*

We use the notation

$$w(g, z_1, \ldots, z_b) = \sum_{k,l_1,\ldots,l_b} w_{k,l_1,\ldots,l_b}\, g^k z_1^{-l_1-1} \cdots z_b^{-l_b-1} \tag{61}$$

for the generating function with an extra factor $z_1^{-1} \ldots z_b^{-1}$, i.e. we make the identifications

$$g = e^{-\mu}, \quad z_i = e^{\lambda_i}. \tag{62}$$

The reason for this particular choice of variables in the generating function is motivated by its analytic structure, which will be revealed below.

In the following we consider a particular class of triangulations which includes degenerate boundaries. It may be defined as the class of complexes homeomorphic to the sphere with a number of holes that one obtains by successively gluing together a collection of triangles and a collection of double links which we consider as (infinitesimally narrow) strips, where links, as well as triangles, can be glued onto the boundary of a complex both at vertices and along links. Gluing a double link along a link makes no change in the complex. An example of such a complex is shown in Fig. 2.13. The reason we use this class of triangulations is that they match the "triangulations" we obtain from the so-called matrix models to be considered below. We call this class of complexes "unrestricted triangulations".

One could have chosen a more regular class of triangulations, corresponding more closely to our intuitive notion of a surface. However, the degenerate structures present in the unrestricted triangulations appear on a slightly larger scale in the regular triangulations in the form of narrow strips consisting of triangles. Since we want to take the lattice side a of a triangle to zero in the continuum limit, there should be no difference in that limit between various classes of triangulations, unless more severe constraints are introduced. We say that "the continuum limit is universal". But at some point the constraint can be so strong that the continuum limit is changed. We will meet precisely such a change below, leading from (Euclidean) dynamical triangulations (DT) to causal dynamical triangulations (CDT).

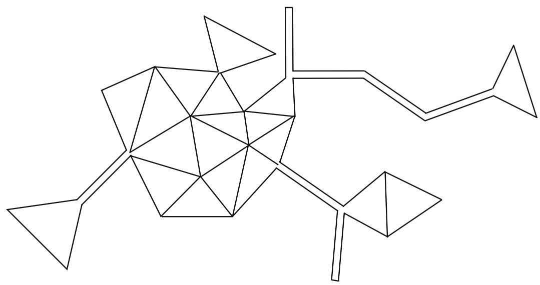


Fig. 2.13 A typical unrestricted "triangulation"

Let $w(g, z)$ denote the generating function for the (unrestricted) triangulations with one boundary component. Then we have

$$w(g, z) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} w_{k,l}\, g^k\, z^{-(l+1)} \equiv \sum_{l=0}^{\infty} \frac{w_l(g)}{z^{l+1}}. \tag{63}$$

We have included the triangulation consisting of one point. It gives rise to the term $1/z$ and we have $w_0(g) = 1$. The function $w_1(g)$ starts with the term g, which corresponds to an unrestricted triangulation with a boundary consisting of one (closed) link with one vertex and containing one triangle. The coefficients $w_{k,1}$ in the expansion $w_{1,1}g + w_{3,1}g^3 + \cdots$ of $w_1(g)$ are the numbers of unrestricted triangulations with a boundary consisting of one link.

The coefficients of $w(g, z)$ fulfil a recursion relation which has the simple graphical representation shown in Fig. 2.14. The diagrams indicate two operations that one can perform on a marked link on the boundary to produce a triangulation which has either fewer triangles or fewer boundary links. The first term on the right-hand side of Fig. 2.14 corresponds to the removal of a triangle. The second term corresponds to the removal of a double link. Note that removing a triangle creates a new double link if the triangle has two boundary links. In addition, note that we count triangulations *with one marked link on each boundary component* and adopt the notation introduced above for the corresponding quantities.

The equation associated with the diagrams is

$$\left[w(g, z)\right]_{k,l} = \left[gzw(g, z)\right]_{k,l} + \left[\frac{1}{z} w^2(g, z)\right]_{k,l}. \tag{64}$$

The subscripts k, l indicate the coefficient of g^k/z^{l+1}. Let us explain the equation in some detail. The factor gz in (64) is present since the triangulation corresponding to the first term on the right-hand side of Fig. 2.14 has one triangle less and one boundary link more than the triangulation on the left-hand side. The function $w^2(g, z)$ in the last term in (64) arises from the two blobs connected by the double link in Fig. 2.14 and the $1/z$ in front of $w^2(g, z)$ is inserted to make up for the decrease by two in the length of the boundary when removing the double link.

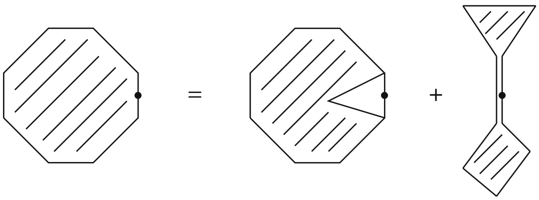


Fig. 2.14 Graphical representation of relation (64): The boundary contains one marked link which is part of a triangle or a double link. Associated to each triangle is a weight g, and to each double link a weight 1

As the reader may have discovered, (64) is not correct for the smallest values of l. Consider Fig. 2.14. The first term on the left-hand side of (64) (a single vertex) has no representation on the diagram. In order for (64) to be valid for $k = l = 0$ we have to add the term $1/z$ on the right-hand side of (64). Furthermore, it is clear from Fig. 2.14 that the first term on the right-hand side has at least two boundary links. Consequently, the term $gzw(g, z)$ on the right-hand side of (64) should be replaced by $gz(w(g, z) - 1/z - w_1(g)/z^2)$ such that all terms corresponding to triangulations with boundaries of length 0 and 1 are subtracted. It follows that the correct equation is

$$(z - gz^2)w(g, z) - 1 + g(w_1(g) + z) = w^2(g, z). \tag{65}$$

We will refer to (65) as the *loop equation*. It is a second-order equation in $w(g, z)$. As will be clear in the following this algebraic feature allows us to extract asymptotic formulas for the number of triangulations with k triangles in the limit $k \to \infty$.

2.4.3 Counting Graphs

Let us begin by solving (65) in the limit $g = 0$. In this case there are no internal triangles and the triangulations are in one-to-one correspondence with rooted branched polymers.[5] The double links correspond to the links of the branched polymers and the root is the marked link, see Fig. 2.15. If $g = 0$ then (65) reads

$$w^2(z) - zw(z) + 1 = 0. \tag{66}$$

The above equation has two solutions. The one that corresponds to the counting problem has a Taylor expansion in z^{-1} whose first term is z^{-1} (recall that $w_{0,0} = 1$). This solution is given by

$$w(z) = \frac{1}{2}\left(z - \sqrt{z^2 - 4}\right). \tag{67}$$

Expanding in powers of $1/z$ yields

---

[5] One might think that such polymers are not relevant at all for studying real surfaces made of triangles, not to mention higher piecewise linear manifolds, but in fact the branched-polymer structure is quite generic. Surfaces or higher-dimensional manifolds can "pinch", such that two parts of the triangulation are only connected by a minimal "neck". If this happens in many places one can effectively obtain a branched-polymer structure even for higher-dimensional piecewise linear manifolds. Such minimal necks have been used to measure critical exponents of various ensembles of piecewise linear manifolds [45, 46] and in four-dimensional Euclidean quantum gravity one has indeed, as mentioned above, observed a phase where the four-dimensional piecewise linear manifolds degenerate to branched polymers [40, 41]. The same is the case for bosonic strings with central charge $c > 1$ [47].

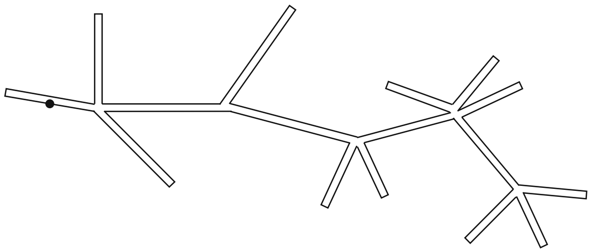


Fig. 2.15 Rooted branched polymers created by gluing of a boundary with one marked link

$$w(z) = \sum_{l=0}^{\infty} \frac{w_{2l}}{z^{2l+1}}, \tag{68}$$

where

$$w_{2l} = \frac{(2l)!}{(l+1)!\,l!} = \frac{1}{\sqrt{\pi}} l^{-3/2}\, 4^l \,(1 + O(1/l)) \tag{69}$$

and w_{2l} is the number of rooted polymers with l links. Note that w_{2l} are the Catalan numbers, known from many combinatorial problems.

The generating function $w(z)$ is analytic in the complex plane C with a cut on the real axis along the interval $[-2, 2]$. The endpoints of the cut determine the radius of convergence of $w(z)$ as a function of $1/z$ or, equivalently, the exponential growth of w_{2l}.

We can solve the second-order equation (65) and obtain

$$w(g, z) = \frac{1}{2}\left(V'(z) - \sqrt{(V'(z))^2 - 4Q(z)}\right), \tag{70}$$

where, anticipating generalizations, we have introduced the notation

$$V'(z) = z - gz^2, \quad Q(z) = 1 - gw_1(g) - gz. \tag{71}$$

The sign of the square root is determined as in (67) by the requirement that $w(g, z) = 1/z + O(1/z^2)$ for large z (since $w_{0,0} = 1$). If $g = 0$ then $V'(z)^2 - 4Q(z) = z^2 - 4$. For $g > 0$, on the other hand, $V'(z)^2 - 4Q(z)$ is a fourth-order polynomial of the form

$$\begin{aligned} V'(z)^2 - 4Q(z) = &\{z - (2 + 2g) + O(g^2)\} \\ &\times\{z + (2 - 2g) + O(g^2)\}\{gz - (1 - 2g^2) + O(g^3)\}^2 \end{aligned} \tag{72}$$

in a neighbourhood of $g = 0$ since the analytic structure of $w(g, z)$ as a function of z cannot change discontinuously at $g = 0$. We can therefore write

$$V'(z)^2 - 4Q(z) = (z - c_+(g))(z - c_-(g))(c_2(g) - gz)^2, \tag{73}$$

and, by (70),

$$w(g, z) = \frac{1}{2}\left(z - gz^2 + (gz - c_2)\sqrt{(z - c_+)(z - c_-)}\right), \tag{74}$$

where c_-, c_+ and c_2 are functions of g, analytic in a neighbourhood of $g = 0$. We label the roots so that $c_- \leq c_+$. *The numbers c_-, c_+ and c_2 are uniquely determined by the requirement that $w(g, z) = 1/z + O(1/z^2)$, again originating from $w_{0,0} = 1$.* This requirement gives three equations for the coefficients of z, z^0, z^{-1}.

We can generalize the above counting problem to planar complexes made up of polygons with an arbitrary number $j \leq n$ of sides, including "one-sided" and "two-sided" polygons. If we attribute a weight gt_j to each j-sided polygon and a weight z to each boundary link, and adopt the notation

$$V'(z) = z - g(t_1 + t_2 z + t_3 z^2 + \cdots + t_n z^{n-1}), \tag{75}$$

$$Q(z) = 1 - g\sum_{j=2}^{n} t_j \sum_{l=0}^{j-2} z^l w_{j-2-l}(g), \quad w_0(g) = 1, \tag{76}$$

the analogue of (65) is

$$w(g, z)^2 = V'(z)w(g, z) - Q(z) \tag{77}$$

or

$$\begin{aligned} w(g, z) = g\left(t_1\frac{1}{z} + t_2 + t_3 z + \cdots + t_n z^{n-2}\right) w(g, z) \\ + \frac{1}{z}Q(z) + \frac{1}{z}w^2(g, z). \end{aligned} \tag{78}$$

The graphical representation of (78) is shown in Fig. 2.16. The subtraction of the polynomial $Q(z)$ in (77) reflects the fact that the term with a j-sided polygon in Fig. 2.16 must have a boundary of length at least $j - 1$ for $j > 1$. The constant term 1 in Q corresponds to the complex consisting of a single vertex.

The solution can be written as

$$w(g, z) = \frac{1}{2}\left(V'(z) - M(z)\sqrt{(z - c_+(g))(z - c_-(g))}\right), \tag{79}$$

where $M(z)$ is a polynomial of a degree which is one less than that of $V'(z)$. Again, the polynomial M is uniquely determined by the requirement that $w(g, z)$ falls off at infinity as before, i.e. $w(g, z) = 1/z + O(1/z^2)$, and the additional requirement that $w(z)$ has a single cut. It is sometimes convenient to write (79) as

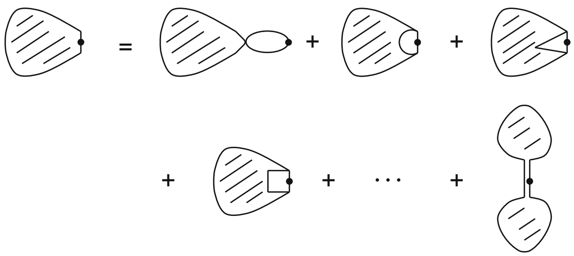


Fig. 2.16 Graphical representation of relation (78): the marked link of the boundary belongs either to an i-gon (associated weight $g\,t_i$) or a double link (associated weight 1). It is also a graphical representation of (114) if instead of weight 1 we associate a weight g_s to the marked double link

$$w(g,z)=\frac{1}{2}\left(V'(z)-\sum_{k=1}^{n-1}M_k(g)(z-c_+)^{k-1}\sqrt{(z-c_+(g))(z-c_-(g))}\right). \quad (80)$$

One can show the following: for $t_1,\dots,t_{n-1}\geq 0$ and $t_n>0$ one has

$$M_k(g)<0, \qquad k<n, \quad (81)$$

while

$$M_1(g)>0 \quad (82)$$

in a neighbourhood of $g=0$. When we increase g we first reach a point g_c, where (81) is still satisfied but

$$M_1(g_c)=0. \quad (83)$$

The coupling constant point g_c is thus the point where the analytical structure of $w(g,z)$ changes from being identical to that of the branched polymer, i.e. it behaves like $(z-c_+)^{1/2}(z-c_-)^{1/2}$, to $(z-c_+)^{3/2}(z-c_-)^{1/2}$. The function $w(g,z)$ is an analytic function around the point $g=0$ ($w(g=0,z)$ is the branched-polymer partition function discussed above). The radius of convergence is precisely g_c. If we return to the expansion in (64) each term $w_l(g)$ has this radius of convergence. It is the generating function for triangulations with one boundary consisting of l links. The singularity of $w_l(g)$ for $g\to g_c$ determines the asymptotic behaviour of the number of such triangulations for a large number of triangles, i.e. the leading behaviour of the numbers $w_{k,l}$ for large k (see (89) and (90) below).

Let us introduce $g_j=gt_j$ as new variables. We have

$$w(g_i,z)=\sum_{l,k_1,\dots,k_n} w_{\{k_j\},l}\,z^{-(l+1)}\prod_{j=1}^{n} g_j^{k_j}, \quad (84)$$

where $w_{\{k_j\},l}$ is the number of planar graphs with k_j j-sided polygons, $j = 1, \ldots, n$, and a boundary of length l.

From $w(g_i, z)$ we can derive the generating function for planar graphs with two boundary components by applying the *loop insertion operator*

$$\frac{d}{dV(z)} = \sum_{j=1}^{\infty} \frac{j}{z^{j+1}} \frac{d}{dg_j}. \tag{85}$$

One should think of this operator as acting on formal power series in an arbitrary number of variables g_j. The action of $d/dV(z_2)$ on $w(g_i, z_1)$ has in each term of the power series the effect of reducing the power k_j of a specific coupling constant g_j by one and adding a factor jk_j/z_2^{j+1}. The geometrical interpretation is that a j-sided polygon is removed, leaving a marked boundary of length j to which the new indeterminate z_2 is associated. The factor k_j is due to the possibility to make the replacement at any of the k_j j-sided polygons present in the planar graph, while j is the number of possibilities to choose the marked link on the new boundary component. The generating function for planar graphs with b boundary components can therefore be expressed as

$$w(g_i, z_1, \ldots, z_b) = \frac{d}{dV(z_b)} \cdots \frac{d}{dV(z_2)} w(g_i, z_1). \tag{86}$$

A most remarkable result is the following: for any potential $V(g_i, z)$ the two-loop function $w(g_i, z_1, z_2)$ has the form

$$\begin{aligned} &w(g_i, z_1, z_2) \\ &= \frac{1}{2(z_1 - z_2)^2} \left(\frac{z_1 z_2 - \frac{1}{2}(z_1 + z_2)(c_+ + c_-) + c_+ c_-}{\sqrt{[(z_1 - c_+)(z_1 - c_-)][(z_2 - c_+)(z_2 - c_-)]}} - 1 \right). \end{aligned} \tag{87}$$

Note that there is no explicit reference to the potential $V(g_i, z)$, but of course c_+ and c_- depend on the potential.

From this formula one can in principle construct the multi-loop function $w(g, z_1, \ldots, z_b)$ by applying the loop insertion operator $b - 2$ times. One can use this formula to find the leading singularity of $w(g, z_1, \ldots, z_b)$ when $g \to g_c$, the critical value of the coupling constant g and the value where $M_1(g) = 0$. One finds

$$w(g, z_1, \ldots, z_b) \sim \left(\frac{1}{\sqrt{g_c - g}} \right)^{2b-5}, \tag{88}$$

as $g \to g_c$. This implies that the generating function $w(g, l_1, \ldots, l_b)$ for the number of triangulations, $w_{k,l_1,\ldots,l_b}$, constructed from k triangles with b boundary components of length $l_1, \ldots, l_b$, has a singularity as $g \to g_c$ that is independent of the length of the boundary components and is given by

$$w_h(g, l_1, \ldots, l_b) \sim \left(\frac{1}{\sqrt{g_c - g}}\right)^{2b-5}. \tag{89}$$

Finally, we obtain from (89) the asymptotic behaviour of $w_{k,l_1,\ldots,l_b}$ as $k \to \infty$:

$$w_{k,l_1,\ldots,l_b} \sim \left(\frac{1}{g_c}\right)^k k^{-\frac{5}{2}+b-1}. \tag{90}$$

We note that these results can be generalized to triangulations which have h handles:

$$w_h(g, l_1, \ldots, l_b) \sim \left(\frac{1}{\sqrt{g_c - g}}\right)^{2b+(h-1)5}. \tag{91}$$

$$w^{(h)}_{k,l_1,\ldots,l_b} \sim \left(\frac{1}{g_c}\right)^k k^{(h-1)\frac{5}{2}+b-1}. \tag{92}$$

For future applications it is important to note that the position of the leading singularity g_c in (91) or, alternatively, the exponential growth of the number of triangles in (92) is independent of the number of handles or the number of boundaries.

2.4.4 The Continuum Limit

We now show how continuum physics is related to the asymptotic behaviour of $w^{(h)}_{k,l_1,\ldots,l_b}$ for $k \to \infty$ *and* $l_1, \ldots, l_b \to \infty$ in a specific way, and we use the results for the generating functions $w_h(g, z_1, \ldots, z_b)$ derived in the previous sections to study this limit.

Before discussing details it is useful to clarify how we expect the continuum wave functionals $W(\Lambda, Z_1, \ldots, Z_b)$ to renormalize. Since the cosmological constants Λ and Z_i have dimensions $1/a^2$ and $1/a$, respectively, a being the length of the lattice cut-off, it is natural to expect that they are subject to an additive renormalization

$$\Lambda_c = \frac{\mu_c}{a^2} + \Lambda, \quad Z_{i,c} = \frac{\lambda_{i,c}}{a} + Z_i, \tag{93}$$

where Λ_c and $Z_{i,c}$ are the *bare* cosmological coupling constants. Since our regularization is represented in terms of discretized two-dimensional manifolds, the bare cosmological constants should be related to the dimensionless coupling constants μ, λ_i by

$$\Lambda_c = \frac{\mu}{a^2}, \quad Z_{i,c} = \frac{\lambda_i}{a}, \tag{94}$$

so that (93) can be written as

$$\mu - \mu_c = a^2\Lambda, \qquad \lambda_i - \lambda_{i,c} = aZ_i. \tag{95}$$

In the following we assume for simplicity that all the $\lambda_{i,c}$s are equal to λ_c. We identify the constants μ_c and λ_c with the critical couplings g_c and $c_+(g_c)$ via the relations

$$\frac{1}{g_c} = e^{\mu_c}, \qquad c_+(g_c) = e^{\lambda_c}. \tag{96}$$

Recalling the relation (62) between μ, g and z, λ, it follows that the $a \to 0$ limit of the functions $w(\mu, \lambda_1, \ldots, \lambda_b)$ is determined by their singular behaviour at g_c. The renormalization (93) has the effect of cancelling the exponential entropy factor for the triangulations, see (90). Note that since we have the same exponential factors for all genera, we expect the renormalization of the cosmological constants to be independent of genus.

We begin by studying the continuum limit for planar surfaces. Then we will discuss how to take higher genera into account, thereby reintroducing the gravitational coupling constant G and also discussing its renormalization. This will lead us to the so-called *double-scaling limit*.

We are interested in a limit of the discretized models where the length a of the links goes to zero while the number k of triangles and the lengths l_i of the boundary components go to infinity in such a way that

$$V = ka^2 \quad \text{and} \quad L_i = l_i a \tag{97}$$

remain finite. The asymptotic behaviour of $w_{k,l_1,\ldots,l_b}$ is given by (90) if the $l_1, \ldots, l_b$ remain bounded. In this case the leading term is of the form

$$e^{\mu_c V/a^2}(V/a^2)^{\beta},$$

where β is a critical exponent. If the boundary lengths $l_1, \ldots, l_b$ diverge according to (97), we expect a corresponding factor

$$e^{\lambda_c L_i/a}(L_i/a)^{\alpha},$$

where α is another critical exponent. This form of the entropy was encountered for branched polymers in (69). We can therefore express the expected asymptotic behaviour of the coefficients $w_{k,l_1,\ldots,l_b}$ as

$$w_{k,l_1,\ldots,l_b} \sim e^{\frac{\mu_c}{a^2}V}\, e^{\lambda_c \sum_i l_i}\, a^{-\alpha b - 2\beta}\, W(V, L_1, \ldots, L_b), \tag{98}$$

as $a \to 0$, with V and L_i defined by (97) fixed. The factor $a^{-\alpha b - 2\beta}$ may be thought of as a wave-function renormalization.

From (98) we deduce that the scaling behaviour of the discretized wave functional $w(\mu, \lambda_1, \ldots., \lambda_b)$ is given by

$$
\begin{aligned}
w(\mu, \lambda_1, \ldots, \lambda_b) &= \sum_{k,l_1,\ldots,l_b} e^{-\mu k}\, e^{-\sum_i \lambda_i l_i}\, w_{k,l_1,\ldots,l_b} \\
&\sim \frac{1}{a^{\alpha b+2\beta}} \sum_{k,l_1,\ldots,l_b} e^{-(\mu-\mu_0)k}\, e^{-\sum_i(\lambda_i-\lambda_0)l_i}\, W(V, L_1, \ldots, L_b) \\
&\sim \frac{1}{a^{(\alpha+1)b+(2\beta+2)}}\, W(\Lambda, Z_1, \ldots, Z_b),
\end{aligned} \tag{99}
$$

where we have used the relation

$$
W(\Lambda, Z_1, \ldots, Z_b) = \int_0^\infty dV \prod_{i=1}^{b} dL_i\; e^{-\Lambda V - \sum_i Z_i L_i}\; W(V, L_1, \ldots, L_b). \tag{100}
$$

Our next goal is to show that we can take a limit as suggested by (95) and (99). In terms of the variables g, z_i we have

$$
g = g_c(1 - \Lambda a^2), \quad z_i = z_c(1 + aZ_i), \tag{101}
$$

where we have introduced the notation

$$
z_c = c_+(g_c) = e^{\lambda_c} \tag{102}
$$

for the critical value $c_+(g_c)$ of z corresponding to the largest allowed value of g. Inserting (101) and (102) in the expression (87) one obtains

$$
w(g, z_1, z_2) \sim a^{-2} W(\Lambda, Z_1, Z_2), \tag{103}
$$

where

$$
W(\Lambda, Z_1, Z_2) = \frac{1}{2}\frac{1}{(Z_1 - Z_2)^2}\left(\frac{\frac{1}{2}(Z_1 + Z_2) + \sqrt{\Lambda}}{\sqrt{(Z_1 + \sqrt{\Lambda})(Z_2 + \sqrt{\Lambda})}} - 1\right). \tag{104}
$$

Similarly one can show, using the loop insertion operator, that when the number of boundaries is larger than two one has

$$
w(g, z_1, \ldots, z_b) \sim \frac{1}{a^{7b/2-5}}\left(-\frac{d}{d\Lambda}\right)^{b-3}\left[\frac{1}{\sqrt{\Lambda}}\prod_{i=1}^{b}\frac{1}{(Z_i + \sqrt{\Lambda})^{3/2}}\right], \tag{105}
$$

i.e.

$$
W(\Lambda, Z_1, \ldots, Z_b) \sim \left(-\frac{d}{d\Lambda}\right)^{b-3}\left[\frac{1}{\sqrt{\Lambda}}\prod_{i=1}^{b}\frac{1}{(Z_i + \sqrt{\Lambda})^{3/2}}\right].
$$

The continuum expressions for the n-loop functions are all universal and independent of the explicit form of the potential $V(z)$ as long as the weights $t_i \geq 0$. For the one-loop function the situation is different. As is seen by formally applying the counting of powers a in (105) to the case $b = 1$ one obtains the power $a^{3/2}$, i.e. a positive power of a. The important point in (105) and (103) is that the power is negative: in the scaling limit $a \to 0$ these terms will dominate. This is how the formulas should be understood: there are other terms too, but they will be subdominant when $a \to 0$, i.e. when $g \to g_c$ and $z \to z_c$ as dictated by (101) and (102). For the one-loop function the term associated with $a^{3/2}$ will vanish when $a \to 0$ and we will be left with a non-universal term explicitly dependent on the potential V. However, the term associated with $a^{3/2}$ is still the leading term which is non-analytic in the coupling constant g, so if we differentiate a number of times with respect to g and *then* take the limit $a \to 0$ it will be dominant. No continuum physics is associated with the analytic terms since they contain g only to some finite positive power, and are thus associated with only a finite number of triangles (of which the lattice length $a \to 0$ when $g \to g_c$) if we recall the interpretation of $w(g, z)$ as the generating function of the number of triangulations. Inserting (101) and (102) in the expression (80) one obtains

$$w(g, z) = \frac{1}{2}(V'(z) + a^{3/2} W(\Lambda, Z) + O(a^5/2)), \tag{106}$$

where

$$W(\Lambda, Z) = \left(Z - \frac{1}{2}\sqrt{\Lambda}\right)\sqrt{Z + \sqrt{\Lambda}}. \tag{107}$$

This ends the calculation of the loop – loop correlation functions for manifolds with topology S^2_b, the sphere with b boundaries. The results agree with continuum calculations using quantum Liouville theory and it follows that one can obtain a continuum, diffeomorphism-invariant theory starting out with a suitable lattice theory, where the lattice link length acts as a diffeomorphism-invariant UV cut-off and simply taking the lattice spacing $a \to 0$, while renormalizing the bare couplings in a standard way, namely, the cosmological and the boundary cosmological coupling constants.

Let us end this description of Euclidean quantum gravity by mentioning the corresponding results for higher-genus surfaces. The generalization of (105) is

$$w_h(g, z_1, \ldots, z_b) \sim \frac{1}{a^{7b/2+(5h-5)}} W_h(\Lambda, Z_1, \ldots, Z_b). \tag{108}$$

In particular, taking $b = 0$ leads to the expression

$$Z_h(g) \sim \frac{\tau_h}{(a^2\Lambda)^{5(h-1)/2}}, \tag{109}$$

for the singular part of the partition function, where the constants τ_h can in principle be (and have been) computed.

In (109) we have actually completed the task of calculating (47). We can now reintroduce the gravitational coupling constant G and try to calculate the complete partition function

$$Z(G,\Lambda) = \sum_{h=0}^{\infty} \tau_h \; e^{\frac{2-2h}{G}} \, a^{5(1-h)} \Lambda^{\frac{5(1-h)}{2}}. \tag{110}$$

The factor a^{-5} present for each genus can be absorbed in a renormalization of the gravitational coupling constant

$$\frac{1}{G_{\mathrm{ren}}} = \frac{1}{G(a)} - \frac{5}{4} \log \frac{\Lambda}{a^2}, \tag{111}$$

where G_{ren} denotes the *renormalized* gravitational coupling. A continuum limit of (110) only exists in the limit $a \to 0$ if we allow G to be a function of the lattice spacing a determined by (111) for fixed G_{ren} and Λ. The continuum limit is then given by

$$Z(G,\Lambda) = \sum_{h=0}^{\infty} \tau_h \; \left(\mathrm{e}^{-1/G} \Lambda^{-\frac{5}{4}}\right)^{2h-2}, \tag{112}$$

and depends only on the variable $x = \Lambda e^{4/(5G)}$. To actually calculate $Z(G,\Lambda)$ we have to perform the summation over the number of handles h in (112), an interesting task which we will not address here. Rather, we will focus on (111), since we can use this equation to calculate the β-function for G using

$$\beta(G) \equiv -a \frac{\mathrm{d}G(a)}{\mathrm{d}a}\bigg|_{\Lambda, G_{\mathrm{ren}}} = -\,52G^2. \tag{113}$$

Two-dimensional Euclidean quantum gravity is asymptotically free as already mentioned in the introduction.

2.5 Two-Dimensional Lorentzian Quantum Gravity

As already mentioned above, Euclidean quantum gravity does not really work in more than two dimensions (in the sense of leading to a continuum theory of higher-dimensional geometry). By contrast, the formalism called CDT, based on causal dynamical triangulations, seems to lead to very interesting results. It is based on the idea that there exists a globally defined (proper-)time variable, which can be used to describe the evolution of the universe. In addition, one assumes that the topology of space is unchanged with respect to the foliation defined by this global time.

These requirements are definitely not satisfied in two-dimensional *Euclidean* quantum gravity. In principle one can also "superimpose" a proper time on Euclidean quantum universes and follow their evolution as first described in the seminal work by Kawai and collaborators. Starting out with a spatial universe of topology S^1, it will immediately split up into many disconnected spatial, one-dimensional universes as a function of proper time. It turns out that the structure is fractal, in the sense that an infinity of spatial universes, most of them of infinitesimal spatial extension, will be created as a function of proper time.

Since two-dimensional Euclidean quantum gravity is explicitly solvable, even on a lattice before the continuum limit is taken, as described above, it is of interest to understand the transition from the Euclidean lattice gravity theory to the CDT lattice gravity theory. Clearly one has to suppress the splitting of a spatial universe into two or more disconnected spatial universes if one wants to move from the spacetime configurations which characterize the Euclidean path integral to the configurations present in the CDT path integral. It makes sense to talk about the splitting of a spatial universe into two if the universe has Lorentzian signature, since such a splitting (in the simplest case) is associated with an isolated point where the metric and its associated light-cone structure are degenerate, which has a diffeomorphism-invariant meaning. This was the motivation for imposing such a constraint in the original CDT model. By working in Lorentzian signature initially, and only later rotating to Euclidean signature, this constraint survives also in (the Euclidean version of) CDT. Going back to Fig. 2.16, this suggests that one should associate a factor g_s instead of a factor 1 with the graph with the double line. Geometrically this figure can be viewed as a process where a triangle is removed at a marked link (and a new link is marked at the new boundary), except in the case where the marked link does not belong to a triangle, but is part of a double link, in which case the double link is removed and the triangulation is separated into two. If one thinks of the recursion process in Fig. 2.16 as a "peeling away" of the triangulation as proper time advances, the presence of a double link represents the "acausal" splitting point beyond which the triangulation splits into two discs with two separate boundary components
(i.e. two separate one-dimensional spatial universes). The interpretation of this process, advocated in [48, 49], is that it represents a split of the spatial boundary with respect to (Euclidean) proper time. Associating an explicit weight g_s with this situation and letting $g_s \to 0$ suppress this process compared to processes where we simply remove an i-gon from the triangulation. Nevertheless, we will see below that there exists an interesting scaling of g_s with a such that the process survives when we let $a \to 0$, but with a result different from the Euclidean quantum gravity theory. We call this new limit *generalized* CDT [26–31].

Let us introduce the new coupling constant g_s in (78). The equation is then changed to

$$w(z) = g\left(\sum_{i=1}^{n} t_i z^{i-2}\right) w(z) + \frac{g_s}{z}\, w^2(z) + \frac{1}{z}\, Q(z, g). \tag{114}$$

In the analysis it will be convenient to keep the coupling constant $t_1 > 0$, although we are usually not so interested in situations with one-gons. It can be motivated as follows. Consider a "triangulation" consisting of T_1 one-gons, T_2 two-gons, T_3 triangles, T_4 squares, etc. up to T_n n-gons. The total coupling-constant factor associated with the triangulation is given by

$$g^{T_1+\cdots+T_n} g_s^{-T_1/2+T_3/2+\cdots+(n/2-1)T_n}. \tag{115}$$

We observe that in the limit $g_s \to 0$, a necessary condition for obtaining a finite critical value $g_c(g_s)$ is $T_1 > 0$. We should emphasize that the analysis described below can be carried out also if we suppress the appearance of any one-gons (by setting $t_1 = 0$) in our triangulations, but it is slightly more cumbersome since then $g_c(g_s) \to \infty$ as $g_s \to 0$, requiring further rescalings.

For simplicity we will consider the simplest nontrivial model with potential[6]

$$V(z) = \frac{1}{g_s}\left(-gz + \frac{1}{2}z^2 - \frac{g}{3}z^3\right) \tag{116}$$

and analyse its behaviour in the limit $g_s \to 0$. The disk amplitude (79) now has the form

$$w(z) = \frac{1}{2g_s}\left(-g + z - gz^2 + g(z-c_2)\sqrt{(z-c_+)(z-c_-)}\right), \tag{117}$$

and the constants c_2, c_+ and c_- are determined by the requirement that $w(z) \to 1/z$ for $z \to \infty$. Compared with the analysis of the previous section, the algebraic condition fixing the coefficient of $1/z$ to be unity will now enforce a completely different scaling behaviour as $g_s \to 0$.

For the time being, we will think of g_s as small and fixed, and perform the scaling analysis for $g_c(g_s)$. As already mentioned above, the critical point g_c is determined by the additional requirement $M_1 = 0$ in the representation (80), i.e. that $c_2(g_c) = c_+(g_c)$, which presently leads to the equation

$$\left(1 - 4g_c^2\right)^{3/2} = 12\sqrt{3}\, g_c^2\, g_s. \tag{118}$$

Anticipating that we will be interested in the limit $g_s \to 0$, we write the critical points as

$$g_c(g_s) = \frac{1}{2}(1 - \Delta g_c(g_s)), \quad \Delta g_c(g_s) = \frac{3}{2} g_s^{2/3} + O\left(g_s^{4/3}\right), \tag{119}$$

and

[6] The rationale for calling V a "potential" will become clear below.

$$z_c(g_s) = c_+(g_c, g_s) = \frac{1}{2g_c(g_s)}\left(1+\sqrt{\frac{1-4g_c(g_s)^2}{3}}\right) = 1 + g_s^{1/3} + O\left(g_s^{2/3}\right), \tag{120}$$

while the size of the cut in (116), $c_+(g_c) - c_-(g_c)$, behaves as

$$c_+(g_c) - c_-(g_c) = 4g_s^{1/3} + 0\left(g_s^{2/3}\right). \tag{121}$$

Thus the cut shrinks to zero as $g_s \to 0$.

Expanding around the critical point given by (119), (120) a nontrivial limit can be obtained if we insist that in the limit $a \to 0$, g_s scales according to

$$g_s = G_s a^3, \tag{122}$$

where a is the lattice cut-off introduced earlier. With this scaling the size of the cut scales to zero as $4\,a\,G_s^{1/3}$. In addition $\sqrt{(z-c_+)(z-c_-)} \propto a$ if we introduce the standard identification (101): $z = c_+(g_c) + a\,Z$. This scaling is different from the conventional scaling in Euclidean quantum gravity where $\sqrt{(z-c_+)(z-c_-)} \propto a^{1/2}$ since in that case $(z-c_+)$ scales while $(z-c_-)$ does not scale.

We can now write

$$g = g_c(g_s)(1-a^2\Lambda) = \bar{g}(1-a^2\Lambda_{\rm cdt} + O(a^4)), \tag{123}$$

with the identifications

$$\Lambda_{\rm cdt} \equiv \Lambda + \frac{3}{2}G_s^{2/3}, \quad \bar{g} = \frac{1}{2}, \tag{124}$$

as well as

$$z = z_c + aZ = \bar{z} + aZ_{\rm cdt} + O(a^2), \tag{125}$$

with the identifications

$$Z_{\rm cdt} \equiv Z + G_s^{1/3}, \quad \bar{z} = 1. \tag{126}$$

Using these definitions one computes in the limit $a \to 0$ that

$$w(z) = \frac{1}{a}\,\frac{\Lambda_{\rm cdt} - \frac{1}{2}Z_{\rm cdt}^2 + \frac{1}{2}(Z_{\rm cdt} - H)\sqrt{(Z_{\rm cdt}+H)^2 - \frac{4G_s}{H}}}{2G_s}. \tag{127}$$

In (127), the constant H (or rather, its rescaled version $h = H/\sqrt{2\Lambda_{\rm cdt}}$) satisfies the third-order equation

$$h^3 - h + \frac{2G_s}{(2\Lambda_{\rm cdt})^{3/2}} = 0, \tag{128}$$

which follows from the consistency equations for the constants c_2, c_+ and c_- in the limit $a \to 0$. We thus define

$$w(z) = \frac{1}{a} W_{\rm cdt}(Z_{\rm cdt}, \Lambda_{\rm cdt}, G_s) \equiv \frac{1}{a} W(Z, \Lambda, G_s) \tag{129}$$

in terms of the continuum Hartle – Hawking wave functions $W_{\rm cdt}(Z_{\rm cdt}, \Lambda_{\rm cdt}, G_s)$ and $W(Z, \Lambda, G_s)$.

Notice that while the cut of $\sqrt{(z-c_+)(z-c_-)}$ goes to zero as the lattice spacing a, it nevertheless survives in the scaling limit when expressed in terms of renormalized "continuum" variables, as is clear from (127). Only in the limit $G_s \to 0$ it disappears and we have

$$w(z) = \frac{1}{a} W_{\rm cdt}(Z_{\rm cdt}, \Lambda_{\rm cdt}, G_s) \underset{G_s \to 0}{\longrightarrow} \frac{1}{a} \frac{1}{Z_{\rm cdt} + \sqrt{2\Lambda_{\rm cdt}}}, \tag{130}$$

which is the original CDT disk amplitude introduced in [25].

Let us make some comments:

(1) We have dealt here directly with a "generalized" CDT model which in the limit $G_s \to 0$ reproduces the "original" CDT disk amplitude (130). The original two-dimensional CDT model was defined according to the principles already outlined in our discussion of four-dimensional quantum gravity. Thus the interpolation between two spatial slices separated by one lattice spacing of proper time is shown in Fig. 2.17. This figure is the two-dimensional analogue of Fig. 2.4. The combinatorial problem of summing over all such surfaces connecting two spatial boundaries separated by a certain number of time steps can be solved [25], and the continuum limit can be taken. The corresponding continuum "propagator" is found to be

$$\begin{aligned} G(X,Y;T) = {} & \frac{4\Lambda_{\rm cdt} \mathrm{e}^{-2\sqrt{\Lambda_{\rm cdt}}\,T}}{\left(\sqrt{\Lambda_{\rm cdt}} + X\right) + \mathrm{e}^{-2\sqrt{\Lambda_{\rm cdt}}\,T}\left(\sqrt{\Lambda_{\rm cdt}} - X\right)} \\ & \times \frac{1}{\left(\sqrt{\Lambda_{\rm cdt}} + X\right)\left(\sqrt{\Lambda_{\rm cdt}} + Y\right) - \mathrm{e}^{-2\sqrt{\Lambda_{\rm cdt}}\,T}\left(\sqrt{\Lambda_{\rm cdt}} - X\right)\left(\sqrt{\Lambda_{\rm cdt}} - Y\right)}, \end{aligned} \tag{131}$$

where X, Y are the boundary cosmological constants associated with the two boundaries of the cylinder and T is the proper time separating the two boundaries. The propagator has an asymmetry between X and Y because we have marked a point (a vertex in the discretized model). By an inverse Laplace transform one can calculate the propagator $G(X, L; T)$ as a function of the length of the unmarked boundary. In particular, we have

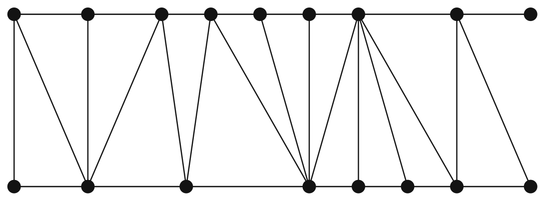


Fig. 2.17 The propagation of a spatial slice from time t to time $t+1$. The end of the strip should be joined to form a band with topology $S^1 \times [0, 1]$

$$G(X, L=0; T) = \int_{-i\infty}^{i\infty} \mathrm{d}Y \; G(X, Y; T), \tag{132}$$

and we define the CDT disk amplitude as

$$W^{(0)}(X) = \int_0^\infty \mathrm{d}T \; G(X, L=0; T), \tag{133}$$

and it is given by the expression on the far right in (130).

The generalized CDT model allows branching of the spatial universes as a function of proper time T, the branching being controlled by the coupling constant G_s. This results in the graphical equation for the generalized CDT disk amplitude shown in Fig. 2.18. The corresponding equation is

$$\begin{aligned} W(X) = W^{(0)}(X) + G_s \int_0^\infty dT \int_0^\infty dL_1 dL_2 \\ (L_1+L_2) G(X, L_1+L_2; T) W(L_1) W(L_2), \end{aligned} \tag{134}$$

where $W(L)$ is the disk amplitude corresponding to boundary length L. It can be solved [26–31] for $W(X)$ and the solution is $W_{\text{cdt}}(X, \Lambda_{\text{cdt}}, G_s)$ found above.

(2) Using (128) we can expand $w(z)$ into a power series in $G_s/(2\Lambda_{\text{cdt}})^{3/2}$ whose radius of convergence is $1/3\sqrt{3}$. For fixed values of Λ_{cdt}, this value corresponds to the largest value of G_s where (128) has a positive solution for h. The existence of

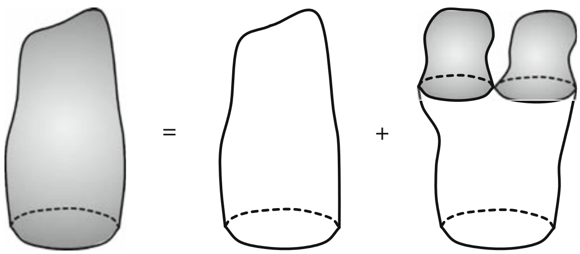


Fig. 2.18 Graphical illustration of (134). *Shaded parts* represent the generalized CDT disk amplitude, *unshaded parts* the original CDT disk amplitude and the original CDT propagator (131)

such a bound on G_s for fixed Λ_{cdt} was already observed in [26–31]. This bound can be re-expressed more transparently in the present Euclidean context, where it is more natural to keep the "Euclidean" cosmological constant Λ fixed, rather than Λ_{cdt}. We have

$$\frac{G_s}{(2\Lambda_{\text{cdt}})^{3/2}} \leq \frac{1}{3\sqrt{3}} \Rightarrow \frac{3G_s^{2/3}}{2\Lambda + 3G_s^{2/3}} \leq 1, \tag{135}$$

which for fixed $\Lambda > 0$ is obviously satisfied for all positive G_s. In order to see that the usual Euclidean two-dimensional quantum gravity (characterized by some *finite* value for g_s) can be re-derived from the disk amplitude (127), let us expand (127) for large G_s. The square root part becomes

$$a^{-1}\, G_s^{-5/6}\, \left(Z - \sqrt{2\Lambda/3}\right)\sqrt{Z + 2\sqrt{2\Lambda/3}}, \tag{136}$$

which coincides with the generic expression $a^{3/2}W(Z, \Lambda)$ in Euclidean two-dimensional quantum gravity (c.f. (106) and (107)) if we take G_s to infinity as g_s/a^3 (and take into account a trivial rescaling of the cosmological constant). However, if we reintroduce the same scaling in the $V'(z)$-part of $w(z)$, it does not scale with a but simply goes to a constant. This term would dominate $w(z)$ in the limit $a \to 0$ if one did not remove it by hand, as is usually done in the Euclidean model.

(3) Why does the potential $V'(z)$ (and therefore the entire disk amplitude $w(z)$) scale (like $1/a$) in the new continuum limit with $g_s = G_s a^3$, $a \to 0$, contrary to the situation in ordinary Euclidean quantum gravity? This is most clearly seen by looking again at the definitions (123) and (125). Because of the vanishing

$$V'(\bar{z}, \bar{g}) = 0, \quad V''(\bar{z}, \bar{g}) = 0 \tag{137}$$

in the point $(\bar{z}, \bar{g}) = (1, 1/2)$, expanding around $(\bar{z}, \bar{g})$ according to (123), (125) leads automatically to a potential which is of order a^2 when expressed in terms of the renormalized constants $(Z_{\text{cdt}}, \Lambda_{\text{cdt}})$, precisely like the square-root term when expressed in terms of $(Z_{\text{cdt}}, \Lambda_{\text{cdt}})$.

The point $(\bar{z}, \bar{g})$ differs from the critical point $(z_c(g_s), g_c(g_s))$, as long as $g_s \neq 0$. In fact, both $1/\bar{z}$ and $\bar{g}$ lie *beyond* the radii of convergence of $1/z$ and g, which are precisely $1/z_c(g_s)$ and $g_c(g_s)$. However, since the differences are of order a and a^2, respectively, they simply amount to *shifts* in the renormalized variables, as made explicit in (124) and (126). Therefore, re-expressing $W(Z, \Lambda, G_s)$ in (129) in terms of the variables Z_{cdt} and Λ_{cdt} simply leads to the expression $W_{\text{cdt}}(Z_{\text{cdt}}, \Lambda_{\text{cdt}}, G_s)$, first derived in [26–31]. Similarly, any geometric quantities defined with respect to Z and Λ can equally well be expressed in terms of Z_{cdt} and Λ_{cdt}. For instance, the average continuum length of the boundary and the average continuum area of a triangulation are given by

$$\langle L \rangle = \frac{\partial \ln W(Z, \Lambda, G_s)}{\partial Z} = \frac{\partial \ln W_{\mathrm{cdt}}(Z_{\mathrm{cdt}}, \Lambda_{\mathrm{cdt}}, G_s)}{\partial Z_{\mathrm{cdt}}}, \tag{138}$$

$$\langle A \rangle = \frac{\partial \ln W(Z, \Lambda, G_s)}{\partial \Lambda} = \frac{\partial \ln W_{\mathrm{cdt}}(Z_{\mathrm{cdt}}, \Lambda_{\mathrm{cdt}}, G_s)}{\partial \Lambda_{\mathrm{cdt}}}. \tag{139}$$

In the limit of $G_s \to 0$, the variables (Z, Λ) and $(Z_{\mathrm{cdt}}, \Lambda_{\mathrm{cdt}})$ become identical and the disk amplitude becomes the original CDT amplitude $(Z_{\mathrm{cdt}} + \sqrt{2\Lambda_{\mathrm{cdt}}})^{-1}$ alluded to in (130).

2.6 Matrix Model Representation

Above we have solved the two-dimensional gravity models by purely combinatorial techniques which emphasize the geometric interpretation: the quantum theory as a sum over geometries. The use of so-called matrix models allows one to perform the summation over the piecewise linear geometries in a relatively simple way. Surprisingly, it turns out that the *scaling limit* of the generalized CDT model has itself a matrix model representation. We will here describe how matrix models can be used instead of the combinatorial methods and how one is led to the CDT matrix model.

Let ϕ be a Hermitian $N \times N$ matrix with matrix elements $\phi_{\alpha\beta}$ and consider for $k = 0, 1, 2, \ldots$ the integral

$$\int d\phi \; e^{-\frac{1}{2}\mathrm{tr}\,\phi^2} \frac{1}{k!} \left(\frac{1}{3} \mathrm{tr}\,\phi^3 \right)^k, \tag{140}$$

where

$$d\phi = \prod_{\alpha \leq \beta} d\,\mathrm{Re}\,\phi_{\alpha\beta} \prod_{\alpha < \beta} d\,\mathrm{Im}\,\phi_{\alpha\beta}. \tag{141}$$

We can regard ϕ as a zero-dimensional matrix-valued field so the integral can be evaluated in the standard way by doing all possible Wick contractions of $(\mathrm{tr}\,\phi^3)^k$ and using

$$\left\langle \phi_{\alpha\beta}\phi_{\alpha'\beta'} \right\rangle = C \int d\phi \; e^{-\frac{1}{2}\sum_{\alpha\beta} |\phi_{\alpha\beta}|^2} \phi_{\alpha\beta}\phi_{\alpha'\beta'} = \delta_{\alpha\beta'}\delta_{\beta\alpha'}, \tag{142}$$

where C is a normalization factor. The evaluation of the expression (140) can be interpreted graphically by associating to each factor $\mathrm{tr}\,\phi^3$ an oriented triangle and to each term $\phi_{\alpha\beta}\phi_{\beta\gamma}\phi_{\gamma\alpha}$ contributing to the trace a labelling of its vertices by α, β, γ in cyclic order, such that the matrix element $\phi_{\alpha\beta}$ is associated with the oriented link whose endpoints are labelled by α and β in accordance with the orientation. Equation (142) can then be interpreted as a gluing of the link labelled by $(\alpha\beta)$ to an oppositely oriented copy of the same link, see Fig. 2.19.

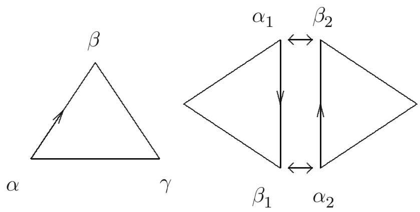


Fig. 2.19 The matrix representation of triangles which converts the gluing along links to a Wick contraction

In this way the integral (140) can be represented as a sum over closed, possibly disconnected, triangulations with k triangles. Triangulations with an arbitrary genus arise in this representation. By summing over k in (140) it follows by standard arguments that the (formal) logarithm of the corresponding integral is represented as a sum over all closed and *connected* triangulations. The contribution of a given triangulation can be determined by observing that in the process of gluing we pick up a factor of N whenever a vertex becomes an internal vertex in the triangulation. Thus, the weight of a triangulation T is simply

$$N^{N_v(T)}.$$

If we make the substitution

$$\mathrm{tr}\,\phi^3 \to \frac{g}{\sqrt{N}}\mathrm{tr}\,\phi^3, \tag{143}$$

the weight of T is replaced by

$$g^k N^{N_v(T)-k/2} = g^k N^{\chi(T)},$$

where $\chi(T)$ is the Euler characteristic of T. Note also that the factor $(k!)^{-1}$ in (140) is cancelled in the sum over different triangulations because of the $k!$ possible permutations of the triangles, except for triangulations with non-trivial automorphisms, in which case the symmetry factor C_T^{-1} survives. With the identifications

$$\frac{1}{G} = \log N, \qquad \mu = -\log g \tag{144}$$

we conclude that

$$Z(\mu, G) = \log \frac{Z(g, N)}{Z(0, N)}, \tag{145}$$

where $Z(\mu, \kappa)$ is defined by (59), (60) and

$$Z(g,N) = \int d\phi \ \exp\left(-\frac{1}{2}\mathrm{tr}\,\phi^2 + \frac{g}{3\sqrt{N}}\mathrm{tr}\,\phi^3\right). \tag{146}$$

The integral (146) is of course divergent and should just be regarded as a shorthand for the formal power series in the coupling constant g.

It is straightforward to generalize the preceding arguments to the case of general unrestricted triangulations, where arbitrary polygons are allowed. Instead of one coupling constant g we have a set $g_1, g_2, g_3, \ldots$, but we will still use the notation g or g_i. In this case (146) is replaced by

$$Z(g_i,N) = \int d\phi \ e^{-N\mathrm{tr}\,V(\phi)}, \tag{147}$$

where the potential V, which depends on all the coupling constants g_i, is given by

$$V(\phi) = \frac{1}{2}\phi^2 - \sum_{j=1}^{\infty} \frac{g_j}{j}\phi^j. \tag{148}$$

In (148) we have scaled $\phi \to \sqrt{N}\phi$ for later convenience. It is not difficult to check that the obvious generalization of (145) holds and the weight of a triangulation T is given by

$$C_T^{-1} N^{\chi(T)} \prod_{j\geq 1} g_j^{N_j(T)},$$

where $N_j(T)$ is the number of j-gons in T. Equation (147) is of course a representation of a formal power series which is obtained by expanding the exponential of the non-quadratic terms as a power series in the coupling constants and then performing the Gaussian integrations term by term.

Differentiating $\log Z(g_i,N)$ with respect to the coupling constants g_j, one obtains the expectation values of products of traces of powers of ϕ. These expectation values have a straightforward interpretation in terms of triangulations. Denoting the expectation with respect to the measure

$$Z(g_i,N)^{-1} e^{-N\mathrm{tr}\,V(\phi)}\, d\phi$$

by $\langle\cdot\rangle$ we see, for example, that $\left\langle N^{-1}\mathrm{tr}\,\phi^n\right\rangle$ is given by the sum over all connected triangulations of arbitrary genus whose boundary is an n-gon with one marked link. Similarly,

$$\frac{1}{N^2}\left\langle \mathrm{tr}\,\phi^n \mathrm{tr}\,\phi^m\right\rangle - \frac{1}{N^2}\left\langle \mathrm{tr}\,\phi^n\right\rangle\left\langle \mathrm{tr}\,\phi^m\right\rangle \tag{149}$$

is given by the sum over all connected triangulations whose boundary consists of two components with n and m links. More generally, the relation to the combinatorial problem is given by

$$w(g_i, z_1, \ldots, z_b) = N^{b-2} \sum_{k_1,\ldots,k_b} \frac{\langle \operatorname{tr} \phi^{k_1} \cdots \operatorname{tr} \phi^{k_b} \rangle_{conn}}{z_1^{k_1+1} \cdots z_b^{k_b+1}}, \tag{150}$$

where the subscript *conn* indicates the connected part of the expectation $\langle \cdot \rangle$. One can rewrite (150) as

$$w(g_i, z_1, \ldots, z_b) = N^{b-2} \left\langle \operatorname{tr} \frac{1}{z_1 - \phi} \cdots \operatorname{tr} \frac{1}{z_b - \phi} \right\rangle_{conn}. \tag{151}$$

The one-loop function $w(g_i, z)$ is related to the density $\rho(\lambda)$ of eigenvalues of ϕ defined by

$$\rho(\lambda) = \left\langle \sum_{i=1}^{N} \delta(\lambda - \lambda_i) \right\rangle, \tag{152}$$

where λ_i, $i = 1, \ldots, N$, denote the N eigenvalues of the matrix ϕ. With this definition we have

$$\frac{1}{N} \langle \operatorname{tr} \phi^n \rangle = \int_{-\infty}^{\infty} d\lambda\ \rho(\lambda) \lambda^n, \quad n \geq 0. \tag{153}$$

Hence,

$$w(g_i, z) = \int_{-\infty}^{\infty} d\lambda\ \frac{\rho(\lambda)}{z - \lambda}. \tag{154}$$

In the limit $N \to \infty$ the support of ρ is confined to a finite interval $[c_-, c_+]$ on the real axis. In this case $w(z)$ will be an analytic function in the complex plane, except for a cut along the interval $[c_-, c_+]$. Note that $\rho(\lambda)$ is determined from $w(z)$ by

$$2\pi i \rho(\lambda) = \lim_{\varepsilon \to 0} \left(w(\lambda - i\varepsilon) - w(\lambda + i\varepsilon) \right). \tag{155}$$

2.6.1 The Loop Equations

A standard method in quantum field theory is to derive identities by a change of variables in functional integrals. Here we apply this method to the matrix models and explore the invariance of the matrix integral (147) under infinitesimal field redefinitions of the form

$$\phi \to \phi + \varepsilon \phi^n, \tag{156}$$

where ε is an infinitesimal parameter. One can show that to first order in ε the measure $d\phi$ defined by (141) transforms as

$$d\phi \to d\phi \left(1 + \varepsilon \sum_{k=0}^{n} \operatorname{tr} \phi^k \operatorname{tr} \phi^{n-k}\right). \tag{157}$$

The action transforms according to

$$\operatorname{tr} V(\phi) \to \operatorname{tr} V(\phi) + \varepsilon \operatorname{tr} \phi^n V'(\phi) \tag{158}$$

to first order in ε. We can use these formulas to study the transformation of the measure under more general field redefinitions of the form

$$\phi \to \phi + \varepsilon \sum_{k=0}^{\infty} \frac{\phi^k}{z^{k+1}} = \phi + \varepsilon \frac{1}{z-\phi}. \tag{159}$$

This field redefinition only makes sense if z is on the real axis outside the support ρ. In the limit $N \to \infty$ this is possible for z outside the interval $[c_-, c_+]$. Under the field redefinitions (159) the transformations of the measure and the action are given by

$$d\phi \to d\phi \left(1 + \varepsilon \operatorname{tr} \frac{1}{z-\phi} \operatorname{tr} \frac{1}{z-\phi}\right), \tag{160}$$

$$\operatorname{tr} V(\phi) \to \operatorname{tr} V(\phi) + \varepsilon \operatorname{tr} \left(\frac{1}{z-\phi} V'(\phi)\right). \tag{161}$$

The integral (147) is of course invariant under this change of the integration variables. By use of (160) and (161) we obtain the identity

$$\int d\phi \left\{ \left(\operatorname{tr} \frac{1}{z-\phi}\right)^2 - N \operatorname{tr} \left(\frac{1}{z-\phi} V'(\phi)\right) \right\} e^{-N \operatorname{tr} V(\phi)} = 0. \tag{162}$$

The contribution to the integral coming from the first term in $\{\cdot\}$ in (162) is, by definition,

$$N^2 w^2(z) + w(z,z). \tag{163}$$

The contribution from the second term inside $\{\cdot\}$ in (162) can be written as an integral over the one-loop function as follows:

$$\frac{1}{N}\left\langle \operatorname{tr} \frac{V'(\phi)}{z-\phi} \right\rangle = \int d\lambda\, \rho(\lambda) \frac{V'(\lambda)}{z-\lambda} = \oint_C \frac{d\omega}{2\pi i} \frac{V'(\omega)}{z-\omega} w(\omega), \tag{164}$$

where the second equality follows from (155). The curve C encloses the support of ρ but not z. It is essential for the existence of C that ρ have compact support. We can then write (162) in the form

$$\oint_C \frac{d\omega}{2\pi i} \frac{V'(\omega)}{z-\omega}\, w(\omega) = w^2(z) + \frac{1}{N^2} w(z,z), \tag{165}$$

where z is outside the interval $[c_-, c_+]$ on the real axis. Since both sides of (165) can be analytically continued to $C \setminus [c_-, c_+]$ the equation holds in this domain.

We recognize (165) as the loop equation already derived by combinatorial means, except that we here have an additional term involving $w(z,z)$ and with a coefficient $1/N^2$. In fact, (165) is the starting point for a $1/N^2$-expansion, i.e. a higher-genus expansion. To leading order in $1/N^2$ we have (as already derived)

$$w(z) = w_0(z) = \frac{1}{2}\left(V'(z) - M(z)\sqrt{(z-c_+)(z-c_-)}\right) \tag{166}$$

and from (155) the corresponding eigenvalue density is

$$\rho(\lambda) = \frac{1}{2\pi} M(\lambda)\sqrt{(c_+ - \lambda)(\lambda - c_-)}. \tag{167}$$

Let us now discuss how the new scaling limit is described in the matrix formalism. Hermitian matrix models are often analysed in terms of the dynamics of their eigenvalues. Since the action in (147) is invariant under the transformation $\phi \to U\phi U^\dagger$, with $U \in U(N)$ a unitary $N \times N$-matrix, one can integrate out the "angular" degrees of freedom. What is left is an integration over the eigenvalues λ_i of ϕ only,

$$Z(g) \propto \int \prod_{i=1}^{N} d\lambda_i \; e^{-N\sum_j V(\lambda_j)} \prod_{k<l} |\lambda_k - \lambda_l|^2, \tag{168}$$

where the last factor, the Vandermonde determinant, comes from integrating over the angular variables and where

$$\mathrm{tr}\, V(\phi) = \sum_{i=1}^{N} V(\lambda_i). \tag{169}$$

Naively one might expect that the large-N limit is dominated by a saddle-point with $V'(\lambda) = 0$. However, this is not the case since the Vandermonde determinant in (168) contributes in the large-N limit. The cut which appears in $w(z)$ is a direct result of the presence of the Vandermonde determinant. In this way one can say that the dynamics of the eigenvalues is "non-classical", deviating from $V'(\lambda) = 0$, the size of the cut being a measure of this non-classicality. To get to the generalized CDT model we introduced a new coupling constant g_s in the matrix model by substituting

$$V(\phi) \to \frac{1}{g_s} V(\phi) \tag{170}$$

and considered the limit $g_s \to 0$. As we have seen the coupling constant g_s controls and reduces the size of the cut and thus brings the system closer to a "classical" behaviour. Thus the quantum fluctuations are reduced in the generalized CDT models.

Let us now consider the matrix potential (116), which formed the starting point of our new scaling analysis. We are still free to perform a change of variables. Inspired by relations (123),(124),(125),(126), let us transform to new "CDT" variables

$$\phi \to \bar{z}\,\hat{I} + a\Phi + O(a^2), \tag{171}$$

at the same time re-expressing g as

$$g = \bar{g}(1 - a^2\Lambda_{\mathrm{cdt}} + O(a^4)), \tag{172}$$

following (123). Substituting the variable change into the matrix potential, and discarding a ϕ-independent constant term, one obtains

$$V(\phi) = \bar{V}(\Phi) \equiv \frac{\Lambda_{\mathrm{cdt}}\Phi - \frac{1}{6}\Phi^3}{2G_s} \tag{173}$$

in the limit $a \to 0$, from which it follows that

$$Z(g, g_s) = a^{N^2} Z(\Lambda_{\mathrm{cdt}}, G_s), \quad Z(\Lambda_{\mathrm{cdt}}, G_s) = \int d\Phi\, \mathrm{e}^{-N \mathrm{tr}\, \bar{V}(\Phi)}. \tag{174}$$

The disk amplitude for the potential $\bar{V}(\Phi)$ is precisely $W(Z_{\mathrm{cdt}}, \Lambda_{\mathrm{cdt}}, G_s)$, and since by definition

$$\frac{1}{z - \phi} = \frac{1}{a} \frac{1}{Z_{\mathrm{cdt}} - \Phi}, \tag{175}$$

the first equal sign in (129) follows straightforwardly from the simple algebraic equation (173). We conclude that the continuum generalized CDT theory is described by the matrix model with potential $\bar{V}(\Phi)$.

The conclusion is that to leading order in N, the combinatorial method which works with a regularized lattice theory with a geometric interpretation and an explicit cut-off a has a matrix model representation, and even in the continuum limit where $a \to 0$ there exists a matrix model representation of the theory. Once this is proved, one can actually "derive" a number of the results known for the generalized CDT model from the matrix model by the formal manipulations of (173),(174), where a appears merely as a parameter without any obvious geometric interpretation as a cut-off. Also, once the matrix model equivalence is established, it is clear that one has automatically a higher-genus expansion available. This higher-genus expansion was first established working with the combinatorial, generalized CDT theory, and based on purely geometric arguments of splitting and joining of space as a function of proper time, as we (partly) described above. Such a theory could be turned into a kind of string field theory, imitating the work of Kawai and collaborators on Euclidean quantum gravity [53–59]. Like in the Euclidean case of Kawai et al., the formulation of the CDT string field theory used entirely a continuum notation (i.e. to cut-off a was already taken to zero). However, contrary to the situation for the Euclidean string field theory, we now have a matrix model representation even in the continuum, and many of the string field theory results derived in [26–31] follow easily from the loop equations (165) for the potential (173).

2.6.2 Summation over All Genera in the CDT Matrix Model

One remarkable application of the CDT matrix model representation is that we can find the Hartle – Hawking wave function summed over all genera.

In [26–31] the matrix model given by (173) and (174) was related to the CDT Dyson – Schwinger equations by (i) introducing into the latter an expansion parameter α, which kept track of the genus of the two-dimensional spacetime, and (ii) identifying this parameter with $1/N^2$, where N is the size of the matrix in the matrix integral. The $1/N$-expansion of our matrix model therefore plays a role similar to the $1/N$-expansion originally introduced by 't Hooft: it reorganizes an asymptotic expansion in a coupling constant t ($t = G_s/(\sqrt{\Lambda_{\text{cdt}}})^{3/2}$ in our case) into convergent sub-summations in which the kth summand appears with a coefficient N^{-2k}. In QCD applications, the physically relevant value is $N = 3$, to which the leading-order terms in the large N-expansion can under favourable circumstances give a reasonable approximation.

As we will see, for the purposes of solving our string field-theoretic model non-perturbatively, an additional expansion in inverse powers of N (and thus an identification of the contributions at each particular genus) is neither essential nor does it provide any new insights. This means that we will consider the entire sum over topologies "in one go", which simply amounts to setting $N = 1$, upon which the matrix integral (174) reduces to the ordinary integral[7]

[7] Starting from a matrix integral for $N \times N$-matrices like (174), performing a *formal* expansion in (matrix) powers commutes with setting $N = 1$, as follows from the following property of expectation values of products of traces, which holds for any $n = 1, 2, 3, \ldots$ and any set of non-negative integers $\{n_k\}$, $k = 1, \ldots, 2n$, such that $\sum_{k=1}^{2n} n_k = 2n$. For any particular choice of such numbers, consider

$$\left\langle \prod_{k=1}^{2n} \left(\frac{1}{N} \operatorname{tr} M^{n_k} \right) \right\rangle \equiv \frac{\int dM\, \mathrm{e}^{-\frac{1}{2}\operatorname{tr} M^2} \prod_{k=1}^{2n} \left(\operatorname{tr} M^{n_k}/N \right)}{\int dM\, \mathrm{e}^{-\frac{1}{2}\operatorname{tr} M^2}} = \sum_{m=-n}^{n} \omega_m N^m, \tag{176}$$

$$Z(G_s, \Lambda_{\rm cdt}) = \int dm \, \exp\left[-\frac{1}{2G_s} \left(\Lambda_{\rm cdt} m - \frac{1}{6}\, m^3 \right) \right], \tag{178}$$

while the disk amplitude can be written as

$$W_{\rm cdt}(X) = \frac{1}{Z(G_s, \Lambda_{\rm cdt})} \int dm \, \frac{\exp\left[-\frac{1}{2G_s} \left(\Lambda_{\rm cdt} m - \frac{1}{6}\, m^3 \right) \right]}{X - m}. \tag{179}$$

These integrals should be understood as formal power series in the dimensionless variable $t = G_s/(\sqrt{\Lambda_{\rm cdt}})^{3/2}$ appearing in (128). Any choice of an integration contour which makes the integral well defined and reproduces the formal power series is a potential non-perturbative definition. However, different contours might produce different non-perturbative contributions (i.e. which cannot be expanded in powers of t), and there may even be non-perturbative contributions which are not captured by any choice of integration contour. As usual in such situations, additional physics input is needed to fix these contributions.

To illustrate the point, let us start by evaluating the partition function given in (178). We have to decide on an integration path in the complex plane in order to define the integral. One possibility is to take a path along the negative axis and then along either the positive or the negative imaginary axis. The corresponding integrals are

$$Z(g, \lambda) = \sqrt{\Lambda_{\rm cdt}}\; t^{1/3} F_{\pm}(t^{-2/3}), \quad F_{\pm}(t^{-2/3}) = 2\pi \; e^{\pm i\pi/6} \mathrm{Ai}(t^{-2/3} \mathrm{e}^{\pm 2\pi i/3}), \tag{180}$$

where Ai denotes the Airy function. Both $F_{\pm}$ have the same asymptotic expansion in t, with positive coefficients. Had we chosen the integration path entirely along the imaginary axis we would have obtained ($2\pi i$ times) $\mathrm{Ai}(t^{-2/3})$, but this has an asymptotic expansion in t with coefficients of oscillating sign, which is at odds with its explicit power expansion in t. We have (using the standard notation of Airy functions)

---

where the last equation *defines* the numbers ω_m as coefficients in the power expansion in N of the expectation value. Now, we have that

$$\sum_{m=-n}^{n} \omega_m = (2n-1)!! \tag{177}$$

independent of the choice of partition $\{n_k\}$. The number $(2n-1)!!$ simply counts the "Wick contractions" of x^{2n} which we could have obtained directly as the expectation value $\langle x^{2n} \rangle$, evaluated with a one-dimensional Gaussian measure. In the model at hand, we will calculate sums of the form $\sum_{m=-n}^{n} \omega_m$ directly, since we are summing over all genera *without* introducing an additional coupling constant for the genus expansion. In other words, the dimensionless coupling constant t in this case already contains the information about the splitting and joining of the surfaces, and the coefficient of t^k contains contributions from two-dimensional geometries whose genus ranges between 0 and $[k/2]$. We cannot disentangle these contributions further unless we introduce N as an extra parameter.

$$F_{\pm}(z) = \pi \left(\mathrm{Bi}(z) \pm i\mathrm{Ai}(z)\right), \tag{181}$$

from which one deduces immediately that the functions $F_{\pm}(t^{-2/3})$ are not real. However, since $\mathrm{Bi}(t^{-2/3})$ grows like $e^{\frac{2}{3t}}$ for small t while $\mathrm{Ai}(t^{-2/3})$ falls off like $e^{-\frac{2}{3t}}$, their imaginary parts are exponentially small in $1/t$ compared to the real part, and therefore do not contribute to the asymptotic expansion in t. An obvious way to *define* a partition function which is real and shares the same asymptotic expansion is by symmetrization,

$$\frac{1}{2}(F_+ + F_-) \equiv \pi \mathrm{Bi}. \tag{182}$$

The situation parallels the one encountered in the double-scaling limit of the "old" matrix model but is less complicated.

Presently, let us collectively denote by $F(z)$ any of the functions $F_{\pm}(z)$ or $\pi\mathrm{Bi}(z)$, leading to the tentative identification

$$Z(G_s, \Lambda_{\mathrm{cdt}}) = \sqrt{\Lambda_{\mathrm{cdt}}}\, t^{1/3}\, F\left(t^{-2/3}\right), \quad F''(z) = zF(z), \tag{183}$$

where we have included the differential equation satisfied by the Airy functions for later reference. Assuming $X > 0$, we can write

$$\frac{1}{X-m} = \int_0^\infty dL\ \exp\left[-(X-m)\,L\right]. \tag{184}$$

We can use this identity in (179) to obtain the integral representation

$$W_{\mathrm{cdt}}(X) = \int_0^\infty dL\ \mathrm{e}^{-XL}\ \frac{F\left(t^{-2/3} - t^{1/3}\sqrt{\Lambda_{\mathrm{cdt}}}L\right)}{F\left(t^{-2/3}\right)}. \tag{185}$$

From the explicit expression of the Laplace transform we can now read off the Hartle – Hawking amplitude as function of the boundary length L:

$$W_{\mathrm{cdt}}(L) = \frac{F(t^{-2/3} - t^{1/3}\sqrt{\Lambda_{\mathrm{cdt}}}\,L)}{F(t^{-2/3})}. \tag{186}$$

Before turning to a discussion of the non-perturbative expression for $W_{\mathrm{cdt}}(L)$ we have just derived, let us remark that the asymptotic expansion in t of course agrees with that obtained by recursively solving the CDT Dyson – Schwinger equations. Using the standard asymptotic expansion of the Airy function one obtains

$$W_{\mathrm{cdt}}(L) = \mathrm{e}^{-\sqrt{\Lambda_{\mathrm{cdt}}}\,L}\ \mathrm{e}^{t\,h(t,\sqrt{\Lambda_{\mathrm{cdt}}}\,L)}\ \frac{\sum_{k=0}^{\infty} c_k\ t^k\ (1 - t\sqrt{\Lambda_{\mathrm{cdt}}}\,L)^{-\frac{3}{2}k-\frac{1}{4}}}{\sum_{k=0}^{\infty} c_k\ t^k}, \tag{187}$$

where the coefficients c_k are given by $c_0 = 1$, $c_k = \frac{1}{k!}\left(\frac{3}{4}\right)^k \left(\frac{1}{6}\right)_k \left(\frac{5}{6}\right)_k$, $k > 0$. In (187), we have rearranged the exponential factors to exhibit the exponential fall-off in the length variable L, multiplied by a term containing the function

$$h\left(t, \sqrt{\Lambda_{\text{cdt}}}\,L\right) = \frac{2}{3t^2}\left[\left(1 - t\sqrt{\Lambda_{\text{cdt}}}\,L\right)^{3/2} - 1 + \frac{3}{2}t\sqrt{\Lambda_{\text{cdt}}}\,L\right], \tag{188}$$

which has an expansion in positive powers of t.

$W_{\text{cdt}}(L)$ has the interpretation of the wave function of the spatial universe according to the hypothesis of Hartle and Hawking. $L \in [0, \infty]$ and the probability of finding a spatial universe with length between L and $L + \mathrm{d}L$ is

$$P(L) = \frac{|W_{\text{cdt}}(L)|^2}{L}, \tag{189}$$

since the integration measure is dL/L. Thus the probability is not normalizable in a conventional way and peaked at $L = 0$ since $W_{\text{cdt}}(L = 0) = 1$. However, for each term in the asymptotic expansion (187) we obtain a finite value $\langle L \rangle \sim 1/\sqrt{\Lambda_{\text{cdt}}}$ as one would naturally expect. Since termination of the series in (187) at a finite k also implies restricting the two-dimensional spacetime to have a finite genus we can say that as long as we restrict spacetime to have finite genus we have $\langle L \rangle \sim 1/\sqrt{\Lambda_{\text{cdt}}}$. However, if we allow spacetimes of arbitrarily large genus to appear, i.e. if topology fluctuations are unconstrained (that means at most suppressed by a coupling constant, but no upper limit on the genus imposed by hand), *a remarkable change appears*: $\langle L \rangle = \infty$ because the full non-perturbative $W_{\text{cdt}}(L)$ does not fall off like $\mathrm{e}^{-\sqrt{\Lambda_{\text{cdt}}}\,L}$ but only as $L^{-1/4}$ (Note that $W_{\text{cdt}}(L)$ is still integrable at infinity since the integration measure is dL/L). This dramatic change in large L behaviour ($W(L)$ also becomes oscillatory for large L, despite the fact that each term in the asymptotic expansion (187) is positive) is clearly to be attributed to surfaces of arbitrarily large genus, i.e. it is a genuinely non-perturbative result.

2.7 Discussion and Perspectives

The four-dimensional CDT model of quantum gravity is extremely simple. It is the path integral over the class of causal geometries with a global time foliation. In order to perform the summation explicitly, we introduce a grid of piecewise linear geometries, much in the same way as when defining the path integral in quantum mechanics. Next, we rotate each of these geometries to Euclidean signature and use as bare action the Einstein – Hilbert action[8] in Regge form. That is all.

[8] Of course, the full, effective action, including measure contributions, will contain all higher-derivative terms.

The resulting superposition exhibits a nontrivial scaling behaviour as function of the four-volume, and we observe the appearance of a well-defined average geometry, that of de Sitter space, the maximally symmetric solution to the classical Einstein equations in the presence of a positive cosmological constant. We are definitely in a quantum regime, since the fluctuations of the three-volume around de Sitter space are sizable, as can be seen in Fig. 2.7. Both the average geometry and the quantum fluctuations are well described in terms of the minisuperspace action (25). A key feature to appreciate is that, unlike in standard (quantum-)cosmological treatments, this description is the *outcome* of a non-perturbative evaluation of the *full* path integral, with everything but the scale factor (equivalently, $V_3(t)$) summed over. Measuring the correlations of the quantum fluctuations in the computer simulations for a particular choice of bare coupling constants enabled us to determine the continuum gravitational coupling constant G as $G \approx 0.42a^2$, thereby introducing an absolute physical length scale into the dimensionless lattice setting. Within measuring accuracy, our de Sitter universes (with volumes lying in the range of 6,000–47,000 ℓ_{Pl}^4) are seen to behave perfectly semi-classically with regard to their large-scale properties.

We have also indicated how we may be able to penetrate into the sub-Planckian regime by suitably changing the bare coupling constants. By "sub-Planckian regime" we mean that the lattice spacing a is (much) smaller than the Planck length. While we have not yet analysed this region in detail, we expect to eventually observe a breakdown of the semi-classical approximation. This will hopefully allow us to make contact with continuum attempts to define a theory of quantum gravity based on quantum field theory. One such attempt has been described in the introduction and is based on the concept of asymptotic safety. It uses renormalization group techniques in the continuum to study scaling violations in quantum gravity around an UV fixed point [2–7]. Other recent continuum field-theoretic models of quantum gravity which are not in disagreement with our data are the so-called Lifshitz gravity model [8, 9] and the so-called scale-invariant gravity model [10–12]. In principle it is only a question of computer power to decide if any of the models agree with our CDT model of quantum gravity.

On the basis of these results two major issues suggest themselves for further research. First, we need to establish the relation of our effective gravitational coupling constant G with a more conventional gravitational coupling constant, defined directly in terms of coupling matter to gravity. In the present work, we have defined G as the coupling constant in front of the effective action, but it would be desirable to verify directly that a gravitational coupling defined via the coupling to matter agrees with our G. In principle it is easy to couple matter to our model, but it is less straightforward to define in a simple way a set-up for extracting the semi-classical effect of gravity on the matter sector. Attempts in this direction were already undertaken in the "old" Euclidean approach [60, 61], and it is possible that similar ideas can be used in CDT quantum gravity.

The second issue concerns the precise nature of the "continuum limit". Recall our discussion in the Introduction about this in a conventional lattice-theoretic setting. The continuum limit is usually linked to a divergent correlation length at a critical point. It is unclear whether such a scenario is realized in our case. In general, it is rather unclear how one could define at all the concept of a divergent length related to correlators in quantum gravity, since one is integrating over all geometries, and it is the geometries which dynamically give rise to the notion of "length".

This has been studied in detail in two-dimensional (Euclidean) quantum gravity coupled to matter with central charge $c \leq 1$ [62–65]. It led to the conclusion that one could associate the critical behaviour of the matter fields (i.e. approaching the critical point of the Ising model) with a divergent correlation length, although the matter correlators themselves had to be defined as non-local objects due to the requirement of diffeomorphism invariance. On the other hand, the two-dimensional studies do not give us a clue of how to treat the gravitational sector itself, since they do not possess gravitational field-theoretic degrees of freedom. As we have seen the two-dimensional lattice models can be solved analytically and the only fine-tuning needed to approach the continuum limit is an additive renormalization of the cosmological constant. Thus, fixing the two-dimensional spacetime volume N_2 (the number of triangles), such that the cosmological constant plays no role, there are no further coupling constants to adjust and the continuum limit is automatically obtained by the assignment $V_2 = N_2 a^2$ and taking $N_2 \to \infty$. This situation can also occur in special circumstances in ordinary lattice field theory. A term like

$$\sum_i c_1(\phi_{i+1} - \phi_i)^2 + c_2(\phi_{i+1} + \phi_{i-1} - 2\phi_i)^2 \tag{190}$$

(or a higher-dimensional generalization) will also go to the continuum free field theory simply by increasing the lattice size and using the identification $V_d = L^d a^d$ (L denoting the linear size of the lattice in lattice units), the higher-derivative term being subdominant in the limit. It is not obvious that in quantum gravity one can obtain a continuum quantum field theory without fine-tuning in a similar way, because the action in this case is multiplied by a dimensionful coupling constant. Nevertheless, it is certainly remarkable that the infrared limit of our effective action apparently reproduces – within the cosmological setting – the Einstein – Hilbert action, which is the unique diffeomorphism-invariant generalization of the ordinary kinetic term, containing at most second derivatives of the metric. A major question is whether and how far our theory can be pushed towards an ultraviolet limit. We have indicated how to obtain such a limit by varying the bare coupling constants of the theory, but the investigation of the limit $a \to 0$ with fixed G has only just begun and other scenarios than a conventional UV fixed point might be possible. One scenario, which has often been discussed as a possibility, but which is still missing an explicit implementation is the following: when one approaches sub-Planckian scales the theory effectively becomes a topological quantum field theory where the metric plays no role. Also in our very explicit implementation of a quantum gravity model it is unclear how such a scenario would look.

Acknowledgements The material presented in this review is based on collaborations with numerous colleagues. In particular, we would like to thank Andrzej G´orlich, Willem Westra and Stefan

Zohren who have contributed in an essential way to the more recent part of the material presented. All authors acknowledge support by ENRAGE (European Network on Random Geometry), a Marie Curie Research Training Network, contract MRTN-CT-2004-005616, and JJ by COCOS (Correlations in Complex Systems), a Marie Curie Transfer of Knowledge Project, contract MTKD-CT-2004-517186, both in the European Community's Sixth Framework Programme. RL acknowledges support by the Netherlands Organisation for Scientific Research (NWO) under their VICI program. JJ acknowledges a partial support by the Polish Ministry of Science and Information Technologies grant 1P03B04029 (2005–2008).

References

1. S. Weinberg, "Ultraviolet divergences in quantum theories of gravitation", in: S.W. Hawking and W. Israel, (eds.), *General Relativity: Einstein Centenary Survey*, Cambridge University Press, Cambridge, UK (1979) pp. 790–831.
2. A. Codello, R. Percacci, C. Rahmede, *Investigating the ultraviolet properties of gravity with a Wilsonian renormalization group equation*. Annals Phys. **324** (2009) 414–469, [arXiv:0805.2909 [hep-th]].
3. M. Reuter, F. Saueressig, *Functional Renormalization Group Equations, Asymptotic Safety, and Quantum Einstein Gravity*, Lectures given at First Quantum Geometry and Quantum Gravity School, Zakopane, Poland, 23 March – 3 April 2007. 57 [arXiv:0708.1317 [hep-th]].
4. M. Niedermaier, M. Reuter, *The asymptotic safety scenario in quantum Gravity*, Living Rev. Rel. **9** (2006) 5; 173.
5. H.W. Hamber, R.M. Williams, *Nonlocal effective gravitational field equations and the running of Newton's G*. Phys. Rev. D **72** (2005) 044026, 16 [arXiv:hep-th/0507017].
6. D.F. Litim, *Fixed points of quantum gravity*, Phys. Rev. Lett. **92** (2004) 201301, 4 pages [arXiv:hep-th/0312114].
7. H. Kawai, Y. Kitazawa, M. Ninomiya, *Renormalizability of quantum gravity near two dimensions*. Nucl. Phys. B **467** (1996) 313–331 [arXiv:hep-th/9511217].
8. P. Hořava, *Spectral dimension of the universe in quantum gravity at a Lifshitz point*, 11 [arXiv:0902.3657 [hep-th]].
9. P. Hořava, *Quantum gravity at a Lifshitz point*. Phys. Rev. D **79** (2009) 084008, 15 [arXiv:0901.3775 [hep-th]].
10. M.E. Shaposhnikov, I.I. Tkachev, *Quantum scale invariance on the lattice*. 5 [arXiv:0811.1967 [hep-th]].
11. M. Shaposhnikov, D. Zenhausern, *Quantum scale invariance, cosmological constant and hierarchy problem*. Phys. Lett. B **671** (2009) 162–166 [arXiv:0809.3406 [hep-th]].
12. M. Shaposhnikov, D. Zenhausern, *Scale invariance, unimodular gravity and dark energy*. Phys. Lett. B **671** (2009) 187–192 [arXiv:0809.3395 [hep-th]].
13. J. Ambjørn, J. Jurkiewicz, R. Loll, *Dynamically triangulating Lorentzian quantum gravity*. Nucl. Phys. B **610** (2001) 347–382 [arXiv:hep-th/0105267].
14. J. Ambjørn, J. Jurkiewicz, R. Loll, *Reconstructing the universe*. Phys. Rev. D **72** (2005) 064014, 24 [arXiv:hep-th/0505154].
15. J. Ambjørn, J. Jurkiewicz, R. Loll, *Emergence of a 4D world from causal quantum gravity*. Phys. Rev. Lett. **93** (2004) 131301, 4 [arXiv:hep-th/0404156].
16. J. Ambjørn, J. Jurkiewicz, R. Loll, *Semiclassical universe from first principles*, Phys. Lett. B **607** (2005) 205–213 [arXiv:hep-th/0411152].
17. J. Ambjørn, A. Görlich, J. Jurkiewicz, R. Loll, *Planckian birth of the quantum de Sitter universe*. Phys. Rev. Lett. **100** (2008) 091304, 4 [arXiv:0712.2485 [hep-th]].
18. J. Ambjørn, A. Görlich, J. Jurkiewicz, R. Loll, *The nonperturbative quantum de Sitter universe*. Phys. Rev. D **78** (2008) 063544, 17 [arXiv:0807.4481 [hep-th]].
19. C. Teitelboim, *Causality versus gauge invariance in quantum gravity and supergravity*. Phys. Rev. Lett. **50** (1983) 705–708.

20. C. Teitelboim, *The proper time gauge in quantum theory of gravitation*. Phys. Rev. D **28** (1983) 297–309.
21. Link to animation of 2d Lorentzian quantum gravity: http://www.nbi.dk/ ambjorn/lqg2/.
22. J. Ambjørn, J. Jurkiewicz, R. Loll, *The universe from scratch*. Contemp. Phys. **47** (2006) 103–117 [arXiv:hep-th/0509010].
23. R. Loll, *The emergence of spacetime, or, quantum gravity on your desktop*. Class. Quant. Grav. **25** (2008) 114006, 17 pages [arXiv:0711.0273 [gr-qc]].
24. J. Ambjørn, A. Görlich, J. Jurkiewicz, R. Loll, *The quantum universe*. Acta Phys. Polon. B **39** (2008) 3309–3341.
25. J. Ambjørn, R. Loll, *Non-perturbative Lorentzian quantum gravity, causality and topology change*. Nucl. Phys. B **536** (1998) 407–434 [arXiv:hep-th/9805108].
26. J. Ambjørn, R. Loll, W. Westra, S. Zohren, *Putting a cap on causality violations in CDT*. JHEP **0712** (2007) 017, 16 [arXiv:0709.2784 [gr-qc]].
27. J. Ambjørn, R. Loll, Y. Watabiki, W. Westra, S. Zohren, *A string field theory based on causal dynamical triangulations*. JHEP **0805** (2008) 032, 26 pages [arXiv:0802.0719 [hep-th]].
28. J. Ambjørn, R. Loll, Y. Watabiki, W. Westra, S. Zohren, *A matrix model for 2D quantum gravity defined by causal dynamical triangulations*. Phys. Lett. B **665** (2008) 252–256 [arXiv:0804.0252 [hep-th]].
29. J. Ambjørn, R. Loll, Y. Watabiki, W. Westra, S. Zohren, *A new continuum limit of matrix models*. Phys. Lett. B **670** (2008) 224–230 [arXiv:0810.2408 [hep-th]].
30. J. Ambjørn, R. Loll, Y. Watabiki, W. Westra, S. Zohren, *A Causal Alternative for c=0 Strings*. Acta Phys. Polon. B **39** (2008) 3355–3364 [arXiv:0810.2503 [hep-th]].
31. J. Ambjørn, R. Loll, W. Westra, S. Zohren, *Summing over all topologies in CDT string field theory*. Phys. Lett. B **678** (2009) 227–232 [arXiv:0905.2108 [hep-th]].
32. J. Ambjørn, J. Jurkiewicz, R. Loll, G. Vernizzi, *Lorentzian 3d gravity with wormholes via matrix models*. JHEP **0109** (2001) 022, 34 [arXiv:hep-th/0106082].
33. J. Ambjørn, J. Jurkiewicz, R. Loll, G. Vernizzi, *3D Lorentzian quantum gravity from the asymmetric ABAB matrix model*. Acta Phys. Polon. B **34** (2003) 4667–4688 [arXiv:hep-th/0311072].
34. J. Ambjørn, J. Jurkiewicz, R. Loll, *Renormalization of 3d quantum gravity from matrix models*. Phys. Lett. B **581** (2004) 255–262 [arXiv:hep-th/0307263].
35. D. Benedetti, R. Loll, F. Zamponi, *(2+1)-dimensional quantum gravity as the continuum limit of causal dynamical triangulations*. Phys. Rev. D **76** (2007) 104022, 26 [arXiv:0704.3214 [hep-th]].
36. J. Ambjørn, B. Durhuus, J. Fröhlich, *Diseases of triangulated random surface models, and possible cures*. Nucl. Phys. B **257** (1985) 433–449;
37. J. Ambjørn, B. Durhuus, J. Fröhlich, P. Orland, *The appearance of critical dimensions in regulated string theories*. Nucl. Phys. B **270** (1986) 457–482.
38. A. Billoire, F. David, *Microcanonical simulations of randomly triangulated planar random surfaces*. Phys. Lett. B **168** (1986) 279–283.
39. D.V. Boulatov, V.A. Kazakov, I.K. Kostov, A.A. Migdal, *Analytical and numerical study of the model of dynamically triangulated random surfaces*. Nucl. Phys. B **275** (1986) 641–686.
40. J. Ambjørn, J. Jurkiewicz, *Four-dimensional simplicial quantum gravity*. Phys. Lett. B **278** (1992) 42–50.
41. J. Ambjørn, J. Jurkiewicz, *Scaling in four-dimensional quantum gravity*. Nucl. Phys. B **451** (1995) 643–676 [arXiv:hep-th/9503006].
42. M.E. Agishtein, A.A. Migdal, *Simulations of four-dimensional simplicial quantum gravity*. Mod. Phys. Lett. A **7** (1992) 1039–1062.
43. J. Ambjørn, J. Jurkiewicz, R. Loll, *Spectral dimension of the universe*. Phys. Rev. Lett. **95** (2005) 171301, 4 [arXiv:hep-th/0505113].
44. B. Dittrich, R. Loll, *Counting a black hole in Lorentzian product triangulations*. Class. Quant. Grav. **23** (2006) 3849–3878 [arXiv:gr-qc/0506035].
45. J. Ambjørn, S. Jain, G. Thorleifsson, *Baby universes in 2-d quantum gravity*. Phys. Lett. B **307** (1993) 34–39, [arXiv:hep-th/9303149].

46. J. Ambjørn, S. Jain, J. Jurkiewicz, C. F. Kristjansen, *Observing 4-d baby universes in quantum gravity*. Phys. Lett. B **305** (1993) 208–213 [arXiv:hep-th/9303041].
47. J. Ambjørn, B. Durhuus, *Regularized bosonic strings need extrinsic curvature*. Phys. Lett. B **188** (1987) 253–257.
48. H. Kawai, N. Kawamoto, T. Mogami, Y. Watabiki, *Transfer matrix formalism for two-dimensional quantum gravity and fractal structures of space-time*. Phys. Lett. B **306** (1993) 19–26 [arXiv:hep-th/9302133].
49. J. Ambjørn, Y. Watabiki, *Scaling in quantum gravity*. Nucl. Phys. B **445** (1995) 129–144 [arXiv:hep-th/9501049].
50. S. Catterall, G. Thorleifsson, M.J. Bowick, V. John, *Scaling and the fractal geometry of two-dimensional quantum gravity*. Phys. Lett. B **354** (1995) 58–68 [arXiv:hep-lat/9504009].
51. J. Ambjørn, J. Jurkiewicz, Y. Watabiki, *On the fractal structure of two-dimensional quantum gravity*. Nucl. Phys. B **454** (1995) 313–342 [arXiv:hep-lat/9507014].
52. H. Aoki, H. Kawai, J. Nishimura, A. Tsuchiya, *Operator product expansion in two-dimensional quantum gravity*. Nucl. Phys. B **474** (1996) 512–528 [arXiv:hep-th/9511117].
53. N. Ishibashi, H. Kawai, *String field theory of noncritical strings*. Phys. Lett. B **314** (1993) 190–196 [arXiv:hep-th/9307045].
54. N. Ishibashi, H. Kawai, *String field theory of $c \leq 1$ noncritical strings*. Phys. Lett. B **322** (1994) 67–78 [arXiv:hep-th/9312047].
55. N. Ishibashi, H. Kawai, *A Background independent formulation of noncritical string theory*. Phys. Lett. B **352** (1995) 75–82 [arXiv:hep-th/9503134].
56. Y. Watabiki, *Construction of noncritical string field theory by transfer matrix formalism in dynamical triangulation*. Nucl. Phys. B **441** (1995) 119–156 [arXiv:hep-th/9401096].
57. M. Ikehara, N. Ishibashi, H. Kawai, T. Mogami, R. Nakayama, N. Sasakura, *String field theory in the temporal gauge*. Phys. Rev. D **50** (1994) 7467–7478 [arXiv:hep-th/9406207].
58. M. Ikehara, N. Ishibashi, H. Kawai, T. Mogami, R. Nakayama, N. Sasakura, *A note on string field theory in the temporal gauge*. Prog. Theor. Phys. Suppl. **118** (1995) 241–258 [arXiv:hep-th/9409101].
59. J. Ambjørn, Y. Watabiki, *Non-critical string field theory for 2d quantum gravity coupled to (p,q) – conformal fields*. Int. J. Mod. Phys. A **12** (1997) 4257–4289 [arXiv:hep-th/9604067].
60. B.V. de Bakker, J. Smit, *Gravitational binding in 4D dynamical triangulation*. Nucl. Phys. B **484** (1997) 476–494 [arXiv:hep-lat/9604023].
61. H.W. Hamber, R.M. Williams, *Newtonian potential in quantum Regge gravity*. Nucl. Phys. B **435** (1995) 361–398 [arXiv:hep-th/9406163].
62. J. Ambjørn, K.N. Anagnostopoulos, U. Magnea, G. Thorleifsson, *Geometrical interpretation of the KPZ exponents*. Phys. Lett. B **388** (1996) 713–719 [arXiv:hep-lat/9606012].
63. J. Ambjørn, K.N. Anagnostopoulos, *Quantum geometry of 2D gravity coupled to unitary matter*. Nucl. Phys. B **497** (1997) 445–478 [arXiv:hep-lat/9701006].
64. J. Ambjørn, K.N. Anagnostopoulos, R. Loll, *A new perspective on matter coupling in 2d quantum gravity*. Phys. Rev. D **60** (1999) 104035, 11 [arXiv:hep-th/9904012].
65. J. Ambjørn, J. Jurkiewicz, Yu.M. Makeenko, *Multiloop correlators for two-dimensional quantum gravity*. Phys. Lett. B **251** (1990) 517–524.

Chapter 3
Lectures on Quantization of Gauge Systems

N. Reshetikhin

Abstract A gauge system is a classical field theory where among the fields there are connections in a principal G-bundle over the space – time manifold and the classical action is either invariant or transforms appropriately with respect to the action of the gauge group. The lectures are focused on the path integral quantization of such systems. Here two main examples of gauge systems are Yang–Mills and Chern–Simons.

3.1 Introduction

Gauge field theories are examples of classical field theories with the degenerate action functional. The degeneration is due to the action of the infinite-dimensional gauge group. Among most known examples are the Einstein gravity and Yang–Mills theory. The Faddeev–Popov (FP) method gives a recipe how to construct a quantization of a classical gauge field theory in terms of Feynman diagrams. Such quantizations are known as perturbative or semiclassical quantizations. The appearance of so-called ghost fermion fields is one of the important aspects of the FP method [20].

The ghost fermions appear in the FP approach as a certain technical tool. Their natural algebraic meaning is clarified in the BRST approach (the letters stand for Becchi, Rouet, Stora, and, independently, Tyutin who discovered this formalism). In the BRST setting, fields and ghost fermions are considered together as coordinates on a super-manifold. Functions on this super-manifold are interpretedas

N. Reshetikhin (✉)
Department of Mathematics, University of California, Berkeley, CA 94720-3840, U.S.A;
KdV Institute for Mathematics, University of Amsterdam, Science Park 904 1098 XH Amsterdam, The Netherlands;
Niels Bohr Visiting Professor, Department of Mathematics, University of Aarhus, Denmark,
reshetik@math.berkeley.edu

Reshetikhin, N.: *Lectures on Quantization of Gauge Systems*. Lect. Notes Phys. **807**, 125–190 (2010)
DOI 10.1007/978-3-642-11897-5_3

elements of the Chevalley complex of the Lie algebra of gauge transformations. In this setting the FP action is a specific cocycle and the fact that the integral with the FP action is equal to the original integral with the degenerate action is a version of the Lefschetz fixed-point formula.

Among all gauge systems the Yang–Mills theory is most interesting for physics because of its role in the standard model in high-energy physics [55]. At the moment there is a mathematically acceptable semiclassical (perturbative) definition of the Yang–Mills theory where the partition functions (amplitudes) are defined as formal power series of Feynman integrals. The ultraviolet divergencies in Feynman diagrams involving FP ghost fields can be removed by the renormalization [33], and the corresponding renormalization is asymptotically free [31]. All these properties make the Yang–Mills theory so important for the high-energy physics.

A mathematically acceptable definition of the path integral in the four-dimensional Yang–Mills theory which goes beyond the perturbation theory is still an open problem. One possible direction which may give such a definition is the constructive field theory, where the path integral is treated as a limit of finite-dimensional approximations.

Nevertheless, even mathematically loosely defined, the path integral remains a powerful tool for *phenomenological* mathematical and physics research in quantum field theory. It predicted many interesting conjectures, many of which were proven later by rigorous methods.

The main goal of these notes is a survey of the semiclassical quantization of the Yang–Mills and of the Chern–Simons theories. These lectures can be considered as a brief introduction to the framework of quantum field theory (along the lines outlined by Atiyah and Segal for topological and conformal field theories). The emphases are given to the semiclassical quantization of classical field theories.

In the Einstein gravity the metric on a space – time is a field. It is well known in dimension four that the semiclassical (perturbative) quantization of Einstein gravity fails to produce renormalizable quantum field theory. It is also known that three-dimensional quantum gravity is related to the Chern– Simons theory for non-compact Lie groups SL_2. In this lectures we will not go as far as to discuss this theory, but will focus on the quantum Chern–Simons field theory for compact Lie groups.

We start with a sketch of classical field theory, with some examples such as a non-linear sigma model, the Yang–Mills theory, and the Chern–Simons theory. Then we outline the framework of quantum field theory following the Atiyah and Segal description of basic structures in topological and conformal field theories. The emphasis is given to the semiclassical quantization. Then Feynman diagrams are introduced in the example of finite-dimensional oscillatory integrals. The Faddeev–Popov and BRST methods are first introduced in the finite-dimensional setting.

The last two sections contain the definition of the semiclassical quantization of the Yang–Mills and of the Chern–Simons theories. The partition functions in such theories are given by formal power series, where the coefficients are determined by Feynman diagrams.

3.2 Local Lagrangian Classical Field Theory

3.2.1 Space – Time Categories

Here we will focus on Lagrangian quantization of Lagrangian classical field theories.

In most general terms *objects* of a *d-dimensional space – time category* are $(d-1)$-dimensional manifolds (space manifolds). In specific examples of space – time categories space manifolds are equipped with a structure (orientation, symplectic structure, metric, etc.).

A *morphism* between two space manifolds Σ_1 and Σ_2 is a d-dimensional manifold M, possibly with a structure (orientation, symplectic, Riemannian, etc.), together with the identification of $\Sigma_1 \sqcup \overline{\Sigma_2}$ with the boundary of M. Here $\overline{\Sigma}$ is the manifold Σ with reversed orientation.

Composition of morphisms is the gluing along the common boundary. Here are examples of space – time categories.

The d-dimensional topological category. Objects are smooth, compact, oriented $(d-1)$-dimensional manifolds. A morphism between Σ_1 and Σ_2 is the homeomorphism class of d-dimensional compact-oriented manifolds with $\partial M = \Sigma_1 \sqcup \overline{\Sigma_2}$ with respect to homeomorphisms constant at the boundary. The orientation on M should agree with the orientations of Σ_i in a natural way.

The composition consists of gluing two morphisms along the common boundary and then taking the homeomorphism class of the result with respect to homeomorphisms constant at the remaining boundary.

The d-dimensional Riemannian category. Objects are $(d-1)$ Riemannian manifolds. Morphisms between two oriented $(d-1)$-dimensional Riemannian manifolds N_1 and N_2 are oriented d-dimensional Riemannian manifolds M, such that $\partial M = N_1 \sqcup \overline{N_2}$. The orientation on all three manifolds should naturally agree, and the metric on M agrees with the metric on N_1 and N_2 on a collar of the boundary. The composition is the gluing of such Riemannian cobordisms. For details see [49].

This category is important for many reasons. One of them is that it is the underlying structure for statistical quantum field theories [34].

The d-dimensional metrized cell complexes. Objects are $(d-1)$-dimensional oriented metrized cell complexes (edges have length, 2-cells have area, etc.). A morphism between two such complexes C_1 and C_2 is a metrized complex C together with two embeddings of metrized cell complexes $i : C_1 \hookrightarrow C$, $j : \overline{C_2} \hookrightarrow C$ where i is orientation reversing and j is orientation preserving. The composition is the gluing of such triples along the common $(d-1)$-dimensional subcomplex.

This category has a natural subcategory which consists of metrized cell approximations of Riemannian manifolds.

It is the underlying category for all lattice models in statistical mechanics.

The pseudo-Riemannian category. The difference between this category and the Riemannian category is that morphisms are pseudo-Riemannian with the signature $d-1, 1$. This is the most interesting category for physics. When $d = 4$ it represents the space – time structure of our universe.

3.2.2 Local Lagrangian Classical Field Theory

The basic ingredients of a d-dimensional local Lagrangian classical field theory are the following:

- For each space – time we assign the space of fields. Fields can be sections of a fiber bundle on a space – time, connections on a fiber bundle over a space – time, etc.
- The dynamics of the theory is determined by a local Lagrangian. It assigns to a field a volume form on M which depends locally on the field. Without giving a general definition we will give illustrating examples of local actions. Assume that fields are functions $\phi : M \to F$, and that F is a Riemannian manifold. An example of an (ultra)local Lagrangian for a field theory in a Riemannian category with such fields is

$$\mathcal{L}(\phi(x), d\phi(x)) = \left(\frac{1}{2}(d\phi(x), d\phi(x))_F - V(\phi(x))\right) dx, \tag{1}$$

  where $(.,.)_F$ is the metric on F, the scalar product on forms is induced by the metric on M, and dx denotes the Riemannian volume form on M.
  The action functional is the integral

$$S_M[\phi] = \int_M \mathcal{L}(\phi, d\phi).$$

  Solutions to the Euler–Lagrange equations for S_M form a (typically infinite-dimensional) manifold X_M.
- A boundary condition is a constraint on boundary values of fields which in "good cases" intersects with X_M over a discrete set. In other words, there is a discrete set of solutions to the Euler-Lagrange equations with given boundary conditions.

A d-dimensional classical field theory can be regarded as a functor from the space – time category to the category of sets. It assigns to a $(d-1)$-dimensional space the set of possible boundary values of fields, and to a space – time the set of possible solutions to the Euler-Lagrange equations with these boundary values.

Some examples of local classical field theories are outlined in the next sections.

3.2.3 Classical Mechanics

In classical mechanics the space – time is a Riemannian one-dimensional manifold with flat metric, that is, an interval. Fields in classical Lagrangian mechanics are smooth mappings of an interval of the real line to a smooth finite-dimensional manifold N, called the *configuration space* (parametrized paths).

The action in classical mechanics is determined by a choice of the Lagrangian function $\mathcal{L} : TN \to \mathbb{R}$ and is

$$S_{[t_2,t_1]}[\gamma] = \int_0^t \mathcal{L}(\dot{\gamma}(\tau), \gamma(\tau))d\tau,$$

where $\gamma = \{\gamma(t)\}_{t_1}^{t_2}$ is a parametrized path in N.

The Euler–Lagrange equations in terms of local coordinates $q = (q^1, \ldots, q^n) \in N$ and $\xi = (\xi^1, \ldots, \xi^n) \in T_q N$ are

$$-\sum_{i=1}^{n} \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \xi^i}(\dot{\gamma}(t), \gamma(t)) + \frac{\partial \mathcal{L}}{\partial q^i}(\dot{\gamma}(t), \gamma(t)) = 0,$$

where $\mathcal{L}(\xi, q)$ is the value of the Lagrangian at the point $(\xi, q) \in TN$.

The Euler–Lagrange equations are a non-degenerate system of second-order differential equations, if $\frac{\partial^2 \mathcal{L}}{\partial \xi^i \partial \xi^j}(\xi, q)$ is non-degenerate for all (ξ, q). In realistic systems it is assumed to be positive.

Even when the Euler–Lagrange equations are satisfied, the variation of the action is still not necessarily vanishing. It is given by boundary terms:

$$\delta S_{[t_2,t_1]}[\gamma] = \frac{\partial \mathcal{L}}{\partial \xi^i}(\dot{\gamma}(t), \gamma(t))\delta\gamma(t)^i \Big|_{t_1}^{t_2} . \tag{2}$$

Imposing Dirichlet boundary conditions means fixing boundary points of the path: $\gamma(t_1) = q_1 \in N$, and $\gamma(t_2) = q_2 \in N$. With these conditions the variation of γ at the boundary of the interval is zero and the boundary terms in the variation of the action vanish.

A concrete example of classical Lagrangian mechanics is the motion of a point particle on a Riemannian manifold in the potential force field. In this case

$$\mathcal{L}(\xi, q) = \frac{m}{2}(\xi, \xi) + V(q), \tag{3}$$

where (., .) is the metric on N and $V(q)$ is the potential.

3.2.4 *First-Order Classical Mechanics*

The non-degeneracy condition of $\frac{\partial^2 \mathcal{L}}{\partial \xi^i \partial \xi^j}(\xi, q)$ is violated in an important class of first-order Lagrangians.

Let α be a 1-form on N and b be a function on N. Define the action

$$S_{[t_2,t_1]}[\gamma] = \int_{t_1}^{t_2} (\langle \alpha(\gamma(t)), \dot{\gamma}(t)\rangle + b(\gamma(t)))dt,$$

where γ is a parametrized path.

The Euler–Lagrange equations for this action are

$$\omega(\dot{\gamma}(t)) + db(\gamma(t)) = 0,$$

where $\omega = d\alpha$. Naturally, the first-order Lagrangian system is called *non-degenerate*, if the form ω is non-degenerate. It is clear that a non-degenerate first-order Lagrangian system defines a symplectic structure on a manifold N. The Euler–Lagrange equations for such system are equations for flow lines of the Hamiltonian on the symplectic manifold (N, ω) generated by the Hamiltonian $H = -b$. It is also clear that the action of a non-degenerate first-order system is exactly the Hamilton–Jacobi action for this Hamiltonian system.

Assuming that γ satisfies the Euler–Lagrange equations the variation of the action does not yet vanish. It is given by the boundary terms (2):

$$\delta S_{[t_2,t_1]}[\gamma] = \langle \alpha(\gamma(t)), \delta\gamma(t) \rangle \Big|_{t_1}^{t_2}.$$

If $\gamma(t_1)$ and $\gamma(t_2)$ are constrained to Lagrangian submanifolds in $L_{1,2} \subset N$ with $TT_{1,2} \subset \ker(\alpha)$, these terms vanish.

Thus, constraining boundary points of γ to such a Lagrangian submanifold is a natural boundary condition for non-degenerate first-order Lagrangian systems. As we will see, this is a part of the more general concept where Lagrangian submanifolds define natural boundary conditions for Hamiltonian systems.

3.2.5 Scalar Fields

The space – time in such theory is a Riemannian category. Fields are smooth mappings from a space – time to $\mathbb{R}$ (sections of the trivial fiber bundle $M \times \mathbb{R}$). The action functional is

$$S_M[\phi] = \int_M \left(\frac{1}{2}(d\phi(x), d\phi(x)) - V(\phi(x)) \right) dx,$$

where the first term is determined by the metric on M and dx is the Riemannian volume form. The Euler – Lagrange equations are

$$\Delta\phi + V'(\phi) = 0. \tag{4}$$

The Dirichlet boundary conditions fix the value of the field at the boundary $\phi|_{\partial M} = \eta$ for some $\eta : \partial M \to \mathbb{R}$. The normal derivative of the field at the boundary varies for these boundary conditions.

3.2.6 Pure Euclidean d-Dimensional Yang–Mills

3.2.6.1 Fields, the Classical Action, and the Gauge Invariance

The space – time is a Riemannian d-dimensional manifold. Fields are connections on a principle G-bundle P over M, where G is a compact Lie group (see, for example, [24] for basic definitions). Usually it is a simple (or Abelian) Lie group.

The action functional is given by the integral

$$S_M[A] = \int_M \frac{1}{2} \operatorname{tr} \langle F(A), F(A) \rangle \, dx,$$

where $\langle ., . \rangle$ is the scalar product of two-forms on M induced by the metric, $\operatorname{tr}(AB)$ is the Killing form on the Lie algebra $\mathfrak{g} = Lie(G)$, $F(A)$ is the curvature of A, and dx is the volume form.

The Euler–Lagrange equations for the Yang–Mills action are

$$d_A^* F(A) = 0.$$

The Yang–Mills action is invariant with respect to gauge transformations. Recall that gauge transformations are bundle automorphisms (see, for example, [24]). Locally, a gauge transformation acts on a connection as

$$A \mapsto A^g = g^{-1} A g + g^{-1} dg.$$

Here we assume that G is a matrix group and $g^{-1}dg$ is the Maurer–Cartan form on G. Now let us describe the Dirichlet boundary conditions for the Yang–Mills theory. Fix a connection A^b on $P|_{\partial M}$. The Dirichlet boundary conditions on the connection A for the Yang–Mills theory require that A^b is the pullback of A to the boundary induced by the embedding $i : \partial M \to M$, i.e., $i^*(A) = A^b$. Gauge classes of Dirichlet boundary conditions define gauge classes of solutions to the Yang–Mills equations. See [26] for more details about classical Yang–Mills theory.

3.2.7 Yang–Mills Field Theory with Matter

Let V be a finite-dimensional representation of the Lie group G, and $V_P = P \times_G V$ be the vector bundle over M associated to a principal G-bundle P. Assume that V has an invariant scalar product $(., .)$.

The classical Yang–Mills theory with matter fields, which are sections of V_P, has the action functional

$$S[\Phi, A] = \int_M \left(\frac{1}{2} \operatorname{tr} \langle F(A), F(A) \rangle + \frac{1}{2} (\langle d_A \Phi, d_A \Phi \rangle) + U(\Phi) \right) dx,$$

where U is a G-invariant function on V and $\langle .,. \rangle$ is the scalar product on forms defined by the metric on M. The function U describes the self-interaction of the scalar field Φ.

The Euler–Lagrange equations in this theory are

$$*d_A F(A) + j_A = 0, \quad d_A^* d_A \Phi - U'(\Phi) = 0,$$

where $j_A \in \Omega^1(M, \mathfrak{g})$ is the one-form defined as $\mathrm{tr}\,\langle \omega, j_A \rangle = \langle \omega\Phi, d_A\Phi \rangle$.

Dirichlet boundary conditions in this theory are determined by the gauge class of the boundary values of the connection A and for the scalar field Φ.

3.2.8 Three-Dimensional Chern–Simons Theory

In this case the space – time category is the category of three-dimensional topological cobordisms. Fix a smooth three-dimensional manifold M. The space of fields of the Chern–Simons theory is the space of connections on a trivial principal G-bundle P over M (just as in the Yang–Mills theory). The choice of a simple compact Lie group G is part of the data.

The Chern–Simons form is the 3-form on P:

$$\alpha(A) = \mathrm{tr}\left(A \wedge dA - \frac{2}{3} A \wedge [A \wedge A] \right).$$

Because the bundle is trivial, $\alpha(A)$ defines a 3-form on M which we will also denote by $\alpha(A)$. The Chern–Simons action is

$$CS_M(A, p) = \int_M p^* \alpha(A).$$

This action is of the first order (in derivatives of A). It is very different from the Yang–Mills theory where the action is of the second order.

The variation of the Chern–Simons action is

$$\delta CS_M(A, p) = \int_M \mathrm{tr}\,(F(A) \wedge \delta A) + \int_{\partial M} \mathrm{tr}\,(A_\tau \wedge \delta A_\tau)\,,$$

where $A_\tau, \delta A_\tau$ are pullbacks to the boundary of A and δA.

The Euler–Lagrange equations for this Lagrangian are

$$F(A) = 0.$$

They guarantee that the first term (the bulk) in the variation vanishes. Solutions to the Euler–Lagrange equations are flat connections in P over M. On the space of solutions to the Euler–Lagrange equations we have

$$\delta CS_M(A, p) = (\Theta, \delta A_\tau),$$

where Θ is the one form on the space $C_{\partial M}$ of connections on $P|_{\partial M} \to \partial M$. Let D be the differential acting on forms on the space $C_{\partial M}$. The form $\omega = D\Theta$ is non-degenerate and defines a symplectic structure on $C_{\partial M}$:

$$\omega(\delta A, \delta B) = \int_{\partial M} \operatorname{tr} \delta A \wedge \delta B. \tag{5}$$

The Chern–Simons action is gauge invariant (for details see [24]). The action of the gauge group is Hamiltonian on $(C_{\partial M}, \omega)$. The result of the Hamiltonian reduction of this symplectic space with respect to the action of the gauge group is the finite-dimensional moduli space $F(\partial M)$ of gauge flat connections together with reduced symplectic structure.

Gauge orbits through flat connections from $C_{\partial M}$ which continue to flat connections on P over M form a Lagrangian submanifold $L_M \subset F(\partial M)$. The corresponding first-order Hamiltonian system describes the reduced Chern–Simons theory as a classical Hamiltonian field theory. For more details see, for example, [24, 7], and references therein.

3.3 Hamiltonian Local Classical Field Theory

3.3.1 The Framework

An n-dimensional Hamiltonian field theory in a category of space – time is an assignment of the following data to manifolds which are the objects and morphisms of this category:

- A symplectic manifold $S(M_{n-1})$ to an $(n-1)$-dimensional manifold M_{n-1}.
- A Lagrangian submanifold $L(M_n) \subset S(\partial M_n)$ to each n-dimensional manifold M_n.

These data shall satisfy the following axioms:

1. $S(\emptyset) = \{0\}$.
2. $S(M_1 \sqcup M_2) = S(M_1) \times S(M_2)$.
3. $L(M_1 \sqcup M_2) = L(M_1) \times L(M_2)$ with $L(M_i) \subset S(\partial M_i)$.
4. $(S(\overline{M}), \omega) = (S(M), -\omega)$.
5. An orientation preserving diffeomorphism $f : M_1 \to M_2$ of $(n-1)$-dimensional manifolds lifts to a symplectomorphism $s(f) : S(M_1) \to S(M_2)$.
6. Assume that $\partial M = (\partial M)_1 \sqcup (\partial M)_2 \sqcup (\partial M)'$ and that there is an orientation reversing diffeomorphism $f : (\partial M)_1 \to \overline{(\partial M)_2}$. Denote by M_f the result of gluing M along $(\partial M)_1 \simeq \overline{(\partial M)_2}$ via f:

$$M_f = M / \left\langle (\partial M)_1 \simeq \overline{(\partial M)_2} \right\rangle.$$

The Lagrangian submanifold corresponding to the result of the gluing should be

$$L(M_f) = \{x \in S((\partial M)') | \text{ such that there exists } y \in S(\partial M)_1 \\ \text{with}(y, s(f)(y), x) \in L(M)\}. \quad (6)$$

Notice that $\partial M_f = (\partial M)'$ by definition. This axiom is known as the gluing axiom. In classical mechanics the gluing axiom is the composition of the evolution at consecutive intervals of time.[1]

A *boundary condition* in the Hamiltonian formulation is a Lagrangian submanifold $L^b(\partial M)$ in the symplectic manifold $S(\partial M)$, assigned to the boundary ∂M of the manifold M, $L^b(\partial M) \subset S(\partial M)$. It factorizes into the product of Lagrangian submanifolds corresponding to connected components of the boundary:

$$L^b((\partial M)_1 \sqcup (\partial M)_2) = L^b((\partial M)_1) \times L^b((\partial M)_2).$$

Classical solutions with given boundary conditions are intersection points $L^b(\partial M) \cap L(M)$.

In order to glue classical solutions along the common boundary (composition of classical trajectories in classical mechanics) let us assume that boundary Lagrangian submanifolds are fibers of Lagrangian fiber bundles. That is, we assume that for each connected component $(\partial M)_i$ of the boundary a symplectic manifold $S((\partial M)_i)$ is given together with a Lagrangian fiber bundle $\pi_i : S((\partial M)_i) \to B((\partial M)_i)$ over some base space $B((\partial M)_i)$ with fibers defining the boundary conditions.

3.3.2 Hamiltonian Formulation of Local Lagrangian Field Theory

Here again, instead of giving general definitions we will give a few illustrating examples.

3.3.2.1 Classical Hamiltonian Mechanics

1. Let $H \in C^\infty(M)$ be the Hamiltonian function generating Hamiltonian dynamics on a symplectic manifold M.[2] Here is how such a system can be reformulated in the framework of a Hamiltonian field theory.

---

[1] I am grateful to V. Fock for many illuminating discussions of Hamiltonian aspects of field theory, see also [23].

[2] Recall that Hamiltonian mechanics is a dynamical system on a symplectic manifold (M, ω) with trajectories being flow lines of the Hamiltonian vector field v_H generated by a function $H \in C^\infty(M)$, $v_H = \omega^{-1}(dH)$. Here $\omega^{-1} : T^*M \to TM$ is the isomorphism induced by the symplectic structure on M.

Objects of the corresponding space – time category are points; morphisms are intervals $I = [t_1, t_2] \subset \mathbb{R}$ with the flat metric. The symplectic manifold assigned to the boundary of the space – time is

$$S(t_1, t_2) = \overline{M} \times M,$$

where M is the phase space of the Hamiltonian system and $\overline{M}$ is the phase space with the opposite sign of the symplectic form.

The Lagrangian subspace $L(I)$ in $S(t_1, t_2)$ is the set of pairs of points (x, y) where x is the initial point of a classical trajectory generated by H and y is the target point of this trajectory.

A pair of Lagrangian fiber bundles $\pi_1 : M \to B_1$ and $\pi_2 : M \to B_2$ with suitable base spaces B_1, B_2 defines a "complete" family of boundary conditions corresponding to the two components of the boundary of I.

Classical trajectories with such boundary conditions are intersection points in $\big(\pi^{-1}(b_1) \times \pi^{-1}(b_2)\big) \cap L(I)$, where $b_1 \in B_1, b_2 \in B_2$.

2. The Lagrangian mechanics on N (see Sect. 3.2.3) is equivalent (for non-degenerate Lagrangians) to the Hamiltonian mechanics on $M = T^*N$ with the canonical symplectic form. The Hamiltonian functions are given by the Legendre transform of the Lagrangian:

$$H(p, q) = \max_{\xi \in T_q N} (p(\xi) - L(\xi, q)).$$

The boundary conditions $q(t_1) = q_1, q(t_2) = q_2$ correspond to Lagrangian fiber bundles $T^*N \to N$ for each component of the boundary of the interval.

The Hamiltonian of a point particle on a Riemannian manifold is

$$H(p, q) = \frac{m}{2}(p, p) + V(q),$$

where (p, p) is uniquely determined by the metric on N.

3. A non-degenerate first-order Lagrangian defines a symplectic structure on the configuration space M given by $\omega = d\alpha$. Solutions to the Euler–Lagrange equations in such system are flow lines of the Hamiltonian vector field generated by the function $b(q)$, see Sect. 3.2.4. So, first-order non-degenerate Lagrangian systems are simply Hamiltonian systems on exact symplectic manifolds (i.e., on symplectic manifolds where the form ω is exact).

3.3.2.2 Bose Field Theory

In this case the symplectic manifold $S(N)$ assigned to a $(d-1)$-dimensional manifold N is an infinite-dimensional linear symplectic manifold which is the cotangent bundle to the space of real-valued smooth functions on N.

Since $C^\infty(N)$ is a linear space its tangent space at any point (can be thought as the space of infinitesimal variations of functions on N) can be naturally identified with the $C^\infty(N)$ itself.

$$\omega((\delta\eta_1, \delta f_1), (\delta\eta_2, \delta f_2)) = \int_N (\delta\eta_1\delta f_2 - \delta\eta_2\delta f_1)dx,$$

where $\eta_i \in T_{f_i}C^\infty(N)$ and $(\delta\eta_i, \delta f_i)$ are tangent vectors from $T_{(\eta_i, f_i)}(T^*C^\infty(N))$. In this formula we identified the tangent space to $T^*C^\infty(N)$ with $C^\infty(N) \oplus C^\infty(N)$.

The Lagrangian fibration corresponding to the Dirichlet boundary conditions is the standard projection $\pi : T^*C^\infty(N) \to C^\infty(N)$.

The Lagrangian submanifold $L(M) \subset S(\partial M)$ is the space of solutions to the Euler–Lagrange equations. Solutions to the Euler–Lagrange equations with given Dirichlet boundary condition $\phi|_{\partial M} = \eta$ are intersection points of $L(M)$ with the Lagrangian fiber $\pi^{-1}(\eta)$.

3.3.2.3 Yang–Mills Theory

Here we will discuss only the Yang–Mills theory where fields are connections in a trivial principal G-bundle. The symplectic manifold $S(N)$ assigned to the $(d-1)$-dimensional manifold N in such field theory is the cotangent bundle to the space of connections in a trivial principal G-bundle over N with the natural symplectic structure.

As in the case of a scalar Bose field the symplectic manifold is the cotangent bundle to the space of all possible Dirichlet boundary conditions. Since the space $\mathcal{A}(N)$ of all smooth connections on N is linear, its tangent bundle can naturally be identified with $\mathcal{A}(N) \oplus \mathcal{A}(N)$. The symplectic form is

$$\omega((\delta\eta_1, \delta A_1), (\delta\eta_2, \delta A_2)) = \int_N (\mathrm{tr}\,\langle\delta\eta_1, \delta A_2\rangle - \mathrm{tr}\,\langle\delta\eta_2, \delta A_1\rangle)dx.$$

Here tangent vectors $(\delta\eta_i, \delta A_i) \in T_{(\eta_i, A_i)}(T^*\mathcal{A}(N)$ are $\mathfrak{g}$-valued 1-forms on N. The scalar product $\langle ., . \rangle$ is the scalar product on 1-forms induced by the metric on N.

In the Hamiltonian formulation of the Yang–Mills theory, the symplectic manifold $S(\partial M)$ is the Hamiltonian reduction of $T^*\mathcal{A}(\partial M)$ with respect to the action of the gauge group. The Lagrangian submanifold $L(M) \subset S(\partial M)$ is the subspace of gauge orbits through boundary values of solutions to the Yang–Mills equation with their normal derivatives.

3.3.2.4 Chern–Simons

The main difference between the Yang–Mills theory and the Chern–Simons field theory is that the YM theory is a second-order theory while the CS is a first-order theory. Solutions to the Euler–Lagrange equations are flat connections on M, and their pullbacks to the boundary are flat connections on the boundary ∂M.

In the Hamiltonian formulation of the Chern–Simons theory, the symplectic manifold assigned to the boundary is the moduli space of flat connections on $P_{\partial M}$ and the Lagrangian submanifold $L(M)$ is the space of gauge classes of flat connections on $P_{\partial M}$ which continue to flat connections on P.

3.4 Quantum Field Theory Framework

3.4.1 General Framework of Quantum Field Theory

We will follow the framework of local quantum field theory which was outlined by Atiyah and Segal for topological and conformal field theories. In a nutshell it is a functor from a category of cobordisms to the category of vector spaces (or, more generally to some "known" category).

All known local quantum field theories can be formulated in this way at some very basic level. It does not mean that this is a final destination of our understanding of quantum dynamics at the microscopical scale. But at the moment this general setting includes the standard model, which agrees with most of the experimental data in high-energy physics. In this sense this is the accepted framework at the moment, just as at different points of history classical mechanics, classical electro-magnetism, and quantum mechanics were playing such a role.[3]

A *quantum field theory* in a given space – time category can be defined as a functor from this category to the category of vector spaces (or to another "standard," "known" category). It assigns a vector space to the boundary and a vector in this vector space to the manifold:

$$N \mapsto H(N), \quad M \mapsto Z_M \in H(\partial M).$$

The vector space assigned to the boundary is the space of pure states of the system on M. It may depend on the extra structure at the boundary (it can be a vector bundle over the moduli space of such structures). The vector $Z(M)$ is called the *partition function* or *the amplitude*.

These data should satisfy natural axioms, such as

$$H(\emptyset) = \mathbb{C}, \quad H(N_1 \sqcup N_2) = H(N_1) \otimes H(N_2), \text{ and} \tag{7}$$

$$Z_{M_1 \sqcup M_2} = Z_{M_1} \otimes Z_{M_2} \in H(\partial M_1) \otimes H(\partial M_2). \tag{8}$$

An isomorphism $f : N_1 \to N_2$ lifts to a linear isomorphism

[3] The string theory goes beyond such framework and beyond scales of present experiments. It is a necessary step further, and it already produced a number of outstanding mathematical ideas and results. One of the differences between the string theory and the quantum field theory is that the concept of non-perturbative string theory is still developing.

$$\sigma(f) : H(N_1) \to H(N_2).$$

The pairing

$$\langle ., . \rangle_N : H(\overline{N}) \otimes H(N) \to \mathbb{C}$$

is defined for each N. This pairing should agree with partition functions in the following sense. Let $\partial M = N \sqcup \overline{N} \sqcup N'$, then

$$(\langle ., . \rangle \otimes id) Z_M = Z_{M_N} \in H(N'), \tag{9}$$

where M_N is the result of gluing of M along N. The operation is known as the gluing axiom. We outlined its structure. The precise definition involves more details (see [5, 47]). The gluing axiom in particular implies the functoriality of Z:

$$Z_{M_1 \circ M_2} = Z_{M_1} * Z_{M_2} .$$

Originally this framework was formulated by Atiyah and Segal for topological and conformal field theories, but it is natural to extend it to more general and more realistic quantum field theories, including the standard model.

This framework is very natural in models of statistical mechanics on cell complexes with open boundary conditions, also known as lattice models.

The main physical concept behind this framework is the locality of the interaction. Indeed, we can cut our space – time manifold in small pieces and the resulting partition function Z_M in such framework is expected to be the composition of partition functions of small pieces. Thus, the theory is determined by its structure on "small" space – time manifolds or at "short distances." This is the concept of *locality*.

3.4.2 Constructions of Quantum Field Theory

3.4.2.1 Quantum Mechanics

Quantum mechanics fits into the framework of quantum field theory as a one-dimensional example. One-dimensional space – time category is the same as in classical Lagrangian mechanics.

In quantum mechanics of a point particle on a Riemannian manifold N the vector space assigned to a point is $L_2(N)$ with the usual scalar product. The quantized Hamiltonian is the second-order differential operator acting in $L_2(N)$

$$\hat{H} = -\frac{mh^2}{2} \Delta + V(q),$$

where Δ is the Laplace operator on N, $V(q)$ is the potential, and h is the Planck constant.

The operator

$$U_{t_2-t_1} = \exp\left(\frac{i}{h}\hat{H}(t_2-t_1)\right) \tag{10}$$

is known as the *propagator* or *evolution operator* in quantum mechanics. It is a unitary operator in $L_2(N)$ (assume N is compact and $V(q)$ is sufficiently good). It can be written as an integral operator:

$$U_{t_2-t_1}(f)(q) = \int_N U_{t_2-t_1}(q,q')f(q')dq', \tag{11}$$

where dq' is the volume measure on N induced by the metric.

The kernel $U_t(q,q')$ is a solution to the Schrödinger equation

$$\left(ih\frac{\partial}{\partial t} - \frac{h^2}{2m}\Delta + V(q)\right)U_t(q,q') = 0 \tag{12}$$

for $t > 0$ with the initial condition

$$\lim_{t\to+0} U_t(q,q') = \delta(q,q').$$

Quantum mechanics of a point particle on a Riemannian manifold N viewed as a one-dimensional quantum field theory assigns the vector space $L_2(N)$ to a point and the vector $Z(I)(q_1,q_2) = U_{t_2-t_1}(q_2,q_1) \in H(\partial I) = \overline{L_2(N)\otimes L_2(N)}$ to the interval $[t_2,t_1]$. Here $\overline{L_2(N)\otimes L_2(N)}$ is a certain completion of the tensor product which can be identified with a space of operators in $L_2(N)$, for details see any mathematically minded textbook on quantum mechanics, for example [50]. For a variety of reasons it is better to think about the space attached to a point not as $L_2(N)$ but as the space of 1/2-densities on N. Given two 1/2-densities a and b, their scalar product is

$$(a,b) = \int_N \bar{a}b,$$

where $\bar{a}b$ is now a density and can be integrated over N (for details see for example [10]). In terms of 1/2-densities the kernel of the evolution operator is a 1/2-density on $N \times N$ and

$$U_t(a)(q) = \int_N U_t(q,q')a(q'),$$

where $U_t(q,q')a(q')$ is a density in q' and can be integrated over N.

3.4.2.2 Statistical Mechanics

Lattice models in statistical mechanics also fit naturally in the framework of quantum field theory. The space – time category corresponding to these models is a combinatorial space category of cell complexes.

A simple combinatorial example of combinatorial quantum field theory with the dimer partition function can be found in [18].

The combinatorial construction of the TQFT (topological quantum field theory) based on representation theory of quantized universal enveloping algebras at roots of unity is given in [44] or, more generally, on any modular category.

Another combinatorial construction of TQFT based on triangulations is given in [52]. This TQFT is the double of the construction from [44], for details see, for example, [53].

3.4.2.3 Path Integral and the Semiclassical Quantization

If we were able to integrate over the space of fields in a Lagrangian classical field theory (as in lattice models in statistical mechanics) we could construct a quantization of a d-dimensional classical Lagrangian system as follows:

- To a $(d-1)$-dimensional manifold we assign the space of functionals on boundary values of fields. Here we assume that a choice of boundary conditions was made.
- To a d-dimensional manifold we assign the functional on boundary fields given by the integral

$$Z_M(b) = \int_{\phi|_{\partial M}=b} \exp\left(\frac{iS[\phi]}{h}\right) D\phi.$$

  If one treats the integral as a formal symbol which satisfies Fubini's theorem (the iterated integral is equal to the double integral), such assignment satisfies all properties of QFT. The problem is that the integral is usually not defined, unless the space of fields is finite or finite dimensional (as in statistical mechanics of cell complexes). Thus, one should either make sense of the integral and check whether the definition satisfies Fubini's theorem or define the QFT by some other means.

There are two approaches on how to make sense of *path integrals*. The approach of *constructive field theory* is based on approximating the path integral by a finite-dimensional integral and then proving that the finite QFT has a limit, when the mesh of the approximation goes to zero. For details of this approach see, for example, in [30].

Another approach is known as *perturbation theory*, or *semiclassical limit*. The main idea is to define the path integral in the way its asymptotic expansion as $h \to 0$ would look like, if the integral were defined. The coefficients of this asymptotic expansion are given by Feynman diagrams. Under the right assumptions the first few coefficients would approximate the desired quantity sufficiently well. The numbers derived from this approach are the base for the comparison of quantum field theoretical models of particles with the experiment.

In the next sections we will outline this approach on several examples.

When M is a cylinder $M = [t_1, t_2] \times N$, the partition function Z_M is an element of $H(N) \times H(N)^*$[4] and therefore can be regarded as an operator in $H(N)$. Classical observables become operators acting in $H(N)$. Thus, a quantization of classical field theories for space – time cylinders can be regarded as passing from classical commutative observables to quantum non-commutative observables. The partition function for the torus has a natural interpretation as a trace of the partition function for the cylinder (see, for example, [30] for more details).

3.5 Feynman Diagrams

3.5.1 Formal Asymptotic of Oscillatory Integrals

Let $\mathcal{M}$ be a compact smooth manifold with a volume form on it. In this section we will recall the diagrammatic formula for the asymptotic expansion of the integral

$$I_h(f) = \int_{\mathcal{M}} \exp\left(i\frac{f(x)}{h}\right) dx, \tag{13}$$

where f is a smooth function on $\mathcal{M}$ with finitely many isolated critical points.

Lemma 1 *We have the following identity:*

$$\lim_{\varepsilon\to 0} \int_{\mathbb{R}^N} \exp(i(x, Bx)/2 - \varepsilon(x, x)) x_{i_1} \cdots x_{i_n} d^N x$$
$$= (2\pi)^{\frac{N}{2}} i^{\frac{n}{2}} \frac{1}{\sqrt{|\det(B)|}} \exp\left(\frac{i\pi}{4}\operatorname{sign}(B)\right) \sum_m B^{-1}_{i_{m_1} i_{m_2}} B^{-1}_{i_{m_1} i_{m_2}} \dots B^{-1}_{i_{m_{n-1}} i_{m_n}} . \tag{14}$$

Here the sum is taken over perfect matchings m on the set $\{1, 2, \dots, n\}$, $\operatorname{sign}(B)$ denotes the signature of the real symmetric matrix B (the number of positive eigenvalues minus the number of negative eigenvalues).

Moreover, if n is odd, this integral is zero.

Proof First notice that

$$\lim_{\varepsilon\to 0} \int e^{\frac{i}{2}(x,Bx) - \varepsilon(x,x)} x_{i_1} \cdots x_{i_n} d^N x = \frac{\partial}{\partial y_{i_1}} \cdots \frac{\partial}{\partial y_{i_n}} \int e^{\frac{i}{2}(x,Bx) + (y,x)} d^N x \Big|_{y=0} .$$

[4] In this rather general discussion of the basic structures of a local quantum field theory we are deliberately somewhat vague about such details as the completion of the tensor product and similar topological questions. Such questions are better answered on a case-by-case basis.

After change of variables $z = x - iB^{-1}y$ in the Gaussian integral

$$\int_{\mathbb{R}^N} \exp\left(\frac{i}{2}(x, Bx)\right) d^N x = (2\pi)^{\frac{N}{2}} \frac{1}{\sqrt{|\det(B)|}} \exp\left(\frac{i\pi}{4}\,\mathrm{sign}(B)\right),$$

we have

$$\begin{aligned} \int_{\mathbb{R}^N} \exp\left(\frac{i}{2}(x, Bx) + (y, x)\right) d^N x \quad &= (2\pi)^{\frac{N}{2}} \frac{1}{\sqrt{|\det(B)|}} \exp\left(\frac{i\pi}{4}\,\mathrm{sign}(B)\right) \\ \times \exp\left(\frac{i}{2}(B^{-1}y, y)\right) . \end{aligned}$$

Expanding the right side in powers of y we obtain the contribution of monomials of degree $2k$:

$$\begin{aligned} \frac{i^k}{2^k!} \sum_{(i)(j)} (B^{-1})_{i_1 j_1} \dots (B^{-1})_{i_{2k} j_{2k}} y_{i_1} \dots y_{i_k} y_{j_1} \dots y_{j_k} = \frac{i^k}{2^k!} \\ \times \sum_{i_1 \le \dots \le i_{2k}} \frac{y_{i_1} \dots y_{i_{2k}}}{m_1(i)! \dots m_{2k}(i)!} \\ \times \sum_{\sigma \in S_{2k}} (B^{-1})_{\sigma(i_1)\sigma(i_2)} (B^{-1})_{\sigma(i_3)\sigma(i_4)} \dots (B^{-1})_{\sigma(i_{2k-1})\sigma(i_{2k})} . \end{aligned}$$

Here $m_1(i)$ is the number of the smallest entries in the sequence $i_1, \dots, i_{2k}$, $m_2(i)$ is the number of the smallest entries after the elimination of i_1, etc.

Taking derivatives with respect to y and taking into account that

$$\begin{aligned} \frac{1}{2^k!} \sum_{\sigma \in S_{2k}} (B^{-1})_{\sigma(i_1)\sigma(i_2)} (B^{-1})_{\sigma(i_3)\sigma(i_4)} \dots (B^{-1})_{\sigma(i_{2k-1})\sigma(i_{2k})} = \\ \times \sum_m (B^{-1})_{i_{m_1} i_{m_2}} (B^{-1})_{i_{m_1} i_{m_2}} \dots (B^{-1})_{i_{m_{2k-1}} i_{m_{2k}}} , \end{aligned}$$

where the sum is taken over perfect matchings on the set $\{1, 2, , \dots, 2k\}$, we obtain the desired formula.

For example when $n = 4$, then this integral is equal to

$$(B^{-1})_{12}(B^{-1})_{34} + (B^{-1})_{13}(B^{-1})_{24} + (B^{-1})_{14}(B^{-1})_{23}.$$

These three terms correspond to the perfect matching shown in Fig. 3.1.

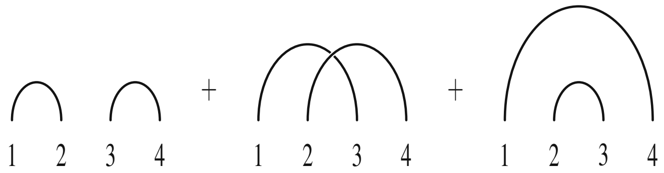


Fig. 3.1 Perfect matching for $n = 4$

Theorem 1 *We have the following identity of power series*

$$\int_{\mathbb{R}^N} \exp\left(i(x, Bx)/2 - \sum_{n\geq 3} \frac{i}{n!} V^{(n)}(x) h^{n/2-1} \right) d^N x =$$
$$(2\pi)^{\frac{N}{2}} \frac{1}{\sqrt{|\det(B)|}} \exp\left(\frac{i\pi}{4}\operatorname{sign}(B)\right) \sum_{\Gamma} \frac{(ih)^{-\chi(\Gamma)} F(\Gamma)}{|\operatorname{Aut}(\Gamma)|},$$

where sum is taken over graphs with vertices of valency ≥ 3*,* $F(\Gamma)$ *is the state sum corresponding to* Γ *described below,* $|\operatorname{Aut}(\Gamma)|$ *is the number of elements in the automorphism group of* Γ*,* $\chi(\Gamma)$ *is the Euler characteristic of the graph* $\chi(\Gamma) = |V| - |E|$*, where* $|E|$ *is the number of edges of* Γ*, and* $|V|$ *is the number of vertices of* Γ*.*

Proof Expanding the integral in formal power series in h we have

$$\int_{\mathbb{R}^N} e^{\frac{i}{2}(x,Bx)+\sum_{n\geq 3}\frac{i}{n!}V^{(n)}(x)h^{n/2-1}} dx = \sum_{n_3\geq 0, n_4\geq 0\cdots} \frac{h^{(3n_3+4n_4+\cdots)/2-n_3-n_4-\cdots} i^{n_3+n_4+\ldots}}{n_3!(3!)^{n_3} n_4!(4!)^{n_4}\cdots}$$
$$\int_{\mathbb{R}^N} e^{i(x,Bx)/2} (V^{(3)}(x))^{n_3} (V^{(4)}(x))^{n_4} \cdots d^N x. \tag{15}$$

Here

$$V^{(n)}(x) = \sum_{i_1,\ldots,i_n} V^{(n)}_{i_1,\ldots,i_n} x^{i_1} \ldots x^{i_n}.$$

For a graph Γ define the state sum $F(\Gamma)$ as follows:

- Enumerate vertices, for each vertex enumerate edges adjacent to it. This defines a total ordering on endpoints of edges (the ordering from left to right in Fig. 3.2).
- The graph Γ defines a perfect matching between edges adjacent to vertices as it is shown in Fig. 3.2. Denote by Γ_m the graph corresponding to the perfect matching m.
- Assign indices $i_1, i_2, \ldots$ to endpoints of edges, $i_\alpha = 1, 2, \ldots, N$.

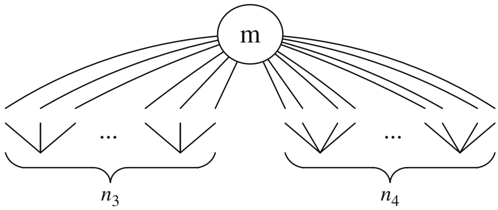


Fig. 3.2 Perfect matchings and Feynman diagrams

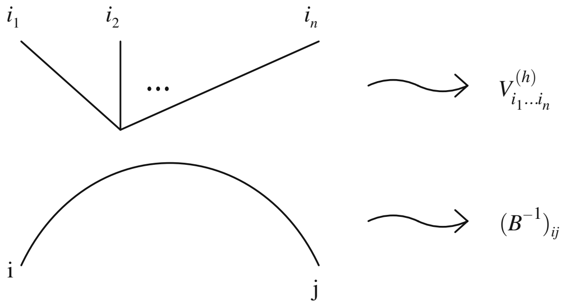


Fig. 3.3 Weights of vertices and edges in Feynman diagrams

- Define $F(\Gamma)$ as

$$F(\Gamma) = \sum_{\{i\}} \prod_{e \in E(\Gamma_m)} (B^{-1})_{e_l, e_r} V^{(n_1)}_{i_1, \ldots, i_{n_1}} V^{(n_1)}_{i_{n_1+1}, \ldots, i_{n_1+n_2}} V^{(n_1+n_2+1)}_{i_1, \ldots, i_{n_1+n_2+n_3}} \cdots$$

where e_l is the index corresponding to the left end of the edge e, e_r corresponds to the right side. The state sum $F(\Gamma)$ is the sum over $\{i\}$ of the product of weights assigned to vertices and edges according to the rules from Figs. 3.3 and 3.4.[5]

Lemma 1 gives the following expression for (15):

$$(2\pi)^{\frac{N}{2}} \frac{1}{\sqrt{|\det(B)|}} e^{\frac{i\pi}{4}\operatorname{sign}(B)}\, i^{|V|} \sum_{n_3 \geq 0, n_4 \geq 0 \cdots} \frac{(ih)^{|E|-|V|}}{n_3!(3!)^{n_3} n_4!(4!)^{n_4} \cdots} \sum_m F(\Gamma_m). \tag{16}$$

---

[5] Equivalently $F(\Gamma)$ can be defined as follows. Assign elements $1, \ldots, N$ to endpoints of edges of Γ. This defines an assignment of indices to endpoints of stars of vertices. The state sum is defined as

$$F(\Gamma) = \sum_{\{i\}} \prod_{e \in E(\Gamma)} (B^{-1})_{i_e, j_e} \prod_{v \in V(\Gamma)} (\text{weight of } v)_i \,.$$

Here weights of vertices are defined as in Fig. 3.3, the indices i_e, j_e correspond to two different endpoints of e (since B is symmetric, it does not matter that this pair is defined up to a permutation).

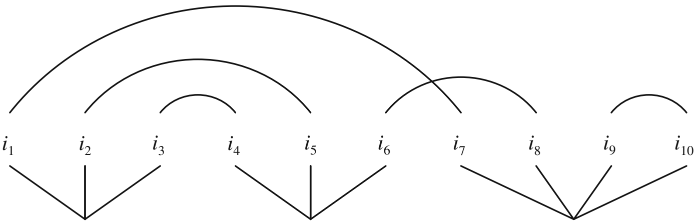


Fig. 3.4 An example of a perfect matching with a state $\{i\}$

Here the sum is taken over perfect matchings, and Γ_m is the graph corresponding to the matching m, see Fig. 3.2, $|E|$ is the number of edges and $|V|$ is the number of vertices of the graph Γ_m.

Some perfect matchings produce the same graphs. Denote by $N(\Gamma)$ the number of perfect matchings corresponding to Γ. In the formula (16) the contribution from the diagram Γ will have the combinatorial factor

$$\frac{N(\Gamma)}{n_3!(3!)^{n_3}n_4!(4!)^{n_4}\cdots} = \frac{1}{|\operatorname{Aut}(\Gamma)|}.$$

This finishes the proof.

There is a simple rule how to check powers of $i = \sqrt{-1}$. These factors disappear, if we replace $B \mapsto iB$ and $V^{(n)} \mapsto iV^{(n)}$.

A Feynman diagram *has order n* if it appears as a coefficient in h^n, i.e., when $n = |E| - |V|$ (or $n = -\chi(\Gamma)$) in the expansion above. As an example, order one Feynman diagrams are given in Figs. 3.5 and 3.6.

Now let us focus on the asymptotic expansion of the integral (13). The standard asymptotic analysis applied to this integral shows that the leading contributions to the asymptotics of the integral as $h \to 0$ come from the infinitesimal (of the diameter of order $h^{-1/2}$) neighborhoods of critical points of $f(x)$. The contribution

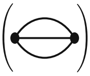
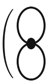
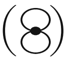
$\frac{1}{2^3}F\Big($  $\Big) + \frac{1}{3!2}F\Big($  $\Big) + \frac{1}{2^2}F\Big($  $\Big)$

Fig. 3.5 Contributions from Feynman diagrams of order one

$F\Big($  $\Big) = \sum_{\{i\}\{k\}} V^{(2)}_{i_1j_1k_1} V^{(2)}_{i_2j_2k_2} (B^{-1})_{i_1j_1} (B^{-1})_{i_2j_2} (B^{-1})_{k_1k_2}$

$F\Big($  $\Big) = \sum_{i_1i_2i_3i_4} V^{(4)}_{i_1i_2i_3i_4} V^{(2)}_{i_2j_2k_2} (B^{-1})_{i_1i_1} (B^{-1})_{i_3i_4}$

$F\Big($  $\Big) = \sum_{\{i\}\{k\}} V^{(2)}_{i_1i_2k_1} V^{(2)}_{i_2j_2k_2} (B^{-1})_{i_1i_2} (B^{-1})_{j_1j_2} (B^{-1})_{k_1k_2}$

Fig. 3.6 Weights of Feynman diagrams of order one

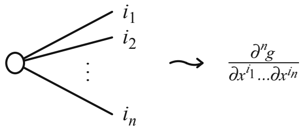


Fig. 3.7 Extra vertices in Feynman diagrams for the integral (18)

to the integral (13) from the critical point a "localizes" to the integral (15) with $(B_a)_{ij} = \frac{\partial^2 f}{\partial x^i \partial x^j}(a)$ and $(V_a^{(n)})_{i_1,\ldots,i_n} = -\frac{\partial^n f}{\partial x^{i_1} \ldots x^{i_n}}(a)$.

Choose local coordinates such that $dx = dx_1 \ldots dx_N$. Denote by $F_a(\Gamma)$ the state sum on the graph Γ with such matrices B and $V(n)$. The asymptotic expansion of the integral (13) has the following form:

$$\int_{\mathcal{M}} \exp\left(i\frac{f(x)}{h}\right)_{+1} dx \simeq \sum_a (2\pi h)^{\frac{N}{2}} \frac{1}{\sqrt{|\det(B_a)|}} e^{\frac{if(a)}{h} + \frac{i\pi}{4}\,\mathrm{sign}(B_a)} \times \sum_{\Gamma} \frac{(ih)^{-\chi(\Gamma)} F_a(\Gamma)}{|\operatorname{Aut}(\Gamma)|} \tag{17}$$

Here $\simeq$ is the asymptotical equivalence when $h \to 0$. A similar argument applied to the integral

$$\int_{\mathcal{M}} \exp\left(i\frac{f(x)}{h}\right) g(x)dx \tag{18}$$

gives the asymptotic expansion as $h \to 0$. It looks exactly as (17) with the only difference that in each Feynman diagram there will be exactly one of the vertices given in Fig. 3.7. The order of the diagram is still $|E| - |V|$, where V is the number of vertices given by derivatives of f, i.e., $-\chi(\Gamma) + 1$.

3.5.2 *Integrals Over Grassmann Algebras*

The *Grassmann algebra* G_n is the exterior algebra of $\mathbb{C}^n$, $G_n = \wedge^{\cdot}\mathbb{C}^n$ with the multiplication $(a, b) \to a \wedge b$. As an algebra defined in terms of generators and relations G_n is generated by $c^1, \ldots, c^n$ with defining relations $c^i c^j + c^j c^i = 0$. The Grassmann algebra G_n can also be regarded as the space of polynomial functions on the super-vector space $\mathbb{C}^{0|n}$.

Left derivatives with respect to c^i are defined as

$$\partial_{c^i} c^{i_1} \cdots c^{i_n} = \begin{cases} 0 & i \notin \{i_1, \ldots, i_n\} \\ (-1)^k c^{i_1} \cdots \hat{c}^{i_k} \cdots c^{i_n} & i = i_k\,. \end{cases}$$

The right derivatives are defined similarly with the sign $(-1)^{n-k}$ instead.

Recall that an *orientation* of $\mathbb{C}^n$ is defined by a basis in $\bigwedge^n \mathbb{C}^n$. Choose $c^1 \wedge \cdots \wedge c^n$ as such orientation. Any element $P \in G_n$ can be written as $p^{\text{top}} c^1 \wedge \cdots \wedge c^n +$ lower terms. The integral of P over the super-vector space $\mathbb{C}^{0|n}$ with the orientation $c_1 \wedge \cdots \wedge c_n$ is

$$\int_{\mathbb{C}^{0|n}} P\, dc := p^{\text{top}}.$$

Lemma 2 *Let* $(c, Bc) = \sum_{ij=1}^n c^i B_{ij} c^j$*, where* B *is skew-symmetric* $B_{ij} = -B_{ji}$*. If* n *is even, then*

$$\int_{\mathbb{C}^{0|n}} \exp\Big(\frac{1}{2}(c, Bc)\Big)\, dc = \mathrm{Pf}(B), \tag{19}$$

where Pf *is the Pfaffian of the matrix* B*. If* n *is odd, the integral is zero.*

Proof Recall that

$$\mathrm{Pf}(B) = \sum_m (-1)^m B_{i_1 j_1} B_{i_2 j_2} \ldots B_{i_{n/2} j_{n/2}}$$

where the sum is taken over perfect matchings m. A perfect matching m is the equivalence class of a collection of pairs $((i_1, j_1), \ldots, (i_{n/2}, j_{n/2}))$ obtained by a permutation σ of $(1, 2, \ldots, n)$ with respect to permutations of pairs $((i_a, j_a)$ with $(i_b, j_b))$ and permutations in a pair $((i_a, j_a)$ to $(j_a, i_a))$. The sign $(-1)^m$ is the sign of the permutation σ, which is constant on the equivalence class m.

Now let us prove the formula (19). It is clear that only monomials of degree n in c will give a non-zero contribution to the integral and that they all come from the term

$$(c, Bc)^{n/2} = \sum_{i_1,\ldots,i_{n/2}, j_1,\ldots,j_{n/2}=1}^{n} B_{i_1 j_1} \cdots B_{i_{n/2} j_{n/2}} c^{i^1} c^{j_1} \cdots c^{i_{n/2}} c^{j_{n/2}}.$$

Reordering factors we get

$$c^{i_1} c^{j_1} \cdots c^{i_{n/2}} c^{j_{n/2}} = (-1)^{\sigma(i|j)} c^1 \cdots c^n\,,$$

where $\sigma(i|j)$ is the permutation which brings $i_1, j_1, \ldots, i_{n/2}, j_{n/2}$ to $1, 2, \ldots, n$. Thus for the Gaussian Grassmann integral we have

$$\int_{\mathbb{C}^{0|n}} \exp\Big(\frac{1}{2}(c, Bc)\Big)\, dc = \frac{(1/2)^{n/2}}{(n/2)!} \sum_{\sigma(i|j)} B_{i_1 j_1} \cdots B_{i_{n/2} j_{n/2}} (-1)^{\sigma(i|j)}\,.$$

Note that the sign does not change when i_a is switched with j_a because the signs come in pairs. Also, the sign does not change when pair (i_a, j_a) and (i_b, j_b) are permuted. But such equivalence classes of permutations are exactly perfect matchings and therefore the formula becomes

$$\sum_{\substack{\sigma(i|j)\\ i_a<j_a\\ i_{a_1}<\cdots<i_{a_n}}} (-1)^{\sigma(i|j)} B_{i_1 j_1} \cdots B_{i_{n/2} j_{n/2}} = \mathrm{Pf}(B),$$

which is the Pfaffian of B.

This lemma is equivalent to the identity

$$\Big(\sum_{i<j} x^i \wedge x^j B_{ij}\Big)^{\wedge \frac{n}{2}} = \mathrm{Pf}(B) x^1 \wedge \cdots \wedge x^n$$

in the exterior algebra $\bigwedge^{\cdot} \mathbb{C}^n$.

Two important identities for Pfaffians:

$$\det B = \mathrm{Pf}(B)^2, \qquad \mathrm{Pf}\big(\begin{smallmatrix} 0 & A \\ -A^t & 0 \end{smallmatrix}\big) = \det A.$$

The following formula is a Grassmann analog of the formula from Lemma 1 for integrating monomials with respect to the Gaussian measure:

$$\int_{\mathbb{C}^{0|n}} \exp\Big(\frac{1}{2}(c, Bc)\Big) c^{i_1} \ldots c^{i_k}\, dc = \mathrm{Pf}(B)(-1)^{\frac{k}{2}} \sum_m (-1)^m (B^{-1})^{i_{m_1} i_{m_2}} \ldots \times (B^{-1})^{i_{m_{k-1}}, i_{m_k}}. \tag{20}$$

Here the sum is taken over perfect matchings m of $1, \ldots, k$, and B is assumed to be non-degenerate. The proof of this formula is parallel to the one for Gaussian oscillating integrals. The only difference is the factor $(-1)^m$ which appears when left derivatives are applied to the exponent.

Let $P(c)$ be an even element of G_n with monomials of degree at least 4, $P(c) = \sum_{k\geq 4} \frac{1}{k!} P^{(k)}(c)$ where $P^{(k)}(c) = \sum_{i_1,\ldots,i_k=1}^{n} P^{(k)}_{i_1,\ldots,i_k} c^{i_1} \ldots c^{i_k}$.

Theorem 2 *The following identity holds:*

$$\int_{\mathbb{C}^{0|n}} \exp\Big(-\frac{1}{2}(c, Bc) + P(c)\Big)\, dc = \mathrm{Pf}(-B) \sum_{\Gamma} (-1)^{c(D(\Gamma))} \frac{F(D(\Gamma))}{|\,\mathrm{Aut}(\Gamma)|}, \tag{21}$$

where the summation is taken over finite graphs, $D(\Gamma)$ is a mapping of Γ to $\mathbb{R}^2$, with the only singular points being crossings of edges ($D(\Gamma)$ is a diagram of the graph Γ), and $c(D(\Gamma))$ is the number of crossings of edges in the diagram $D(\Gamma)$. The number $F(D(\Gamma))$ is computed by the same rules as in the previous section. The product $(-1)^{c(D(\Gamma))} F(D(\Gamma))$ does not depend on the choice of the diagram.

Proof Expand the integral in $P(c)$:

$$\int_{\mathbb{C}^{0|n}} \exp\Big(\frac{1}{2}(c, Bc) + P(c)\Big)\, dc = \sum_{n_4, n_6, \dots \geq 0} \frac{1}{n_4!(4!)^{n_4} n_6!(6!)^{n_6} \dots}$$
$$\sum_{i_1, i_2, i_3, \dots} P^{(4)}_{i_1, i_2, i_3, i_4} \dots P^{(6)}_{i_{4n_4+1}, i_{4n_4+2}, i_{4n_4+3}, i_{4n_4+4}, i_{4n_4+5}, i_{4n_4+6}} \dots$$
$$\int_{\mathbb{C}^{0|n}} \exp\Big(\frac{1}{2}(c, Bc)\Big) c^{i_1} c^{i_2} c^{i_3} \dots dc. \tag{22}$$

Using the identity (20) we arrive at the formula

$$\mathrm{Pf}(B) \sum_{n_4, n_6, \dots \geq 0} \frac{1}{n_4!(4!)^{n_4} n_6!(6!)^{n_6} \dots}$$
$$\sum_{i_1, i_2, i_3, \dots} P^{(4)}_{i_1, i_2, i_3, i_4} \dots P^{(6)}_{i_{4n_4+1}, i_{4n_4+2}, i_{4n_4+3}, i_{4n_4+4}, i_{4n_4+5}, i_{4n_4+6}} \dots$$
$$\sum_m (-1)^m (B^{-1})_{i_{m_1} i_{m_2}} (B^{-1})_{i_{m_3}, i_{m_4}} \dots, \tag{23}$$

where m is a perfect matching on $1, 2, \dots, k$, $k = \sum_{i \geq 3} i n_i$. The summation over $\{i\}$ gives the number $F(D_m)$, where D_m is the diagram from Fig. 3.2. Some of the diagrams D_m represent projections of the same graph. It is easy to show that the combination $(-1)^m F(D_m)$ depends only on the graph, but not on its diagram and is equal to $(-1)^{c(D(\Gamma))} F(D(\Gamma))$ for any diagram $D(\Gamma)$ of Γ. Thus, if we will change the summation from n_i and m to the summation over graphs, the factorials together with the number of perfect matchings corresponding to the same graph produce the combinatorial factor $1/|\operatorname{Aut}(\Gamma)|$.

Having in mind applications to oscillatory integrals, it is convenient to have the formula (21) in the form

$$\int_{\mathbb{C}^{0|n}} \exp\Big(\frac{i}{2h}(c, Bc) - \frac{i}{h} P(c)\Big)\, dc = h^{-\frac{n}{2}} \mathrm{Pf}(iB) \sum_{\Gamma} (ih)^{-\chi(\Gamma)} (-1)^{c(D(\Gamma))} \frac{F(D(\Gamma))}{|\operatorname{Aut}(\Gamma)|}. \tag{24}$$

3.5.3 Formal Asymptotics of Oscillatory Integrals Over Super-manifolds

There are a number of equivalent definitions of super-manifolds. For our goals a super-manifold $M_{(n|m)}$ is a trivial vector bundle over a smooth n-dimensional manifold M (even part) with the fiber which is the exterior algebra of an m-dimensional vector space V (odd part). The algebra of functions on such a super-manifold is the algebra of sections of this vector bundle with the point-wise exterior multiplication on fibers, i.e., if $f, g : M \to M \times \wedge V$ are two sections $x \mapsto (x, f(x))$ and $x \mapsto (x, g(x))$, their product is the section

$$x \mapsto (x, f(x) \wedge g(x)).$$

In other words, it is the tensor product of the Grassmann algebra of the fibers with the algebra of smooth functions on M, i.e.,

$$C^\infty(M_{(n|m)}) = C^\infty(M) \otimes_{\mathbb{R}} \left\langle c^1, \dots, c^m | c^\alpha c^\beta = -c^\beta c^\alpha \right\rangle.$$

Elements of the algebra are polynomials in anti-commuting variables $c^1, \cdots, c^m$ with coefficients in smooth functions on M:

$$f(x, c) = f_0(x) + \sum_{k=1}^{m} \sum_{\alpha_1 < \cdots < \alpha_k} f_{\alpha_1, \dots, \alpha_k}(x) c^{\alpha_1} \dots c^{\alpha_k}. \tag{25}$$

Let dx be a volume form for the manifold M. Choose the orientation $c_1 \dots c_m$ on the fibers. By definition, the integral of the function $f(x, c)$ with respect to the volume form dx and the orientation $c_1 \dots c_m$ is

$$\int_{M_{(n|m)}} f dxdc = \int_M f_{1,\dots,m}(x) dx.$$

An even function on such a super-manifold has only terms of even degree in (25). *Critical points of an even function* f on $M_{(n|m)}$ are, by definition, critical points of f_0 on M.

Let f be an even function on $M_{(n|m)}$. Consider the following integral

$$\int_{M_{(n|m)}} \exp\left(\frac{if(x, c)}{h}\right) g(x, c) dxdc. \tag{26}$$

Here we assume that M is compact and that all functions are smooth.

Combining asymptotic analysis and the asymptotic expansion for oscillating integrals with the formulae for Grassmann integrals obtained in the previous section we arrive at the following asymptotic expansion for the integral (26):

$$\int_{M_{(n|m)}} \exp\left(\frac{if(x,c)}{h}\right) g(x,c)dxdc \simeq h^{\frac{n-m}{2}} (2\pi)^{\frac{n}{2}}$$
$$\sum_a \frac{1}{\sqrt{|\det(B(a))|}} \mathrm{Pf}(iL(a)) \exp\left(\frac{i}{h} f(a) + \frac{i\pi}{4} \operatorname{sign}(B(a))\right)$$
$$\left(1 + \sum_{\Gamma \neq \emptyset} \frac{(ih)^{-\chi(\Gamma)} (-1)^{c(D(\Gamma))} F_a(D(\Gamma))}{|\operatorname{Aut}(\Gamma)|}\right). \tag{27}$$

Here $B(a)_{ij} = \frac{\partial^2 f}{\partial x^i \partial x^j}(a)$ and $L(a)_{\alpha\beta} = f_{\alpha\beta}(a)$, the summation is over finite graphs with two types of edges: fermionic edges (dashed), and bosonic edges (solid), $c(D(\Gamma))$ is the number of crossings of fermionic edges in the diagram. Weights of edges (propagators) and of vertices are given in Fig. 3.8. An example is given in Fig. 3.9.

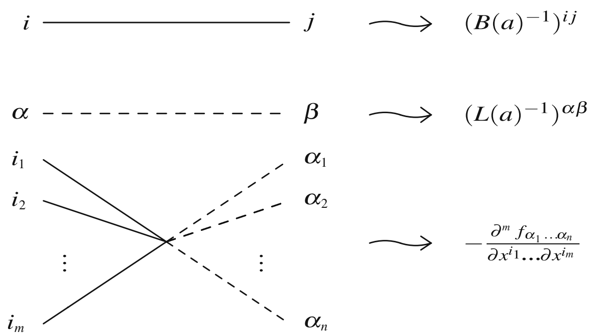


Fig. 3.8 Weights for Feynman diagrams in (27)

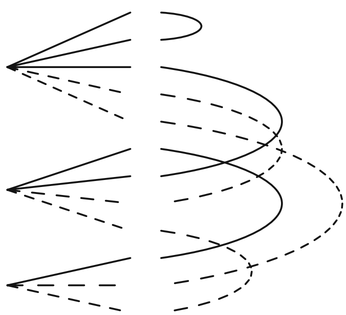


Fig. 3.9 An example of the Feynman diagram for super-integrals

3.5.4 Charged Fermions

Assume that $m = 2k$. Denote $c^i = c^i, \bar{c}^i = c^{k+i}$ for $i = 1, \dots, k$. Assume that the function f in (26) has the form

$$f(x, c, \bar{c}) = f_0(x) + \sum_{\alpha,\beta=1}^{k} f_{\alpha\bar{\beta}}(x) c^{\alpha} \bar{c}^{\bar{\beta}} + \cdots,$$

where ... denote terms of higher order in $c, \bar{c}$.

In this case the asymptotic expansion of the integral (26) is given by Feynman diagrams with *oriented* fermionic edges:

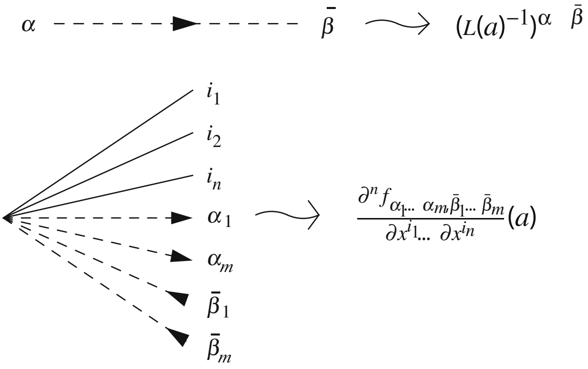


Fig. 3.10 Weights for Feynman diagrams in (28)

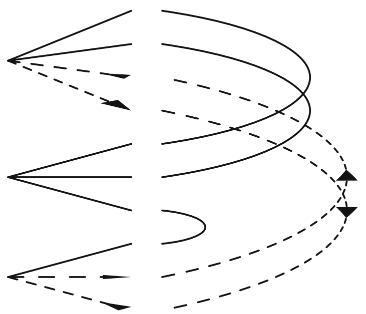


Fig. 3.11 Pairing with oriented edges, producing a Feynman diagram

$$\int_{M_{(n|2k)}} \exp\left(\frac{if(x,c)}{h}\right) g(x,c)dxdc = h^{\frac{n-2k}{2}}(2\pi)^{\frac{n}{2}}$$
$$\sum_a \frac{1}{\sqrt{|\det(B(a))|}} \det(L(a)) \exp\left(\frac{i}{h}f(a) + \frac{i\pi}{4}\operatorname{sign}(B(a))\right)$$
$$\left(1 + \sum_{\Gamma\neq\emptyset} \frac{(ih)^{-\chi(\Gamma)}(-1)^{c(D(\Gamma))}F_a(D(\Gamma))}{|\operatorname{Aut}(\Gamma)|}\right), \tag{28}$$

where all ingredients are the same as in (27) except that the summation is taken over the graphs with oriented fermionic edges and with weights from Fig. 3.10. An example is given in Fig. 3.11.

3.6 Finite-Dimensional Faddeev–Popov Quantization and the BRST Differential

In this section we will study the integral (13) when a Lie group G acts on X faithfully (with no stabilizers) and the function f is invariant with respect to this action.

3.6.1 Faddeev–Popov Trick

Let X be a manifold with the action of a Lie group G. We assume here that the action is free, i.e., that the stabilizer of every point in X is trivial. Assume also that X/G is a manifold. (Note that what is really important is the assumption that X/G is smooth near orbits where f is critical.) In this case

$$\dim(X/G) = \dim(X) - \dim(G).$$

Assume that the manifold X has a G-invariant volume form and that X is compact. It is clear that such restrictions are too strong, but we will see in the next section how they can be relaxed to reasonable assumptions.

Let $f(x)$ be a G-invariant real analytic function. The goal of this section is to prepare the setup for the description of the asymptotic expansion of the integral

$$I_h = \int_X \exp\left(i\frac{f(x)}{h}\right) dx \tag{29}$$

as the sum of Feynman diagrams, just as it was done in section for functions on X with simple critical points.

Since the function f is G-invariant, its critical points are not simple, except when a critical point is a fixed point of the G-action, but since we assume faithfulness, there are no such points.

Instead of assuming the simplicity of critical points of f we assume that critical variety $C_f = \{x \in X|df(x) = 0\}$ of f is the disjoint union of finitely many G-orbits.

We want to change the integration over X to the integration over the orbits of the G-action. In practice, it is convenient to describe the space of orbits in terms of a cross section.

Let us assume that the surface

$$S_\phi = \{x \in X|\phi^a(x) = 0, a = 1, \ldots, n\},$$

where $\phi^a(x)$, $a = 1, \ldots, n$, with $n = \dim(G)$ are independent functions, is a cross-section, i.e., intersects every G-orbit exactly once.

Let x^i, $i = 1, \ldots, d$, be local coordinates on $U \subset X$, e_a, $a = 1, \ldots, n$ be a linear basis in the Lie algebra $\mathfrak{g} = Lie(G)$. Denote by $D^i_a(x)$ the matrix describing the action of e_a as a vector field on X in terms of local coordinates x^i:

$$(e_a f)(x) = \sum_{i=1}^{d} D^i_{a'}(x)\frac{\partial f}{\partial x^i}(x)$$

and by $L^b_c(x)$ the matrix:

$$L^b_a(x) = \sum_{i=1}^{d} D^i_a(x)\frac{\partial \phi^b}{\partial x^i}(x) = e_a\phi^b(x).$$

Since we assume that S_ϕ is a cross section, $\det(L) \neq 0$ on this surface. Later we will relax this condition requiring only that the determinant is not vanishing in a vicińity of critical points of f.

A coordinate free way to formulate this non-degeneracy condition can be phrased as follows. For $x \in S_\phi \subset X$ let $L_x \subset T_x X$ be the subspace of the tangent space spanned by vector field describing the action of the Lie algebra $\mathfrak{g}$ on X, and $T_x S_\phi \subset T_x X$ be the tangent space to S_ϕ at this point. The non-degeneracy of L is equivalent to linear independence of L_x and $T_x S_\phi$ in $T_x X$.

Theorem 3 *(Faddeev–Popov)*[6] *The integral in question is given by*

$$\int_X \exp\left(i\frac{f(x)}{h}\right) dx = h^n|G| \int_L \exp\left(i\frac{f_{FP}(x, c, \bar{c}, \lambda)}{h}\right) dx\, d\bar{c}\, dc\, d\lambda, \qquad (30)$$

where the super-manifold L is $X \times \mathfrak{g}_{odd} \times \mathfrak{g}_{odd} \times \mathfrak{g}^*$, $|G|$ is the volume of G with respect to a left invariant measure dg, and

[6] Faddeev and Popov derived the formula (30) in the setting of the Yang–Mills theory, where the symmetry group is infinite-dimensional and only the integration over gauge classes may have a meaning, see [20].

$$f_{FP}(x, c, \bar{c}, \lambda) = f(x) - ih \sum_{a,b=1}^{n} c^a L_a^b(x)\bar{c}_b + \sum_{a=1}^{n} \lambda_a \phi^b(x). \tag{31}$$

Proof From the G-invariance of f

$$\int_X \exp\left(i\frac{f(x)}{h}\right) dx = \int_{X\times G} \exp\left(i\frac{f(x)}{h}\right) \Delta(x)\delta(\phi(x))dxdg, \tag{32}$$

where $\Delta(x)$ is determined by the identity

$$\Delta(x)\int_G \delta(\phi(gx))dg = 1. \tag{33}$$

Here dg is a right-invariant measure on G, i.e. $d(gh) = dg$. Because of this $\Delta(hx) = \Delta(x)$. The G-orbit through x intersects the cross section S_ϕ only once (since it is a cross section). Denote this point g_0x (such g_0 depends on x, it exists because S_ϕ is a cross section and, in particular, intersects all orbits). Then, we have

$$\phi^a(g_0x) = 0.$$

In a vicinity of this point

$$\phi^a\left(\exp\left(\sum_b t^b e_b\right) g_0x\right) = \sum_{b,i} t^b D_b^i(g_0x)\frac{\partial\phi^a(g_0x)}{\partial x^i} + O(t^2) = \sum_b t^b L_b^a(g_0x) + O(t^2).$$

Thus, the identity (33) is equivalent to

$$\Delta(x)\int_{\mathbb{R}^n} \delta(L(g_0x)t)dt = 1,$$

i.e.

$$\Delta(x) = \det(L(g_0x)).$$

Here we identified $T_{g_0}G \simeq \mathbb{R}^n$. Taking this into account we arrive at the formula

$$\int_X \exp\left(i\frac{f(x)}{h}\right) dx = |G| \int_X \exp\left(i\frac{f(x)}{h}\right) \det(L(g_0x))\delta(\phi(g_0x))dx.$$

Taking into account that $g_0 = 1$ when $\phi(x) = 0$, we obtain

$$\int_X \exp\left(i\frac{f(x)}{h}\right) dx = |G| \int_X \exp\left(i\frac{f(x)}{h}\right) \det(L(x))\delta(\phi(x))dx.$$

Expressing $\det(L(x))$ as a fermionic integral and taking into account

$$\delta(\phi) = \int_{\mathbb{R}^n} \exp(i(\phi, \lambda))d\lambda,$$

we arrive at the formula (30).

3.6.2 Feynman Diagrams with Ghost Fermions

Now let us use the formula (30) to derive the Feynman diagram expansion of the integral (29).

Critical points of the function f_{FP} on the super-manifold L are, by definition, critical points of

$$\tilde{f}(x, \lambda) = f(x) + \sum_a \lambda_a \phi^a(x).$$

This is simply the Lagrange multiplier method and by the assumption which we made above critical points of this function on $X \times \mathfrak{g}^*$ are simple. In particular the matrix of second derivatives is non-degenerate near each critical point of this function.

Thus, we can describe the asymptotic expansion of the integral (29) by Feynman diagrams. Applying the formula (28) to the integral (30) we obtain the following asymptotic expansion:

$$\begin{aligned} \int_L \exp\left(\frac{if_{FP}(x, \bar{c}, c, \lambda)}{h}\right) g(x, c, \bar{c}) dx dc d\bar{c} d\lambda \simeq |G| h^{\frac{d-n}{2}} (2\pi)^{\frac{d+n}{2}} \\ \sum_a \frac{1}{\sqrt{|\det(B(a))|}} \det(-iL(a)) \exp\left(\frac{i}{h} f(a) + \frac{i\pi}{4} \operatorname{sign}(B(a))\right) \\ \left(1 + \sum_{\Gamma \neq \emptyset} \frac{(ih)^{-\chi(\Gamma)} (-1)^{c(D(\Gamma))} F_a(D(\Gamma))}{|\operatorname{Aut}(\Gamma)|}\right), \end{aligned} \tag{34}$$

Here the first summation is over the set of critical points of $\tilde{f}$. Feynman diagrams in this formula have bosonic edges and fermionic-oriented edges, $c(D(\Gamma))$ is the number of crossings of fermionic edges. The structure of Feynman diagrams is the same as in (28). The propagators corresponding to Bose and Fermi edges are shown in Fig. 3.12. The weights of vertices are shown in Fig. 3.13.[7]

The asymptotic expansion (34) depends only on how the cross section S_ϕ intersects G-orbits in the infinitesimal neighborhood of critical points of f. In other words, the expansion is defined as long as the linear operators $B(a)$ and $L(a)$ are invertible at all critical points of the function $\tilde{f}(x, \lambda)$. This is equivalent to the condition $T_a S_\phi \cap \mathfrak{g}_a = \{0\}$ where $T_a S_\phi \subset T_a X$ is the tangent space to S_ϕ at a, and $\mathfrak{g}_a$ is the subspace in $T_a X$ spanned by vector fields describing the infinitesimal action of the Lie algebra of G.

The main moral of this observation is that in order to have the asymptotic expansion of the integral in terms of Feynman diagrams we just have to choose a constraint which is a cross section through the orbits in an infinitesimal neighborhood of critical orbits.

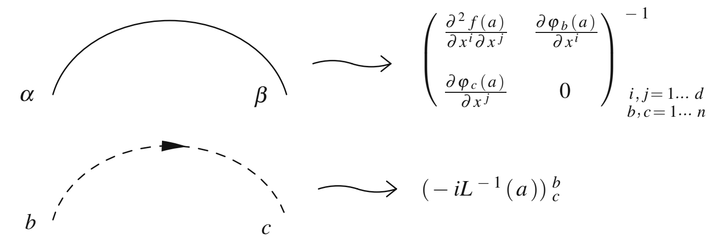


Fig. 3.12 Weights of edges for Feynman diagrams in (34)

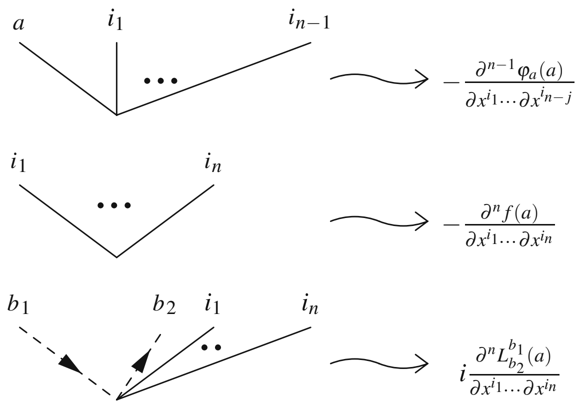


Fig. 3.13 Weights of vertices for Feynman diagrams in (34)

[7] Each fermionic propagator contributes to the weight of the diagram an extra factor h^{-1}. Each vertex with two adjacent fermionic (dashed) edges contributes the factor of h. Because fermionic lines form loops, these factors cancel each other.

3.6.3 Gauge Independence

The asymptotic expansion of the integral (29) does not depend on the choice of the constraint ϕ (as long as it is a cross section through the G-orbits of tangent spaces at critical points).

However, it is not obvious from the Feynman diagram formula for the asymptotic expansion. Let us check that the semiclassical term of the expansion does not depend on ϕ. Till the end of this section we work in a vicinity of a critical point of f_{FP}. The semiclassical term is

$$\det(B)^{-\frac{1}{2}}\det(L),$$

where

$$B = \begin{pmatrix} \frac{\partial^2 f}{\partial x^i \partial x^j} & \frac{\partial \phi^b}{\partial x^i} \\ \frac{\partial \phi^a}{\partial x^j} & 0 \end{pmatrix} \tag{35}$$

and

$$L_c^b = \sum_i l_c^i \frac{\partial \phi^b}{\partial x^i}. \tag{36}$$

Let us make an infinitesimal variation of the constraint $\phi^a(x) \mapsto \phi^a(x) + \varepsilon^a(x)$. The product of the determinants will change as

$$\det(B)^{-\frac{1}{2}}\det(L) \mapsto \det(B)^{-\frac{1}{2}}\det(L)\left(1 - \frac{1}{2}\operatorname{tr}(B^{-1}\delta B) + \operatorname{tr}(L^{-1}\delta L) + \cdots\right),$$

where ... are higher order terms. We have to prove that the first-order terms vanish. The matrix B has the block form, so is the matrix B^{-1}. Both of these matrices are symmetric, therefore

$$-\frac{1}{2}\operatorname{tr}(B^{-1}\delta B) = -\operatorname{tr}((B^{-1})_{12}\delta B_{21}) = -\sum_{i,c} b_c^i \frac{\partial \varepsilon^c}{\partial x^i},$$

where b_a^i are matrix elements of the block $(B^{-1})_{12}$. They satisfy the identity $\sum_i \frac{\partial \phi^b}{\partial x^i} b_c^i = \delta_c^b$.

The second term can be written as

$$\mathrm{tr}(L^{-1}\delta L) = \sum_{b,c,i} (L^{-1})^c_b l^i_c \frac{\partial \varepsilon^b}{\partial x^i}.$$

Using the identity $\sum_i (L^{-1}l)^i_b \frac{\partial \phi^c}{\partial x^i} = \delta^c_b$ and the corresponding identity for b we conclude that

$$-\frac{1}{2}\,\mathrm{tr}(B^{-1}\delta B) + \mathrm{tr}(L^{-1}\delta L) = 0,$$

which proves that the semiclassical factor does not depend on the choice of the gauge condition.

We will leave the exercise of verifying this fact in all orders ≥ 1 to the reader.

3.6.4 Feynman Diagrams for Linear Constraints

Because the asymptotic expansion depends only on the formal neighborhood of critical points of $f(x)$ on the surface of the constrains and does not depend on the particular choice of the constraint (as long as it is a local cross section in the neighborhood of each critical point), we can choose them at our convenience at each neighborhood.

In particular, if X is linear, we can deform ϕ to a linear cross section in a formal neighborhood of each critical point. Now let us find the asymptotic expansion of the integral

$$\int_{X_a} \exp\left(i\frac{f(x)}{h}\right) \det(L(x))\delta(\phi(x))dx, \tag{37}$$

where X_a is an infinitesimal neighborhood of $a \in X$.

For the integral (37) we obtain

$$\int_{X_a} \exp\left(i\frac{f(x)}{h}\right)\det(L(x))\delta(\phi(x))dx = \int_{\ker(\phi)} \exp\left(i\frac{f(s)}{h}\right)\det(L(s))\det(L(a))^{-1}ds,$$

Since the constraint is linear, $L^b_a(s) = \sum_i l^i_a(s)\varphi^a_i$. Here $l^c_a(s)$ is the matrix describing the action of $\mathfrak{g}$ on $\mathfrak{g}(a) \subset T_aX$.

Change coordinates to $x^i = a^i + \sum_a l^i_a(a)X^a + \sum_\alpha \psi^i_\alpha s^\alpha$ where ψ_α is a basis in $\ker(\phi)$. We assume $a \epsilon \ker(\phi)$

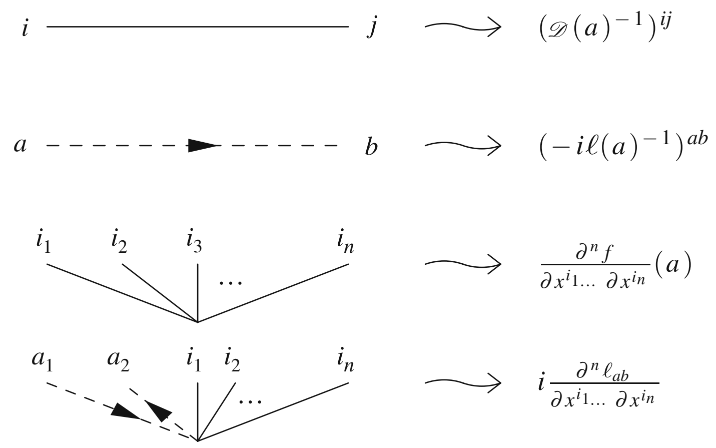


Fig. 3.14 Weights of Feynman diagrams in (38)

Its contribution to the asymptotic expansion is given by the formal power series where coefficients are determined by Feynman diagrams with rules described in Fig. 3.14. This power series does not depend on the choice of X_ϕ. Indeed different choices of ϕ are related by linear transformations in X_ϕ. The contribution from each Feynman diagram is invariant with respect to linear transformations of s-coordinates and therefore does not depend on the choice of ϕ.

Finally, we can write the asymptotic expansion of (30) as

$$\int_X \exp\left(i\frac{f(x)}{h}\right)dx \simeq |G|h^{\frac{d-n}{2}}(2\pi)^{\frac{d+n}{2}}$$
$$\sum_a \frac{1}{\sqrt{|\det(D(a))|}}\det(-il(a))\exp\left(\frac{i}{h}f(a)+\frac{i\pi}{4}\operatorname{sign}(D(a))\right)$$
$$\left(1+\sum_{\Gamma\neq\emptyset}\frac{(ih)^{-\chi(\Gamma)}(-1)^{c(D(\Gamma))}F_a(D(\Gamma))}{|\operatorname{Aut}(\Gamma)|}\right). \tag{38}$$

Here $D(a)_{ij}=\frac{\partial^2 f}{\partial s^i\partial s^j}$ where s^i are coordinates on X_ϕ. The coefficients are given by Feynman diagrams with weights of edges and vertices described in Fig. 3.14, and all other ingredients are as before.

The factor $\exp\left(\frac{i\pi}{4}\operatorname{sign}(B)\right)$ can also be written as $i^N\exp(-\frac{i\pi}{2}n_-(B))$ where $n_-(B)$ is the number of negative eigenvalues of B. This is more or less how the *Morse index* appears in the semiclassical asymptotic of the propagator in quantum mechanics.

3.6.5 The BRST Differential

The appearance of fermionic variables (Faddeev – Popov ghost fields) in the asymptotic expansion of (30) looks as a bit of a mystery and as a technical trick. In the BRST approach these non-commutative variables attain a natural meaning.

The key observation of Becchi et al. [12] and of Tuytin [54][8] is that the odd operator Q

$$Q = \sum_{a,i=1}^{n,d} c^a L_a^i \frac{\partial}{\partial x^i} - \frac{1}{2} \sum_{a,b,c} f_{bc}^a c^b c^c \frac{\partial}{\partial c^a} + \sum_a \lambda_a \frac{\partial}{\bar{c}_a}$$

acting on the space $C^\infty(L) = Fun(X \times \mathfrak{g}^*) \otimes \mathbb{C}[c^a, \bar{c}_a] = C^\infty(X \times \mathfrak{g}^*) \otimes \wedge(\mathfrak{g} \oplus \mathfrak{g}^*)$ of functions on the super-manifold $L = X \times \mathfrak{g}_{\mathrm{odd}} \times \mathfrak{g}_{\mathrm{odd}} \times \mathfrak{g}^*$ has the properties

$$Q^2 = 0,$$
$$Qf_{FP} = 0.$$

The first property means that the pair $(C^\infty(L), Q)$ is a co-chain complex. Because we assumed that the action of G on X is faithful, its zero cohomology can be naturally identified with $C^\infty(X/G)$, and the other cohomologies vanish. Note that $Q = Q_{Ch} + Q_K$, where the first term is the differential in the *Chevalley complex* for $\mathfrak{g}$ with the coefficients in $C^\infty(X)$. The second term $Q_K = \sum_a \lambda_a \frac{\partial}{\bar{c}_a}$ is the *Koszul differential* for functions on $\mathfrak{g}^*$.

The second property means that the Faddeev – Popov action is a cocycle in the BRST complex. The function f_{FP} is not a co-boundary and therefore defines a non-trivial zero cohomology class in $H^0(L) \simeq C^\infty(X/G)$. This class is simply the initial function f considered as a function on G-orbits. Indeed, the function f_{FP} can be written as $f_{FP} = f + Q\left(\sum_a \phi^a \bar{c}_a\right)$.

To see how the integral over the super-space L appears in this setting, consider first a simple fact in linear algebra.

Let C be a super-vector space and $d : C \to C$ be an odd linear operator with $d^2 = 0$. Assume D is another super-vector space with an odd differential $d^* : D \to D, d^{*2} = 0$ and a non-degenerate pairing $\langle ., . \rangle : D \otimes C \to \mathbb{C}$ such that $\langle d^* l, a \rangle = (-1)^{\bar{l}} \langle l, da \rangle$.

We will think of (D, d^*) and (C, d) as co-chain complexes and say that $l \in D$ and $a \in C$ are cocycles, if $d^* l = 0$ and $da = 0$. Denote by $[l] \in H(D) = \mathrm{Ker}(d^*)/\mathrm{Im}(d^*)$ and $[a] \in H(C) = \mathrm{Ker}(d)/\mathrm{Im}(d)$ the cohomology classes of the cocycles l and a.

Lemma 3 *If l and a are cocycles, then*

$$\langle l, a \rangle = \langle [l], [a] \rangle \,,$$

[8] The original formulation uses the supersymmetry concept and has a slightly different appearance.

where $\langle [l], a] \rangle$ is the induced pairing on the cohomology spaces.

Indeed, the cocycle properties imply that

$$\left\langle l + d^* m, a + dc \right\rangle = \langle l, a \rangle \,,$$

which defines the pairing on the cohomology spaces and proves the lemma.

Now we should identify ingredients of this lemma in the FP-BRST setting. The G-invariance of the measure of integration $dxd\bar{c}dcd\lambda$ in the FP integral which we will denote by dl implies[9]

$$\int_L Qgdl = 0$$

Considering the integral as a linear functional on $C = C^\infty(L)$ with the differential we can think of it as an element of D which is annihilated by d^*.

Applying the lemma to a cocycle $g \in C^\infty(L)$, i.e., to a function, such that $Qg = 0$ we arrive at the identity

$$\int_L gdl = \int_Y [g]dy. \tag{39}$$

Here Y is the super-manifold such that $H^0(C^\infty(L)) = C^\infty(Y)$, i.e., the appropriate topological version of X/G. If we would be in algebro-geometric setting, the variety Y would be the spectrum of the commutative algebra H^0. We also made an assumption that all cohomologies except H^0 are vanishing, which is in our setting equivalent to the faithfulness of the G-action on X.

Equation (39) implies, in particular, that if $Qg = 0$ (i.e., if g is G-invariant) and if the measure is G-invariant, then

$$\int_L \exp\left(\frac{if_{FP}}{h}\right) gdl = \int_Y \exp\left(\frac{if}{h}\right) [g]dy.$$

This puts the Faddeev–Popov method into a natural algebraic setting and "explains" the algebraic meaning of fermionic ghost fields. It also shows that the method can be extended to any complex which has $C^\infty(X/G)$ as its cohomology space. Because of the formula (39) it does not matter with which complex $(C^\infty(L), Q)$ we started, as long as the cohomology space is $C^\infty(X/G)$. This observation leads to

---

[9] The operator Q can be regarded as an super-vector field on L. The invariance of the measure dl is equivalent to the zero-divergence condition of the vector-field (with respect to the measure dl). Recall that for any vector field Q we have

$$\int_L Qgdl = \int_L g \,\mathrm{div}_{dl}(Q)dl$$

where $\mathrm{div}_{dl}(Q)$ is the divergence of the vector field Q with respect to the volume measure dl.

an important notion of cohomological field theories [57] and to a natural notion of quasi-isomorphic field theories.

Perhaps one of the most important developments along these lines is the extension of the BRST observations to a more general class of degenerate Lagrangians (i.e., degenerate critical points of f). This generalization known as Batalin–Vilkovisky quantization (BV) works even in the case when the Lagrangian is invariant with respect to the action of vector fields, which do not necessarily form a Lie algebra. One of the most striking applications of this technique was the quantization of the Poisson sigma model and the construction of the *-product for an arbitrary Poisson manifold. But this subject goes beyond the goal of the present lectures.

3.7 Semiclassical Quantization of a Scalar Bose Field

The classical theory of a scalar Bose field is described in Sect. 3.2.5. Let us define the amplitude $Z(M)$ as a semiclassical expansion of a (non-existing) path integral given by Feynman diagrams similar to how the asymptotic expansion looks for finite-dimensional integrals.

This definition can be motivated by finite-dimensional approximations to the path integral, which provide an acceptable definition of *infinite-dimensional integrals* such as Wiener integral and path integrals in low-dimensional Euclidean quantum field theories [30].

In semiclassical quantum field theory, path integrals are defined as formal power series which have the same structure as if they were asymptotical expansions of existing integrals. The coefficients in these expansions are given by Feynman integrals. We will show how it works in quantum mechanics, and how it compares with the semiclassical analysis of the Schrödinger equation for $d = 1$. We will have a brief discussion of the $d > 1$ case, as well.

3.7.1 Formal Semiclassical Quantum Mechanics

3.7.1.1 Semiclassical Asymptotics from the Schrödinger Equation

To be specific, we will consider here quantum mechanics of a point particle on a Riemannian manifold N in a potential $V(q)$ (see Sect. 3.2.3).

Let $\{\gamma_c(t)\}_{t_1}^{t_2}$ be a solution to the Euler – Lagrange equations for a classical Lagrangian $\mathcal{L}(\xi, q)$ with Dirichlet boundary conditions $\gamma(t_1) = q_1, \gamma(t_2) = q_2$. Denote by $S^{(c)}_{t_2-t_1}(q_2, q_1)$ the value of the classical action functional on γ_c:

$$S^{(c)}_{t_2-t_1}(q_2, q_1) = \int_{t_1}^{t_2} \mathcal{L}(\dot{\gamma}_c(t), \gamma_c(t))dt.$$

Let $U_t(q_2, q_1)$ be the kernel of the integral operator representing the evolution operator (11). Solving Schrödinger equation (12) in the limit $h \to 0$ we obtain the following asymptotics of the evolution kernel as $h \to 0$

$$U_t(q_2, q_1) \sim \sum_{\gamma_c} (2\pi i)^{-\frac{n}{2}} \exp\left(\frac{i}{h} S_t^{(c)}(q_2, q_1) + \frac{i\pi\mu(\gamma_c)}{2}\right)$$
$$\left| \wedge^n \left(\frac{\partial^2 S_t^{(c)}(q_2, q_1)}{\partial q_2 \partial q_1} dq_2 \wedge dq_1 \right) \right|^{\frac{1}{2}} \left(1 + \sum_{n \geq 1} h^n U_c^{(n)}(q_2, q_1) \right). \quad (40)$$

Here

$$\wedge^n \left(\frac{\partial^2 S(a, b)}{\partial a \partial b} da \wedge db \right)$$
$$= \wedge^n d_a d_b S(a, b) = \det \left(\frac{\partial^2 S(a, b)}{\partial a^i \partial b^j} \right) da^1 \wedge \dots da^n \wedge db^1 \wedge \dots db^n. \quad (41)$$

$\mu(\gamma_c)$ is the Morse index of γ_c (i.e., the number of focal points of the trajectory in $T^*\mathbb{R}^n$ induced by γ_c relative to fibers of the cotangent bundle). The coefficients $a_k^{(c)} = (2\pi i)^{-\frac{n}{2}} \left(\det \left(\frac{\partial^2 S(a,b)}{\partial a^i \partial b^j} \right) \right)^{\frac{1}{2}} U_c^{(n)}$ satisfy the transport equation

$$\frac{\partial a_k^{(c)}}{\partial t} + \frac{1}{2m} \Delta S^{(c)} a_k^{(c)} + \frac{1}{m} \sum_{j=1}^n \frac{\partial S}{\partial q_j} \frac{\partial a_k^{(c)}}{\partial q_j} + \frac{i}{2m} \Delta a_{k-1}^{(c)} = 0.$$

However, the initial condition $\lim_{t \to +0} U_t(q, q') = \delta(q, q')$ can no longer be imposed since we consider the asymptotical expansion when $h << t$. Instead, to determine the coefficients $a_k^{(c)}$, one should use the semigroup property of the propagator:

$$U_t U_s = U_{s+t}.$$

The kernel of the integral operators representing the evolution operator satisfies the identity

$$\int_N U_t(q_3, q_2) U_s(q_2, q_1) = U_{s+t}(q_3, q_1). \quad (42)$$

Here the first factor is a half-density in q_3, q_2, the second is a half-density in q_2, q_1. The product is a density in q_2 and it is integrated over N.

As $h \to 0$ the semigroup property implies that the asymptotical expansion should satisfy the identity

$$\sum_{k,l\geq 0}\int_N \exp\left(\frac{i(S_t^{(c')}(q_3,q_2)+S_s^{(c'')}(q_2,q_1))}{h}\right)a_k^{(c')}a_l^{(c'')}$$
$$=\sum_k \exp\left(\frac{iS_t^{(c)}(q_3,q_1)}{h}\right)a_k^{(c)}.$$

Here by the integral of the product of two half-densities on N we mean the formal asymptotic expansion (17), and $\gamma_{c'}, \gamma_{c''}$ are parts of the path $\{\gamma_c\}_{t=0}^{t+s}$ when $0 < \tau < s$ and $s < \tau < s+t$, respectively.

It is not difficult to derive the first coefficients of the asymptotical expansion (40) from this equation. Moreover, this equation alone defines all higher order terms in the semiclassical expansion.

For more details on the semiclassical analysis see, for example, [50].

3.7.1.2 Semiclassical Expansion from the Path Integral

Looking at the expression (40), it is natural to imagine that it may be interpreted as a semiclassical asymptotics of an oscillating integral over the space of paths connecting the points q_1 and q_2. Critical points in this integral are classical trajectories.

This point of view was put forward in quantum mechanics by R. Feynman and it can be supported by many very convincing arguments [22]. Eventually, a mathematically meaningful definition of a path integral for the Euclidian version (when the integral is rapidly decaying instead of oscillating) emerged and was developed further in the framework of constructive field theory. The *Wiener integral*, which was introduced in probability theory, even earlier, is an example of such an object.

Here we will not try to make the definition of the integral rigorous. Instead of this we will define its semiclassical expansion in such a way that it has an appearance of a semiclassical expansion of an infinite-dimensional integral. After this we will check that it satisfies the semigroup property. This is an illustration of a semiclassical quantum field theory, where the partition function Z_M depends on the boundary condition, and integrating over possible boundary conditions has the replicating gluing property (9). The difference is that in quantum mechanics we have the Schrödinger equation as a reference point to compare any definition of the path integral. In the more complicated models of quantum field theory, the gluing axioms seem to be the only major structural requirement (beyond unitarity and causality, which we do not discuss here).

So, we are looking for a formal power series which would look like the asymptotic expansion of the integral

$$Z_t(q_2,q_1)=\int_{\gamma(0)=q_1,\gamma(t)=q_2}\exp\left(\frac{i}{h}S[\gamma]\right)D\gamma.$$

We will focus in this section on the point particle of mass m in $\mathbb{R}^n$ in the potential $V(q)$ (3). By analogy with the finite-dimensional case, we define the asymptotic expansion as

$$Z_t(q_2, q_1) = C \sum_{\gamma_c} \exp\left(\frac{i}{h} S_t^{(c)}(q_2, q_1) - \frac{i\pi}{2} n_-(K^{(c)})\right)$$
$$|\det{}'(K^{(c)})|^{-\frac{1}{2}} |dq_1|^{\frac{1}{2}} |dq_2|^{\frac{1}{2}} \left(1 + \sum_{\Gamma \neq \emptyset} (ih)^{-\chi(\Gamma)} \frac{F_c(\Gamma)}{|\operatorname{Aut}(\Gamma)|}\right). \quad (43)$$

Here

$$(K^{(c)})_{ij} = -mh^2 \frac{d^2}{d\tau^2} \delta_{ij} + \frac{\partial^2 V}{\partial x^i \partial x^j}(\gamma_c(\tau))$$

is the matrix differential operator which acts on the space of functions on $[0, t]$ with values in $\mathbb{R}^n$ (trivialized tangent bundle to N restricted to $\gamma^{(c)}$ in local coordinates) with the Dirichlet boundary conditions $f(0) = f(t) = 0$. The half-density $|dq|^{\frac{1}{2}}$ is the "square root" of the Riemannian volume density on N. The sum is taken over classical trajectories connecting q_1 and q_2, and $n_-(K^{(c)})$ denotes the number of negative eigenvalues of the operator $K^{(c)}$, C is some constant. The weights for Feynman diagrams in (43) are given in Fig. 3.15, where $G^{ij}(x, y)$ is the kernel of the integral operator which is the inverse to $K^{(c)}$.

The expansion is not the result of computation. It is a definition, which is based on the idea that the path integral exists in some sense and its asymptotical expansion as $h \to 0$ is given by a formula similar to the finite-dimensional case. It turns out that despite very different appearance the semiclassical expansion of U_t coincides with this series.

One can show easily (see, for example, [50]) that

$$|\det{}'(K^{(c)})| = \left|\det\left(\frac{\partial^2 S(a, b)}{\partial a_i \partial b_j}\right)\right|^{-1},$$

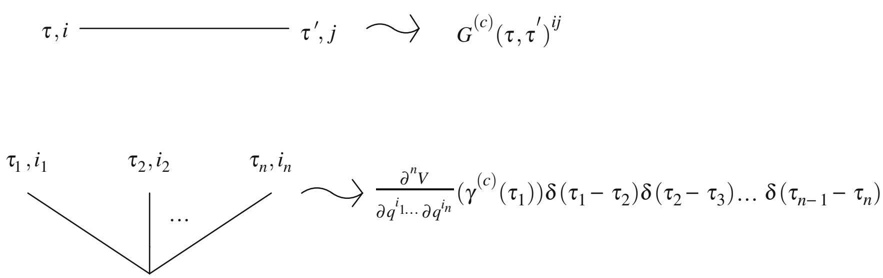


Fig. 3.15 Weights of Feynman diagrams in (43)

as well as that $\mu(\gamma^{(c)})$ can be identified with $n_-(K^{(c)})$. This shows that the leading terms of (43) and (40) are the same. Now the question is whether the two power series are the same.

We will state without proof the following theorem.

Theorem 4 *The expansion $Z_t(q_2, q_1)$ is equal to the asymptotic expansion of the kernel of the propagator and it satisfies the gluing property.*

The details will appear in a paper by T. Johnson-Freyd [38] when $N = \mathbb{R}^d$ with flat metric.

As an immediate corollary to this theorem we have

Corollary 1 *The functions*

$$U_c^{(n)}(q_2, q_1)) = \sum_{-\chi(\Gamma)=n} \frac{F_c(\Gamma)}{|\operatorname{Aut}(\Gamma)|}$$

are coefficients of the asymptotical expansion of the propagator and, after being properly normalized, satisfy the transport equation. Here $\chi(\Gamma) = |V| - |E|$ is the Euler characteristic.

Let us write the semiclassical propagator as

$$Z_t(q_2, q_1) = \sum_c \exp\left(\frac{i}{h} S_t^{(c)}(q_2, q_1)\right) J_t^{(c)}(q_2, q_1).$$

The semigroup property of the propagator implies that this power series satisfies the following gluing/cutting identity:

$$\exp\left(\frac{i}{h} S_t^{(c)}(q_3, q_1)\right) J_t^{(c)}(q_3, q_1) =$$
$$\int_{q_2 \in N} \exp\left(\frac{i}{h}\left(S_s^{(c)}(q_3, q_2) + S_{t-s}^{(c)}(q_2, q_1)\right)\right) J_s^{(c)}(q_3, q_2) J_{t-s}^{(c)}(q_2, q_1). \quad (44)$$

Here the integral is taken in a sense of the semiclassical expansion as the sum of corresponding Feynman diagrams. It is easy to check that the identity (44) determines uniquely not only the higher order coefficients but also the leading order factor.

3.7.2 $d > 1$ and Ultraviolet Divergencies

In the semiclassical theory of scalar Bose field on a compact Riemannian manifold the partition function for the theory is given by the formal power series

$$Z_M(b) = C \sum_{\phi_c} \exp\left(\frac{iS_M(\phi_c)}{h} - \frac{i\pi}{2} n_-(K_{\phi_c})\right) |\det{}'(K_{\phi_c})|^{-\frac{1}{2}}$$
$$\left(1 + \sum_{\Gamma \neq \emptyset} (ih)^{-\chi(\Gamma)} \frac{F_{\phi_c}(\Gamma)}{|\operatorname{Aut}(\Gamma)|}\right). \tag{45}$$

Here we assume that there are finitely many solutions ϕ_c to the Euler–Lagrange equation (4) with the Dirichlet boundary conditions $\phi_c|_{\partial M}=b$. The number $n_-(K_{\phi_c})$ denotes the number of negative eigenvalues of the differential operator

$$K_{\phi_c} = \Delta + V''(\phi_c(x))$$

acting on the space of functions on M with the boundary condition $f(x) = 0, x \in \partial M$, and $\det'(K_{\phi_c})$ is its regularized determinant ($-\frac{\pi}{2} n_-(K_{\phi_c})$ is the phase of the square root of the determinant). The ζ-function regularization is one of the standard ways to define $\det'$ (see for example [6]). The weights of Feynman diagrams are given in Fig. 3.16, where $G^{(c)}(x, y)$ is the kernel of the integral operator which is inverse to K_{ϕ_c}.

An example of an order one Feynman diagram is given in Fig. 3.17.

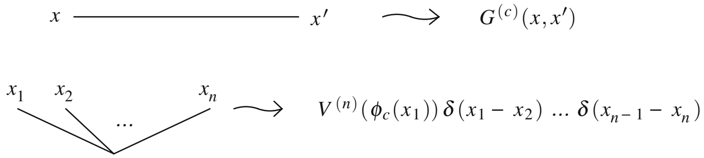


Fig. 3.16 Weights of Feynman diagrams in (45)

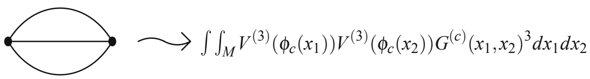


Fig. 3.17 An example of the Feynman diagram of order one

The kernel $G(x, y)$ behaves at the diagonal as

$$G(x, y) \sim |x - y|^{2-d},$$

which means that the Feynman integrals converge for $d = 1$ (quantum mechanics), diverge logarithmically for $d = 2$, and diverge as a power of the distance for $d > 2$.

This is a well-known problem of ultraviolet divergencies in the perturbation theory. The usual way to deal with divergencies is a two-step procedure.

Step 1. The theory is replaced by a family of theories where the Feynman integrals converge (regularized theories). There are several standard ways to do this:

- *Higher derivative regularization* replaces the theory with one where the quadratic part of the action has terms with higher derivatives. In the regularized theory the propagator $G(x, y)$ is not singular at the diagonal. For more details see [35].
- *Lattice regularization* replaces the theory on a smooth Riemannian manifold M by a metrized cell approximation of M. The path integral becomes finite-dimensional and Feynman diagrams describing the semiclassical expansion become finite sums.
- *Dimensional regularization* is more exotic. It replaces Feynman d-dimensional integrals, where d is an integer, by formal expressions, where d is not an integer. It is very convenient computationally for certain tasks (see, for example, [19] and references therein).

Step 2. After the theory is replaced by a family of theories where Feynman integrals converge, one should compute them and pass to the limit corresponding to the original theory. Of course the limit will not exist since some terms will have singularities. In some cases it is possible to make the parameters in the regularized theory (for example, coefficients in $V(\phi)$) depend on the parameters of the regularization in such a way that the sum of Feynman diagrams of order up to n remains finite when the regularization is removed. Such theories are called *renormalizable in orders up to n*.

The compatibility of the gluing/cutting axiom, i.e., an analog of the identity (44), and the renormalization is, basically, an open problem for $d > 1$, which requires further investigation. Notice that for $d = 2$ the integration over the boundary fields does not introduce Feynman diagrams with ultraviolet divergencies, but these diagrams will diverge for $d > 2$. This problem was addressed in the case of Minkowski flat space – time by K. Symanzik in [48].

3.8 The Yang–Mills Theory

The classical Yang–Mills theory with Dirichlet boundary conditions was described in Sect. 3.2.6.

In this section we will define Feynman diagrams for the Yang–Mills theory following the analogy with the finite-dimensional case. In these notes we will do it "half-way," leaving the most important part concerned with the ultraviolet divergencies aside.

Naively, the path integral quantization of the classical d-dimensional Yang–Mills theory can be constructed as follows. Let G be a compact Lie group.

- To a closed oriented $(d - 1)$-dimensional Riemannian manifold with a principal G-bundle P we assign the space of functionals on the space of connections on P.
- To a d-dimensional Riemannian manifold M with a principal G-bundle on it, we define the functional Z on the space of connections on $P|_{\partial M}$ as

$$Z_M(b) = \int_{i^*(A)=b} \exp\left(\frac{i}{h} S_{YM}(A)\right) DA.$$

where $i : \partial M \hookrightarrow M$ is the tautological inclusion of the boundary and $i^*(A)$ is the pullback of the connection A to the boundary.

Now we can use the Faddeev–Popov Feynman diagrams to *define* the semiclassical expansion of this integral. In the finite-dimensional case, Feynman diagrams were derived as an asymptotic expansion of the existing integral. To define such expansion, we should do the gauge fixing and then define the Feynman rules. The Feynman diagrams for the Yang–Mills are divergent because the propagator is singular at the diagonal (ultraviolet divergence). Nevertheless, the theory is *renormalizable*, as in the previous example, even better, it is *asymptotically free* [31]. We will not go into the details of the discussion of renormalization but will make a few remarks at the end of this section. For more details about quantum Yang–Mills theory and Feynman diagrams see [21].

3.8.1 The Gauge Fixing

As we have seen in the finite-dimensional case, the constraint (gauge fixing) needed to construct the asymptotic expansion of the integral (30) has to be a cross section through the orbits only in the vicinity of critical points (critical orbits) of the action functional. To define Feynman diagrams for the Yang–Mills theory, we can follow the same logic. In particular, we can choose a linear Lorentz gauge condition for connections in the vicinity of a classical solution A.

For a connection $A+\alpha$, where α is a 1-form (quantum fluctuation around A), the Lorentz gauge condition is

$$d_A * \alpha = 0, \tag{46}$$

where $*$ is the Hodge operation. This condition defines a subspace in the linear space of $\mathfrak{g}$-valued 1-forms, so we can use the formula (38) which uses no Lagrange multipliers. The contribution to the path integral from a vicinity of A is then an "integral" over the space of 1-forms α from $Ker(d_A^*)$. In other words, to define the semiclassical asymptotic of the partition function for the Yang–Mills theory, we can try the Faddeev–Popov expansion with the Lorentz gauge condition.

3.8.2 The Faddeev–Popov Action and Feynman Diagrams

Following the analogy with the finite-dimensional case define the Faddeev–Popov action for pure Yang–Mills theory as the following action with fields $\alpha(x)$, $\bar{c}(x)$, $c(x)$:

$$S_A(\alpha) = S_{YM}(A) + \int_M \frac{1}{2} \operatorname{tr} \langle F_A(\alpha), F_A(\alpha) \rangle \, dx$$
$$\times - \frac{ih}{2} \int_M * d_A \bar{c} \wedge d_A c - \frac{ih}{2} \int_M * d_A \bar{c} \wedge [\alpha, c]. \quad (47)$$

Here A is a background connection which is the solution to the classical Yang–Mills equations and α is a $\mathfrak{g}$-valued 1-form on M. The bosonic part of this action is simply $S_{YM}(A + \alpha)$.

The quadratic part in α and the quadratic part in $\bar{c}$, c of the action (47) are given by the differential operator $d_A^* d_A$ which is invertible on the space $Ker(d_A^*)$ with Dirichlet boundary conditions. Other terms define the weights of Feynman diagrams. The weights are shown in Fig. 3.18. The functions G_1^A and G_0^A are Green's functions of the Laplace–Beltrami operator $\Delta = d^* d + d d^*$ on 1- and 0-forms, respectively.

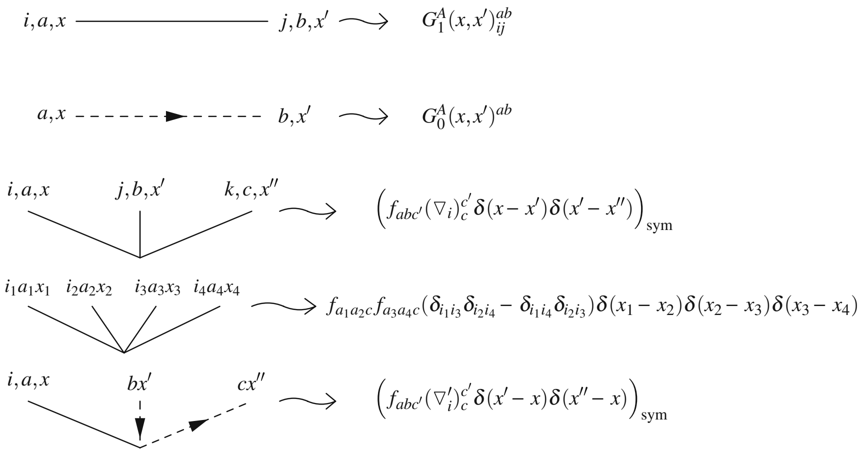


Fig. 3.18 Weights of Feynman diagrams in the semiclassical expansion for the Yang–Mills theory

3.8.3 The Renormalization

The propagator in the Yang–Mills theory is singular at the diagonal for $d > 1$, and just as in the scalar Bose field contributions from Feynman diagrams to the partition function diverge. However, just as in the scalar Bose field, when $d \leq 4$, after the renormalization procedure Feynman diagrams become finite and there is a well-defined semiclassical formal power series for the Yang–Mills given by renormalized diagrams. This fact was discovered by t'Hooft [33] who invented the dimensional regularization of Feynman diagrams and showed that taking into account Faddeev–Popov ghost fields makes Yang–Mills into a renormalizable theory.

Moreover, the renormalization in the Yang–Mills theory is remarkable because it gives an asymptotically free theory. This was discovered in [31] and it means that particles in such theory have to behave as non-interacting, free particles at high energies. This *prediction* perfectly agrees with *experimental data* and this is why the Yang–Mills theory is part of the Standard Model, unifying the theory of strong, weak, and electromagnetic interactions.

The super-symmetric $N = 4$ Yang–Mills theory is expected to have a particularly remarkable renormalization. It turns out that the divergent contributions from Feynman diagrams cancel each other in each order of the expansion in h. This was proven in the light-cone gauge and is believed to be true for other gauges. This Yang–Mills theory is particularly important for Topological Quantum Field Theories [39, 27] and in particular to the quantum field theoretical interpretation of the geometric *Langlands program.*

Finally, few words on *correlation functions*. Since the Yang–Mills theory is gauge invariant, natural observables should also be gauge invariant. Such observables are known as *Wilson loops* or, more generally, as *Wilson graphs*.

Recall the definition of Wilson loops. Let V be a finite-dimensional representation of a Lie group G. The Yang–Mills potential A (the field in the Yang–Mills theory) is a connection in a principal G-bundle P. It induces a connection in the vector bundle $V_P = P \times_G V$. Let

$$h_A(C_x) = P \exp\left(\int_{C_x} A\right) \tag{48}$$

be the parallel transport in V_P along a path C_x which starts and ends at $x \in M$ defined by the connection A. Here P stands for the iterated path-ordered integral.

The Wilson loop observable is

$$W_A^V(C) = \mathrm{Tr}_{V_x}(h_A(C_x)). \tag{49}$$

Here the trace is taken over the fiber V_x of V_P over $x \in M$. The definition of more general gauge invariant observables, Wilson graphs, will be given later, when we will discuss observables in the Chern–Simons theory.

An important conjecture about the Yang–Mills theory, and another fundamental fact expected from this theory, is the *dynamical mass generation*. In terms of expectation values of Wilson loops, this conjecture means that

$$\langle W_A(C)\rangle \propto \exp(-ml(C)), \tag{50}$$

as $l(C) \to \infty$. Here on the left side we have the expectation value of the Wilson loop and on the right side $l(C)$ is the length of C in the Riemannian metric on M. This conjecture is based on the conjecture that the Yang–Mills theory can be defined non-perturbatively.

The parameter m in (50) is characterizing the radius of correlation. In a massless theory, such as Yang–Mills theory, there are no reasons to expect that $m \neq 0$. The appearance of such a parameter with the scaling dimension of the mass is known as dynamical mass generation. For more details about this conjecture see [36].

3.9 The Chern–Simons Theory

In this section M is a compact-oriented manifold. The classical Chern–Simons theory with a compact simple Lie group G was described in Sect. 3.2.8. As in the pure Yang–Mills theory, fields in the Chern–Simons theory are connections in a principal G-bundle over the space–time M. In contrast with the Yang–Mills theory, the Chern–Simons action is the first-order action. One of the implications of this is the difference in Hamiltonian formulations. The other is that the path integral quantization for the Chern–Simons theory for manifolds with boundary is more involved. Some of the aspects of this theory on manifolds with boundaries can be found in references [24, 7].

From now on we assume that the space–time M is a compact, oriented, and closed 3-manifold. The Chern–Simons action is topological, i.e., its definition does not require a choice of metric on M. This is why it is natural to expect that the result of quantization, the partition function $Z(M)$, also depends only on the homeomorphism class of M. This gives a powerful criterium for consistency of the definition of Feynman diagrams: the sum of Feynman diagrams for any given order should depend only on the topology of the manifold.

The path integral formulation of the Chern–Simons theory on manifolds with a boundary is a bit more involved then the one for the Yang–Mills. This is because the Chern–Simons is a first-order theory. The space of states assigned to the boundary is the space of holomorphic sections of the geometric quantization line bundle over the moduli space of flat connections in a trivial principal G-bundle over the boundary (provided we made a choice of complex structure). This space is a quantum counterpart to the boundary conditions for the Chern–Simons theory when the pullback to the boundary is required to be holomorphic. For more details on the quantization of the Chern–Simons theory on manifolds with boundary see for example [7].

So, the goal of this section is to make sense of the expression

$$Z_M = \int \exp(ikCS(A))DA, \tag{51}$$

or, more generally, of

$$Z_{M,\Gamma} = \int \exp(ikCS(A))W_\Gamma(A)DA, \tag{52}$$

where $W_\Gamma(A)$ is a gauge invariant functional (Wilson graph or any other gauge invariant functional) which will be defined later, and k is an integer which guarantees that the exponent is gauge invariant. The integral is supposed to be taken over the space of all connections on a principal G-bundle on M.

The integrals (51, 52) are not defined as mathematical objects. However, one can try (as in the previous examples of the scalar Bose field and of the Yang–Mills theory) to define a combination of formal power series in k^{-1} resembling the expansion of finite-dimensional integrals studied in the previous section. In the case of the Chern–Simons theory, there is a natural requirement for such expansion: every term should be an invariant of 3-manifolds. Remarkably, such a power series exists and is more or less unique. This program was originated by Witten in [56] who outlined the basic structure of the expansion. It was developed in a number of subsequent works, in particular, in [42, 8, 9, 13, 14, 16] for the partition function for closed 3-manifolds and in [11, 32, 3] for (52), and others, when Γ is a link.

3.9.1 The Gauge Fixing

Let us use the same gauge fixing as in the Yang–Mills theory. For this we need to choose a metric on M.

Since classical solutions in the Chern–Simons theory are flat connections, the covariant derivative $d_A = d + A$ is a differential, i.e., $d_A^2 = 0$ (twisted de Rham differential) acting on $\mathfrak{g}$-valued forms on M.[10] Denote the cohomology spaces by H_A^i. Because of the Poincaré duality we have natural isomorphisms $H_A^0 \simeq H_A^3$ and $H_A^1 \simeq H_A^2$.

In a neighborhood of a classical solution A, connections can be written as $A + \alpha$ where α is a $\mathfrak{g}$-valued 1-form on M. As in the Yang–Mills theory, the *Lorentz gauge condition* for such connections is

$$d_A^* \alpha = 0.$$

We will use this gauge condition in the rest of the chapter.

3.9.2 The Faddeev–Popov Action in the Chern–Simons Theory

According to our finite-dimensional example we should add fermionic ghost fields $c(x)$ and $\bar{c}(x)$ and the Lagrange multipliers $\lambda(x)$ to the action, if we want to define Feynman diagrams in this gauge theory. However, as we argued in Sect. 3.6.4, the gauge condition can be chosen linearly near each critical point of the action, and therefore we can use the version without Lagrange multipliers. In this case we just have to add fermionic ghost fields to the action.

According to the rules of Sect. 3.6.4, the Faddeev–Popov action for the Chern–Simons theory is the sum of the classical Chern–Simons action and the ghost terms which are identical to those for the Yang–Mills theory:

---

[10] Because a principal G-bundle over any compact oriented 3d-manifold is trivializable, we choose a trivialization and identify $\Omega(M, ad(E))$ with $\Omega(M, \mathfrak{g})$.

$$\begin{aligned} CS_A(\alpha) =& CS(A) + \int_M \frac{1}{2} tr \left(\alpha \wedge d_A \alpha - \frac{2}{3} \alpha \wedge \alpha \wedge \alpha \right) \\ & \times - \frac{ih}{2} \int_M * d_A \bar{c} \wedge d_A c - \frac{ih}{2} \int_M * d_A \bar{c} \wedge [\alpha, c], \end{aligned} \tag{53}$$

where h stands for $\frac{1}{k}$. We will focus in the discussion below mostly on the special case of isolated flat connections, when $H^1_A = \{0\}$. Quite remarkably [8], the field α and the ghost fields in the Chern–Simons theory can be combined into one odd "super-field":

$$\Psi = c + \alpha + ih * d_A \bar{c}.$$

Here c, α, and $*d_A\bar{c}$ are 0, 1, and 2 forms, respectively. The action (53) can be written entirely in terms of Ψ[11]:

$$CS_A(\alpha) = CS(A) + \frac{1}{2} \int_M \mathrm{tr} \left(\Psi \wedge d_A \Psi - \frac{2}{3} \Psi \wedge \Psi \wedge \Psi \right).$$

The quadratic part of the action is the de Rham differential twisted by the flat connection A.

If $H^2_A(M, \mathfrak{g}) = \{0\}$ (equivalently, $H^1_A = \{0\}$) the gauge condition $d^*_A \alpha = 0$ together with the special form of the last term in Ψ is equivalent to $d^*_A \Psi = 0$. The inverse is also true: $d^*_A \Psi = 0$ implies $d^*_A \alpha = 0$ together with $\Psi^{(2)}$ being the Hodge dual to an exact form.

The quadratic part of this action is $(\Psi, *d_A \Psi)$ where

$$(\Phi, \Psi) = \int_M \mathrm{tr}(\Phi \wedge *\Psi).$$

The surface of the constraint $d_A \Psi = 0$ is the super-space $\Omega^0(M, \mathfrak{g}) \oplus \mathrm{Ker}\left(\left(d^*_A\right)^*_0\right) \oplus \Im((*d_A)_0) \subset \Omega(M, \mathfrak{g})[1]$ where the first and the third summands are odd and the second is even. The operator $D_A = *d_A + d_A*$ restricted to this subspace describes the quadratic part of the action. Indeed, we have

$$\int_M \mathrm{tr}(\Psi \wedge d_A \Psi) = \frac{1}{2} (\Psi, (*d_A + d_A *) \Psi).$$

The operator D_A maps even forms to even and odd form to odd, $D_A : \Omega^i \to \Omega^{2-i} \oplus \Omega^{4-i}$. It plays a prominent place in index theory [6]. It can be considered as a *Dirac operator* in a sense that

---

[11] This form of the Faddeev–Popov action for the Chern–Simons theory has a simple explanation in the framework of the Batalin–Vilkovisky formalism, see, for example, [17]. However, we will not discuss it in these notes.

$$D_A^2 = \Delta_A = d_A^* d_A + d_A d_A^*,$$

where Δ_A is *Hodge Laplace operator*. The operator D_A effectively appears in the quadratic part being restricted to odd forms. This operator will be denoted by

$$D_A^- : \begin{cases} \Omega^1 \to \Omega^1 \oplus \Omega^3 \\ \Omega^3 \to \Omega^1 \end{cases} .$$

Now the question is whether the operator D_A^- is invertible on the surface of the constraint. In other words, whether the Lorentz gauge is really a cross section through gauge orbits.

3.9.2.1 The Propagator

First, assume that the complex $(\Omega^i(M, \mathfrak{g}), d_A)$ is *acyclic*, i.e., $H^i(M, \mathfrak{g}) = \{0\}$ for all $i = 0, 1, 2, 3$ (by Poincaré duality $H^i \simeq H^{3-i}$, so it is enough to assume the vanishing of H^0 and H^1). In this case, the representation of $\pi_1(M)$ in G defined by holonomies of a flat connection A is irreducible (implied by $H^0 = \{0\}$) and isolated (implied by $H^1 = \{0\}$).

Since the spaces H^i can be naturally identified with harmonic forms and therefore with zero eigenspaces of Laplace operators, in this case all Laplace operators are invertible and so is D_A. Denote by G_A the inverse to Δ_A, i.e., the Green's function, then

$$P_A = \left(D_A^-\right)^{-1} = D_A^- G_A = G_A D_A^-.$$

Thus, in this case the quadratic part is non-degenerate and we can write contributions from Feynman diagrams as multiple integrals of the kernel of the integral operator $\left(D_A^-\right)^{-1}$. The analysis of the contributions of Feynman diagrams to the partition function was studied in this case by Axelrod and Singer in [8, 9], and by Kontsevich [42].

Another important special case arises when the flat connection is reducible but still isolated. For example, a trivial connection for rational homology spheres [13, 14, 16] has such property. In this case, we still have $H^1 = H^2 = \{0\}$ and the Lorentz gauge for α together with the exactness of $*\Psi^{(2)}$ is still equivalent to the Lorentz gauge for Ψ, i.e., $d^*\Psi = 0$. However, now there are harmonic forms in $\Omega^0(M)$ and $\Omega^3(M)$ corresponding to the fundamental class of M and because of this, D_A^- is not invertible on the space of all forms.

Nevertheless, in this case (and in a more general case when $H^1 \neq \{0\}$) one can construct an operator which is "almost inverse" to D_A^-. Such an operator is determined by the chain homotopy $K : \Omega^i \to \Omega^{i-1}$ and the Hodge decomposition of Ω. For details about such operator P see [8, 9, 13, 14, 16] and Sect. 3.3.2 of [17].

An important example of a rational homology sphere is S^3 itself. In this case, the inverse to D^- for trivial a connection can be constructed explicitly by puncturing of

S^3 at one point (the infinity). The punctured S^3 is homeomorphic to $\mathbb{R}^3$ where the fundamental class is vanishing and D^- is invertible. The restriction of $(D^-)^{-1}$ to 1-forms is the integral operator with the kernel

$$\omega(x, y) = \frac{1}{8\pi} \sum_{ijk=1}^{3} \varepsilon^{ijk} \frac{(x-y)^i dx^i \wedge dy^k}{|x-y|^3} I, \tag{54}$$

where ε^{ijk} is the totally antisymmetric tensor with $\varepsilon^{123} = 1$, and I is the identity in $\mathrm{End}(V)$. It acts on the form $\sum_i \alpha_i(x) dx^i$ as

$$\omega \circ \alpha(x) = \frac{1}{8\pi} \sum_{ijk=1}^{3} \varepsilon^{ijk} \int_{\mathbb{R}^3} \frac{(x-y)^i}{|x-y|^3} \alpha_j(x) d^3y dx^k \, . \tag{55}$$

In all cases, the propagator P_A is defined as the restriction of the restriction of the "inverse" to D_A^- (a chain homotopy, when D^- is not invertible) to 1-forms. It is an integral operator with the kernel being an element of the skew-symmetric part of $\Omega^2(M \times M, \mathfrak{g} \times \mathfrak{g})$. If e_a is an orthonormal basis in $\mathfrak{g}$ and x^i are local coordinates, we have

$$P_A(x, y) = P_A^{ab}(x, y)_{ij} e_a \otimes e_b dx^i \wedge dy^j, \quad P_A^{ab}(x, y)_{ij} = P_A^{ba}(y, x)_{ji} \, .$$

3.9.3 Vacuum Feynman Diagrams and Invariants of 3-Manifolds

3.9.3.1 Feynman Diagrams

As in other examples of quantum field theories such as the scalar Bose field and the Yang–Mills field we want to define the semiclassical expansion of the partition function and of correlation functions imitating the semiclassical expansion of finite-dimensional integrals.

Following this strategy and the computations of the Faddeev–Popov action for the Chern–Simons theory in Lorentz gauge presented above, it is natural to define the partition function $Z(M)$ (the "integral" (51)) for the Chern–Simons theory as the following combination of formal power series

$$\sum_{[A]} \exp\left(i \frac{CS_M(A)}{h} + \frac{i\pi}{4} \eta(A) \right) \left| \det{}' \left(D_A^- \right) \right|^{-1/2} \det{}' \left(\Delta_A^0 \right)$$
$$\left(1 + \sum_{n \geq 1} (ih)^n I_A^{(n)}(M, g) \right), \tag{56}$$

where h stands for k^{-1}, $\det'$ are regularized determinants of corresponding differential operators, Δ_A^0 is the Laplace–Beltrami operator acting on $C^\infty(M, \mathfrak{g})$, D_A^- is the operator D_A acting on odd forms, and $\eta(A)$ is the index of the operator D_A^-. The sum is taken over gauge classes of flat connections on M (we assume that there is a finite number of such isolated flat connections). The nth order contribution is given by the sum of Feynman diagrams

$$I_A^{(n)}(M,g) = \sum_{\Gamma, -\chi(\Gamma)=n} \frac{I_A(D(\Gamma), M, g)(-1)^{c(D(\Gamma))}}{|\operatorname{Aut}(\Gamma)|}. \tag{57}$$

In the Chern–Simons case, these are graphs with $2n$ vertices (each of them being 3-valent). The contribution $I_A(D(\Gamma), M, g)$ is an appropriate trace of the integral over M^m of the product of propagators. In other words this is the contribution from the Feynman diagram $D(\Gamma)$ with weights from Fig. 3.19.[12]

Because in this case we have only 3-valent vertices, only two first-order diagrams in Fig. 3.5 will survive. Among these two, only the "theta graph" will give a non-zero contribution due to the skew-symmetry of the propagator. The contribution from the theta graph is

$$\int_M \int_M \sum_{\{a\},\{b\}} f_{a_1a_2a_3} f_{b_1b_2b_3} P^{a_1b_1}(x,y) P^{a_2b_2}(x,y) P^{a_3b_3}(x,y) dxdy. \tag{58}$$

According to what we expect from the heuristic formula (51), the expression (56) should depend only on the homeomorphism class of M and should not depend on the choice of the metric (gauge condition). But first of all, we should make sure that every term in this series is defined. The problem is that individual integrals in the definition of $I^{(n)}$ diverge.

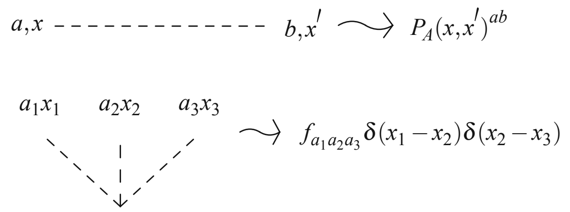


Fig. 3.19 Weights in Feynman diagrams for the Chern–Simons theory, i.e. propagators and vertices for the Ψ-field

[12] The weights in Feynman diagrams for the Chern–Simons theory are the same as we would have without the ghost fields (without the Faddeev–Popov determinant). For the Chern–Simons theory the ghost fields change the bosonic Feynman diagrams (which we would have in the naive perturbation theory) to the fermionic one (with the sign $(-1)^{c(D(\Gamma))}$). It happens because Ψ is an odd field and therefore the Feynman diagrams have fermionic nature. The orientation of graphs used in [42] is another way to encode the fermionic nature of Feynman diagrams for the Chern–Simons theory.

With the fermionic sign the sum of Feynman diagrams is finite in each order [8]. Without this sign the sum would diverge because of the singularity of the propagator at the diagonal. It is similar to the effect of ghost fields in the Yang–Mills theory. Without ghost fields the Yang–Mills theory is not renormalizable. With ghost fields, as it was shown by t'Hooft it becomes renormalizable.

As noticed in Footnote 12, a remarkable property of Feynman diagrams in the Chern–Simons case is that the sum of Feynman diagrams of any given order is finite. It is relatively easy to see that (58) is finite because of the skew-symmetry of the propagator and because it is asymptotically equivalent to (55) near the diagonal (i.e., when $x \to y$). The finiteness of the sum of Feynman diagrams in each order was proven in all orders by Axelrod and Singer [8, 9] for acyclic connections, and by Kontsevich [42] for trivial connections in rational homology spheres. This illustrates that the Chern–Simons theory is very different from the Yang–Mills theory where the renormalization procedure is necessary.

3.9.3.2 Metric Independence

Now let us focus on the metric dependence of (56). Because we expect the quantum field theory to be topological, the leading terms and each coefficient in the expansion in the powers of h should not depend on the metric. First, assume that A is an isolated irreducible flat connection.

The most singular term in the exponent is $CS_M(A)$ which is clearly metric independent. Taking into account that $\Delta = D^2$ and the natural isomorphism $\Omega^3 \simeq \Omega^0$, the absolute value of the determinant of D_A^- can be written as

$$\left|\det{}'\left(D_A^-\right)\right| = \det{}'\left(\Delta_A^1\right)^{\frac{1}{2}} \det{}'\left(\Delta_A^3\right)^{\frac{1}{2}} = \det{}'\left(\Delta_A^1\right)^{\frac{1}{2}} \det{}'\left(\Delta_A^0\right)^{\frac{1}{2}} ,$$

where Δ_A^i is the action of the Laplacian twisted by A on i-forms. Using this identity we can rearrange the determinants as

$$\left|\det{}'\left(D_A\right)\right|^{-1/2} \det{}'\left(\Delta_A^0\right) = \frac{\det{}'\left(\Delta_A^0\right)^{\frac{3}{4}}}{\det{}'\left(\Delta_A^1\right)^{\frac{1}{4}}} .$$

This expression is exactly the square root of the *Ray–Singer analytical torsion* [43], which is also the *Reidemeister torsion*, and is known to be independent of the metric.[13]

---

[13] When A is an isolated irreducible flat connection, the Ray–Singer torsion is defined as the positive number $\tau(M, A)$ such that

$$\tau(M, A) = \prod_{i \geq 1} \det{}'\left(\Delta_A^i\right)^{i(-1)^{i+1}/2} .$$

Here and in the main text $\det'(\Delta)$ is the zeta function regularization of the determinant: $\det'(\Delta) = \exp\left(-\zeta'_\Delta(0)\right)$, where

$$\zeta(s) = \mathrm{Tr}(\Delta^{-s}) = \frac{1}{\Gamma(s)} \int_0^\infty t^{s-1}\, \mathrm{Tr}(e^{t\Delta}) dt,$$

Taking into account that for Riemannian manifolds we have natural isomorphisms $\Omega^0 \simeq \Omega^3$ and $\Omega^1 \simeq \Omega^2$ we obtain

$$\tau(M, A) = \det{}'\left(\Delta_A^0\right)^{\frac{3}{2}} \det{}'\left(\Delta_A^1\right)^{-\frac{1}{2}} .$$

When the H^i are not all zeroes, $\tau(M, A)^{\frac{1}{2}}$ can be regarded as a volume element on "zero" modes, i.e., on the space $H^0 \oplus H^1$.

The exponent $\frac{i\pi}{4}\eta(A)$, which involves the index of D_A^-, can be written as

$$\frac{2\pi\eta(A)}{8} = d\frac{2\pi\eta(g, M)}{4} + c_2(G)CS(A) - \frac{2\pi}{4}I_A - \frac{d\pi(1 + b^1(M))}{4} + \frac{2\pi(\dim(H^0) + \dim(H^1))}{8} \quad (mod\ 2). \tag{59}$$

Here $\eta(g, M)$ is the *index* of the operator $D = *d + d*$ acting on odd forms on M, $d = \dim(G)$, $c_2(G)$ is the value of the *Casimir element* for $\mathfrak{g} = Lie(G)$ on the adjoint representation (also known as the dual *Coxeter number* $h^\vee$ for the appropriate normalization of the *Killing form* on $\mathfrak{g}$), and $b^1(M)$ is the first *Betti number* for M. The quantity $I_A \in \mathbb{Z}/8\mathbb{Z}$ is the *spectral flow* of the operator

$$\begin{pmatrix} *d_{A_t} & -d_{A_t}* \\ d_{A_t}* & 0 \end{pmatrix}$$

acting on $\Omega^1(M, \mathfrak{g}) \oplus \Omega^3(M, \mathfrak{g})$. Here A_t, $t \in [0, 1]$, is a path in the space of connections joining A with the trivial connection. The spectral flow I_A depends neither on the metric on M nor on the choice of the path.

The index $\eta(g, M)$ depends on the metric g on M. Recall that a *framing* of M is the homotopy class of a trivialization of the tangent bundle TM. Given a framing $f : M \to TM$ of M we can define *the gravitational Chern–Simons action*

$$I_M(g, f) = \frac{1}{4\pi}\int_M f^*Tr\left(\omega \wedge d\omega - \frac{2}{3}\omega \wedge \omega \wedge \omega\right), \tag{60}$$

where g is the metric on M, ω is the Levi–Civita connection on M, and the integrand is the pullback via f^* of the Chern–Simons form on TM.

According to the Atiyah–Patodi–Singer theorem the expression

$$\frac{1}{4}\eta(g, M) + \frac{1}{12}\frac{I_M(g, f)}{2\pi}$$

depends only on the homeomorphism class of the manifold M with the framing f, but not on the metric, and this is true for any framing f.

These arguments suggest [56, 25] that for manifolds with only irreducible and isolated flat connections, the leading term in the expression (56) should be proportional to

$$\exp\left(d\frac{i\pi}{4}\eta(g, M) + i\frac{d}{24}I_M(g, f) - \frac{di\pi}{4}\right) \sum_{[A]}\exp\left(-\frac{2\pi i I_A}{4} + i\left(\frac{1}{h} + c_2(G)\right)CS_M(A)\right)\tau(M, A)^{1/2}(1 + O(1/k)), \tag{61}$$

where $\tau(M, A)$ is the Ray–Singer torsion. This expression differs from the original guess (56) by the extra factor $\exp\left(i\frac{d}{24}I_M(g, f)\right)$.

Let us emphasize that this formula is *not a computation of the path integral*, as there is nothing to compute. It is a rearrangement and adjustment of the natural guess for the leading terms of the semiclassical expansion of the quantity to be defined. The adjustment was made on the base of the concept that the expression should not depend on the metric. Remarkably, at the end it does not depend on the metric, though it still depends on the framing.

Now let us look into higher order terms.

In the finite-dimensional case, when Feynman diagrams represent an asymptotic expansion of an existing (convergent) integral, the sum of Feynman diagrams in each order does not depend on the choice of the gauge condition simply by the nature of these coefficients.

In the *infinite-dimensional case* we are defining the integral as a sum of Feynman diagrams. Therefore, the independence of the sum of Feynman diagrams on the choice of the gauge condition (a metric on M in the case of the Lorentz gauge condition for the Chern–Simons theory) should be checked independently in each order. This was done by Axelrod and Singer in [8, 9] for acyclic connections and by Bott and Cattaneo [13, 14] for trivial connections and rational homology spheres. One of the important tools for the proof of such fact is the *graph complex* by Konstevich [42].

More preciselym the following has been proven. First write the sum of higher order contributions as

$$1+\sum_{n\geq 1}(ih)^n I_A^{(n)}(M,g)=\exp\left(\sum_{n\geq 1}(ih)^n J_A^{(n)}(M,g)\right),$$

where

$$J_A^{(n)}(M,g)=\sum_{-\chi(\Gamma)=n}^{(c)}\frac{I_A(D(\Gamma),M,g)(-1)^{c(D(\Gamma))}}{|\operatorname{Aut}(\Gamma)|}.$$

Here the sum is taken over connected graphs only. As it follows from [8, 9, 13, 14] this expression can be written as

$$J_A^{(n)}(M,g)=F_A^{(n)}(M,f)+\beta_n I_M(g,f),$$

for some $F_A^{(n)}(M,f)$ and constants β_n. Here $I(g,f)$ is the gravitational Chern–Simons action (60).

Thus, the sum of contributions of connected Feynman diagrams of fixed order, after the substraction of the gravitational Chern–Simons action with an appropriate numerical coefficient, does not depend on the metric, and, therefore, is an invariant of framed rational homology spheres (in the works of Bott and Cattaneo) or of a 3-manifold with an acyclic flat connection in a trivial principal G-bundle over it (in the works of Axelrod and Singer, and Kontsevich).

Finally, all these results can be summarized as the following proposal for the partition function of the semiclassical Chern–Simons theory. It depends on the framing and is proportional to

$$\exp\left(c(h)\left(\frac{i\pi}{4}\eta(g,M)+i\frac{1}{24}I_M(g,f)\right)\right)e^{-\frac{id\pi\left(1+b^1(M)\right)}{4}}\sum_{[A]}e^{i\left(\frac{1}{h}+c_2(G)\right)CS_M(A)}$$
$$\exp\left(-\frac{2\pi i I_A}{4}\right)\tau(M,A)^{1/2}\exp\left(\sum_{n\geq 1}(ih)^n F_A^{(n)}(M,f)\right), \quad (62)$$

where $c(h)=d+O(h)$. Witten suggested [56] the exact form of $c(h)$:

$$c(h)=\frac{d}{1+hc_2(G)}=\frac{kd}{k+h^\vee},$$

where $k=\frac{1}{h}$. This is the central charge of the corresponding *Wess–Zumino–Witten theory*.

In order to define the full TQFT from this proposition one should define the partition function in the case when flat connections are not necessarily irreducible and when they are not isolated. For most the recent progress in this direction see [17]. Also, in order to have a TQFT we should define partition functions for manifolds with boundaries. In the semiclassical framework this is an open problem.

3.9.4 Wilson Loops and Invariants of Knots

Arguing "phenomenologically" one should anticipate that expectation values of topological[14] gauge invariant observables in Chern–Simons theory, which do not require metric in their definition, should depend only on topological data, and, therefore, give some *topological invariants*.

3.9.4.1 Wilson Graphs

An example of topological observables are Wilson loops (49) or, more generally, Wilson graphs. Let us clarify the notion of the Wilson loop observable in the perturbative Chern–Simons theory. Wilson loops are defined for a collection of circles embedded into M otherwise known as a *link*. Our goal is to define the power series which would be similar to the perturbative expansion (34), as (62) is similar to the perturbative expansion (38). Most importantly, such power series should not depend on the choice of a metric on M (the choice of the gauge condition). As we have

[14] Topological observables do not require a metric in their definition.

seen above, this is possible but one should choose a framing $f : M \to TM$ of the 3-manifold.

A framed Wilson graph observable (or simply a Wilson graph) is a gauge invariant functional on connections defined as follows. Let Γ be a framed graph[15] embedded in M. Here by the framing we mean a section of the co-normal bundle $x \in \Gamma \to TM/T_x\Gamma$ for a generic point $x \in \Gamma$ which agrees on vertices.

Framing together with the orientation of M defines a cyclic ordering of edges adjacent to each vertex. It is illustrated in Fig. 3.20.

To define a Wilson graph we should make the following choices:

1. Choose a total ordering of edges adjacent to each vertex which agrees with the cyclic ordering defined by the framing.
2. Choose an orientation of each edge.
3. Choose a total ordering of vertices of Γ.
4. Choose a finite-dimensional representation V for each edge of Γ.
5. Choose a G-invariant linear mapping $\nu : \mathbb{C} \to V_1^{\varepsilon_1} \otimes \cdots \otimes V_k^{\varepsilon_k}$ for each vertex. Here numbers $1, \dots, k$ enumerate edges adjacent to the vertex, $\varepsilon_i = +$ if the edge i is incoming to the vertex, $\varepsilon_i = -$ if the edge i is outgoing from the vertex, $V_i^+ = V_i$, $V_i^- = V_i^*$, V_i is the representation assigned to the edge i, and V_i^* is its dual.

As in the case of Feynman diagrams, the ordering of vertices, and on the edges adjacent to each vertex, defines a perfect matching on endpoints of edges. Choose such total ordering.

Use the coloring of edges by finite-dimensional G-modules and the orientation of edges to define the space $V_{a_1}^{\alpha_1} \otimes V_{a_2}^{\alpha_2} \otimes V_{a_3}^{\alpha_3} \otimes V_{b_1}^{\beta_1} \otimes V_{b_1}^{\beta_2} \otimes \cdots$. Here indices $1, 2, \dots$ enumerate vertices, letters $a_i, b_i, c_i, \dots$ enumerate edges adjacent to the vertex i, and $\alpha, \beta, \dots = \pm$ indicate the orientations of edges $a, b, c, \dots$ ($+$ if the

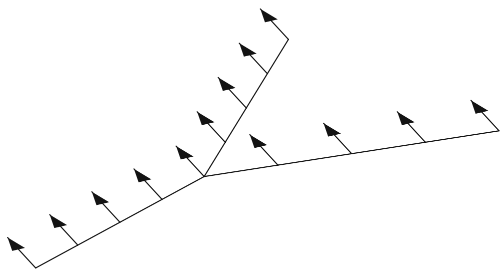


Fig. 3.20 Parallel framing at a trivalent vertex

15 The embedding $C \subset M$ induces the embedding $TC \subset TM$. Therefore a framing on M induces a framing on C, i.e., the mapping $C \to (T_C M/TC)^{\mathrm{perp}}$. A metric on M defines the splitting $T_C M = NC \oplus TC$ where NC is a normal bundle to C. A framing $f : M \to TM$ defines the framing $f_C : C \to NC$ of C by attaching a normal vector $f_C(x) \in N_x C$ for every $x \in C$.

edge is incoming, and $-$ if the edge is outgoing). The number of factors in the tensor product is equal to the number of endpoints of edges.

The coloring of vertices defines the vector

$$\nu_1 \otimes \nu_2 \otimes \cdots \in V_{a_1}^{\alpha_1} \otimes V_{a_2}^{\alpha_2} \otimes V_{a_3}^{\alpha_3} \otimes V_{b_1}^{\beta_1} \otimes V_{b_1}^{\beta_2} \otimes \cdots .$$

The holonomy $h_e(A)$ along the edge e is an element of $\mathrm{End}(V_e)$, and therefore, it is a vector in $V_e \otimes V_e^*$, where V_e is the finite-dimensional G-module assigned to the edge. The tensor product of holonomies defines a vector $\otimes_e h_e(A)$ in the space dual to $V_{a_1}^{\alpha_1} \otimes V_{a_2}^{\alpha_2} \otimes V_{a_3}^{\alpha_3} \otimes V_{b_1}^{\beta_1} \otimes V_{b_1}^{\beta_2} \otimes \cdots$.

The Wilson graph observable is the functional on the space of connections defined as

$$W_\Gamma(A) = \langle \otimes_e h_e(A), \nu_1 \otimes \nu_2 \otimes \cdots \rangle .$$

Here is an example of the Wilson graph observable for the "theta graph":

$$\sum_{i_1,i_2,i_3} \left(h_{e_1}(A)\right)_{j_1}^{i_1} (h_{e_1}(A))_{j_1}^{i_1} (h_{e_1}(A))_{j_1}^{i_1} \nu_{i_1,i_2,i_3} \nu_{j_1,j_2,j_3} .$$

The indices i_k, j_k enumerate a basis in the representation v_k assigned to the edge $k = 1, 2, 3$ and ν, μ are G-invariant vectors in the corresponding tensor products. Here we used an orthonormal basis in $\mathfrak{g}$ which explains upper and lower indices.

3.9.4.2 Feynman Diagrams for Wilson Graphs

As in the case of the partition function, define the expectation value (52) of the Wilson graph Γ as a combination of formal power series, similar to the formula (34) for the asymptotic expansion of corresponding finite-dimensional integrals.

Taking into account all we know for the partition functions of the Chern – Simons theory we arrive to the following proposal. The semiclassical ansatz for the expectation value of the Wilson graph W_Γ is

$$\sum_{[A]} \exp\left(i\frac{CS(A)}{h} + \frac{id\pi}{4}\eta(A)\right) \left|\det{}'(D_A)\right|^{-1/2} \times \det{}'((\Delta_A)_0) \left(W_\Gamma(A) + \sum_{n\geq 1} (ih)^n I_A^{(n)}(M,\Gamma) \right) . \quad (63)$$

Here we assume that all flat connections are irreducible and isolated. All quantities are the same as in (56) except

$$I_A^{(n)}(M,\Gamma) = \sum_{\Gamma',\, \chi(\Gamma)-\chi(\Gamma')=n} \frac{I_A(\Gamma',\Gamma)}{|\,\mathrm{Aut}(\Gamma)|} .$$

The Feynman diagram rules in the presence of Wilson graphs are essentially the same as for the partition function with weights given in Fig. 3.19. The difference is that now there are two types of edges, and two types of propagators (linear operators assigned to edges). As for the partition function we have dashed edges with 3-valent vertices. But now we also have solid edges, see an example in Fig. 3.21, vertices where only solid edges meet, and vertices where two solid edges (with opposite orientations) meet a dashed edge. The subgraph formed by solid edges is always Γ. The weights of vertices where only solid edges meet is given by the coloring of this edge in Γ. The weights of vertices where two solid edges meet a dashed edges and weights of solid edges are described in Fig. 3.22.

One can show [11, 32, 3, 15, 51, 13, 14] that the sum of integrals corresponding to Feynman diagrams of order n is finite for each n. Similarly to the vacuum partition function from the previous section, the semiclassical ansatz for the expectation value of the Wilson graph depends on the framing, but remarkably not on the metric. When flat connections are irreducible and isolated we arrive at the following expression:

$$\exp\left(c(h)\left(\frac{i\pi}{4}\eta(g,M)+\frac{i}{24}I_M(g,f)\right)-\frac{id\pi(1+b^1(M))}{4}\right)$$
$$\sum_{[A]}\exp\left(i\left(\frac{1}{h}+c_2(G)\right)CS_M(A)-\frac{2\pi i I_A}{4}\right)\tau(M,A)^{1/2}$$
$$\left(W_\Gamma(A)+\sum_{n\geq 1}(ih)^n J_A^{(n)}(M,\Gamma,f)\right). \tag{64}$$

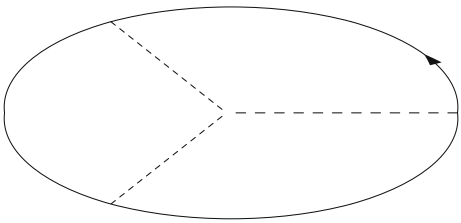


Fig. 3.21 An example of an order one graph

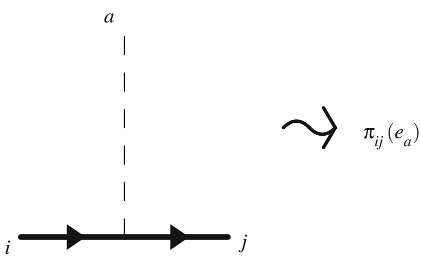


Fig. 3.22 Weights of trivalent vertices where two solid edges meet one dashed edge

Here the coefficients $J_A^{(n)}(M,\Gamma,f)$ do not depend on the metric but depend on the framing f of M. This formula defines the path integral semiclassically. Let us emphasize again that it is not a result of computation of an integral. It is a definition, modeled after the semiclassical expansion of integrals in terms of Feynman graphs. A remarkable *mathematical* fact is that every term is defined (the integrals do not diverge) and that it does not depend on the metric.

More careful analysis includes powers of h. A conjecture for counting powers of h when $H_A^0, H_A^1 \neq \{0\}$ was proposed in [25, 37, 45]. It agrees with the finite-dimensional analysis from previous sections and states that, in general, we should expect that the partition function is proportional to

$$\exp\left(d\frac{i\pi}{4}\eta(g,M)+i\frac{c(h)}{24}I_M(g,f)-\frac{d\pi i(1+b^1(M))}{4}\right)$$
$$\sum_A (2\pi(k+h^\vee))^{\frac{\dim\left(H_A^0\right)-\dim\left(H_A^1\right)}{2}}\frac{1}{Vol(G_A)}$$
$$\exp\left(i(k+h^\vee)CS_M(A)-\frac{2\pi i I_A}{4}-i\pi\frac{\dim\left(H_A^0\right)+\dim\left(H_A^1\right)}{2}\right)$$
$$\int_{M_A}\tau^{1/2}\left(W_\Gamma(A)+\sum_{n\geq 1}(ih)^n J_A^{(n)}(M,\Gamma,f)\right). \tag{65}$$

Here the sum is taken over representatives A of connected components M_A of the moduli space of flat connections in a principal G-bundle over M. The torsion τ is an element of $\otimes_i \det\left(H_A^i\right)^{\otimes(-1)^i} \simeq \left(\det\left(H_A^0\right)\otimes\det\left(H_A^1\right)^*\right)^{\otimes 2}$. The Lie algebra $\mathfrak{g}$ has an invariant scalar product and therefore $H_A^0\subset\mathfrak{g}$ has an induced volume form. Pairing this volume form with the square root of the torsion gives a volume form on H_A^1. Assuming the connected component is smooth we can integrate functions with respect to this volume form. The factor $Vol(G_A)$ is the volume of the stabilizer of the flat connection.

3.9.5 Comparison with Combinatorial Invariants

Invariants of 3-manifolds with framed graphs also can be constructed combinatorially (as a *combinatorial topological quantum field theory*). In [44] such invariants were constructed using modular categories and the representation of 3-manifolds as a surgery on S^3 or on a handlebody along a framed link. Another combinatorial construction, based on the triangulation, was developed in [52]. This construction uses a certain class of monoidal categories which are not necessarily braided.

These two constructions are related:

$$Z_M^{RT}(\mathcal{C})Z_{\overline{M}}^{RT}(\mathcal{C})=Z_M^{TV}(\mathcal{C})=Z_M^{RT}(D(\mathcal{C})).$$

Here $Z_M^{RT}(\mathcal{C})$ is the invariant obtained by the surgery [44], $Z_M^{TV}(\mathcal{C})$ is the invariant obtained by the triangulation, and the category $D(\mathcal{C})$ is the center (the double) of the category $\mathcal{C}$, see, for example, [40], and $\overline{M}$ is the manifold M with the reversed orientation.

Most interesting known examples of modular categories are quotient categories of finite-dimensional modules over quantized universal enveloping algebras at roots of unity, see [44, 2, 29]. Such categories are parametrized by pairs $(\varepsilon, \mathfrak{g})$, where $\varepsilon = \exp\left(\frac{2\pi i m}{r}\right)$ with mutually prime m and r and $\mathfrak{g}$ is a simple Lie algebra. Denote the truncated category of modules over $U_\varepsilon(\mathfrak{g})$ by $\mathcal{C}_\varepsilon(\mathfrak{g})$ (see [44, 2, 29] for details). When $m = 1$ and $r = k + c_2(\mathfrak{g})$ this category is naturally equivalent to the braiding fusion category of the WZW conformal field theory at level k, i.e., to the category of integrable modules over the affine Lie algebra $\hat{\mathfrak{g}}$ at level k with the fusion tensor product [41]. This conformal field theory is directly related to the Chern–Simons theory at level k. The arguments in favor of this are not perturbative [7]. They are based on ideas of geometric quantization.

For other values of m, the category $\mathcal{C}_\varepsilon(\mathfrak{g})$ is also equivalent to the braiding fusion category of a conformal field theory, but this conformal field theory is not directly related to the Chern–Simons theory.

The main conjecture relating the combinatorial and geometric approaches is that the *following power series are identical*:

- The asymptotic expansion of the combinatorial TQFT based on the category $\mathcal{C}_\varepsilon(\mathfrak{g})$ when $\varepsilon = \exp\left(\frac{2\pi i}{k+c_2(\mathfrak{g})}\right)$ and $k \to \infty$.
- The semiclassical expansion for the Chern–Simons path integral in terms of Feynman diagrams.

Of course this is an outline of a number of conjectures rather than a conjecture. The main reason is that the semiclassical partition functions for the Chern–Simons theory in terms of Feynman integrals are not worked out yet.

The precise statement about the correspondence between these formal power series was first outlined in [57, 25] followed by [37, 28, 45, 46, 1].

To compare these invariants one should first choose a canonical 2-framing on M [4]. The 2-framing on M is a section of $TM \times TM$. The Levi–Civita connection on TM defined by the Riemannian structure on M induces a connection on $TM \times TM$. The canonical 2-framing defines the branch of the gravitational Chern–Simons action with the property

$$d\frac{\pi}{4}\eta(g, M) + \frac{c(h)}{24} I_M(g, f) = 0.$$

One should expect that the choice of such 2-framing presumably fixes the framing in higher order corrections, though this part is still conjectural.

When the moduli space of flat connections on a principal G-bundle over M is a collection of isolated points, each corresponding to an irreducible flat connection, one should expect the following:

$$Z_M^{RT} \sim \frac{1}{|Z(G)|} \exp\left(-\frac{d\pi i}{4}\right) \sum_{[A]} \exp\left(\frac{-2i\pi I_A}{4}\right)$$
$$\exp((k + h^\vee) CS_M(A))(1 + O(1/k)),$$

where $|Z(G)|$ is the number of elements in the center of G, and Z^{RT} is the combinatorial invariant corresponding to the category $\mathcal{C}_\varepsilon(\mathfrak{g})$. When connected components of the moduli space have non-zero dimension and are smooth, the expected asymptotic behavior is

$$Z_M^{RT} \sim \exp\left(-\frac{d\pi i(1 + b^1(M))}{4}\right) \sum_{[A]} (2\pi(k + h^\vee))^{\frac{\dim\left(H_A^0\right) - \dim\left(H_A^1\right)}{2}} \frac{1}{Vol(G_A)}$$
$$\exp\left(i(k + h^\vee) CS_M(A) - \frac{2\pi i I_A}{4} - i\pi \frac{\dim\left(H_A^0\right) + \dim\left(H_A^1\right)}{2}\right)$$
$$\int_{M_A} \tau^{1/2} W_\Gamma(A)\,(1 + O(1/k))\,. \tag{66}$$

Many examples confirming this prediction were analyzed in [25, 37, 28, 45, 1].

Acknowledgements These notes are based on lectures given when the author visited Aarhus University in 2006–2007 and on lectures given during the summer school at Roskilde University's Søminestation Holbæk, Denmark, May 2008. The notes benefited from the author's discussions with many people. Special thanks to J. Andersen, L. Faddeev, V. Fock, D. Freed, T. Johnson-Freyd, D. Kazhdan, A. Losev, J. Lott, P. Mnev, and A. Schwarz. The author is also grateful to the referees for many corrections and suggestions which improved the original draft and to the corresponding editor B. Booß-Bavnbek for help and patience. Referees and editor clearly went beyond the call of duty, and the author is in their debt. Finally, the author thanks A.W. Budtz for helping with the figures.

This work was supported by the Danish National Research Foundation through the Niels Bohr initiative. The author is grateful to Aarhus University for the hospitality. The work was also supported by the NSF grant DMS-0601912 and by DARPA.

References

1. J.E. Andersen, S.K. Hansen, *Asymptotics of the quantum invariants for surgeries on the figure 8 knot*. J. Knot Theory Ramifications **15** (2006), no. 4, 479–548.
2. H.H. Andersen, *Tensor products of quantized tilting modules*. Comm. Math. Phys. **149** (1992), no. 1, 149–159.
3. D. Altschuler, L. Freidel, *Vassiliev knot invariants and Chern-Simons perturbation theory to all orders*. Comm. Math. Phys. **187** (1997), no. 2, 261–287.
4. M. Atiyah, *On framings of 3-manifolds*. Topology **29** (1990), no. 1, 1–7.
5. M. Atiyah, *Topological quantum field theories*. Inst. Hautes Etudes Sci. Publ. Math. **No. 68** (1988), 175–186.
6. M. Atiyah I.M. Singer *The index of elliptic operators, III*, Ann. Math., Second Series, **87** (3) (1968) 546–604.

7. S. Axelrod, S. Della Pietra, E. Witten, *Geometric quantization of Chern-Simons gauge theory*. J. Differential Geom. **33** (3) (1991) 787–902.
8. S. Axelrod, I.M. Singer, Chern-Simons perturbation theory. II. Perspectives in mathematical physics, 17–49, Conf. Proc. Lec. Notes Math. Phys., III, Int. Press, Cambridge (1994).
9. S. Axelrod, I.M. Singer, *Chern-Simons perturbation theory. II*. J. Differential Geom. **39** (1) (1994) 173–213.
10. S. Bates, A. Weinstein, Lectures on the Geometry of Quantization. Berkeley Mathematics Lecture Notes, 8. American Mathematical Society, Providence, RI; Berkeley Center for Pure and Applied Mathematics, Berkeley, CA, (1997).
11. D. Bar-Natan, *Perturbative Chern-Simons theory*. J. Knot Theory Ramifications **4** (1995), no. 4, 503–547.
12. C. Becci, A. Rouet, R. Stora, "Renormalizable Theories With Symmetry Breaking", in: E. Tirapegui (ed)., *Field Theory Quantization and Statistical Mechanics*, Reidel, Dordtrecht 1981.
13. R. Bott, A.S. Cattaneo, *Integral invariants of 3-manifolds*. J. Differential Geom. **48** (1998), no. 1, 91–133. 58J28
14. R. Bott, A.S. Cattaneo, *Integral invariants of 3-manifolds. II*. J. Differential Geom. **53** (1999), no. 1, 1–13.
15. R. Bott, C. Taubes, *On the self-linking of knots. Topology and physics*. J. Math. Phys. **35** (1994), no. 10, 5247–5287.
16. A.S. Cattaneo, Configuration space integrals and invariants for 3-manifolds and knots. Low-dimensional topology (Funchal, 1998), 153–165, Contemp. Math., 233, Amer. Math. Soc., Providence, RI, (1999).
17. A.S. Cattaneo P. Mnev Remarks on Chern-Simons invariants, arXiv:0811.2045.
18. D. Cimasoni, N. Reshetikhin, *Dimers on surface graphs and spin structures. II*. Comm. Math. Phys. **281** (2008), no. 2, 445–468.
19. A. Connes, D. Kreimer, *Renormalization in quantum field theory and the Riemann-Hilbert problem. I. The Hopf algebra structure of graphs and the main theorem*. Comm. Math. Phys. **210** (2000), no. 1, 249–273.
20. L. Faddeev, V. Popov, *Feynman diagrams for the Yang-Mills field*, Phys. Lett., **B25** (1967) 29.
21. L. Faddeev, A. Slavnov, Gauge Fields, Benjamin/Cummings, 1980.
22. R.P. Feynman, A.R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York (1965).
23. V. Fock, *Notes in Quantum Field Theory*, preprint.
24. D. Freed, *Classical Chern-Simons theory. I*. Adv. Math. **113** (1995), no. 2, 237–303.
25. D. Freed, R. Gompf, *Computer calculation of Witten's 3-manifold invariant*. Comm. Math. Phys. **141** (1991), no. 1, 79–117.
26. D. Freed, K. Uhlenbeck, *Instantons and Four-manifolds. Second edition*. Mathematical Sciences Research Institute Publications, 1. Springer-Verlag, New York (1991).
27. A. Gerasimov, S. Shatashvili, Two-dimensional gauge theories and quantum integrable systems. From Hodge theory to integrability and TQFT tt*-geometry, 239–262, Proc. Sympos. Pure Math., 78, Amer. Math. Soc., Providence, RI (2008).
28. S. Garoufalidis, *Relations among 3-manifold invariants*. PhD Thesis. University of Chicago, 1991.
29. S. Gelfand, D. Kazhdan, *Examples of tensor categories*. Invent. Math. **109** (1992), no. 3, 595–617.
30. J. Glimm, A. Jaffe, *Quantum Physics. A functional integral point of view*. Springer-Verlag, New York-Berlin, (1981).
31. D.J. Gross, F. Wilczek *Asymptotically free gauge theories. I*, Phys. Rev. **D8** 3633 (1973) 1343–1346.
32. E. Guadagnini, M. Martellini, M. Mintchev, *Wilson lines in Chern-Simons theory and link invariants*. Nuclear Phys. B **330** (1990), no. 2–3, 575–607.
33. G. t'Hooft, *Renormalization of massless Yang-Mills fields*. Nucl. Phys. **B33** (1971) 173–199.

34. C. Itsykson, J.-M. Druffe, *Statistical Field Theory.I., Cambridge Monographs on Mathematical Physics*, Cambridge University Press, Cambridge (1981).
35. C. Itsykson, J.-B. Zuber, *Quantum field theory*. McGraw/Hill, (1980).
36. A. Jaffe, E. Witten, *Quantum Yang-Mills theory. The millennium prize problems*, Clay Math. Inst., Cambridge, MA, (2006) pp. 129–152.
37. L. Jeffrey, *Chern-Simons-Witten invariants of lens spaces and torus bundles, and the semiclassical approximation*. Comm. Math. Phys. **147** (1992), no. 3, 563–604.
38. T. Johnson-Freyd, *The formal path integral and Schrödinger equation*. preprint, arXiv:1003.1156.
39. A. Kapustin, E. Witten, *Electric-magnetic duality and the geometric Langlands program*. Commun. Number Theory Phys. **1**(1) (2007) 1–236.
40. C. Kassel, V. Turaev, *Double construction for monoidal categories*. Acta Math. **175** (1995), no. 1, 1–48.
41. D. Kazhdan, G. Lusztig, *Affine Lie algebras and quantum groups*. Internat. Math. Res. Notices, 2, 1991 21–29.
42. M. Kontsevich, Feynman diagrams and low-dimensional topology. First European Congress of Mathematics, vol. II (Paris, 1992), 97–121, Progr. Math., 120, Birkhäuser, Basel, (1994).
43. D.B. Ray, I.M. Singer, *R-torsion and the Laplacian on Riemannian manifolds*. Advances in Math. **7**, 145–210. (1971).
44. N. Reshetikhin, V.G. Turaev, *Invariants of 3-manifolds via link polynomials and quantum groups*. Invent. Math. **103** (1991), no. 3, 547–597.
45. L. Rozansky, *A large k asymptotics of Witten's invariant of Seifert manifolds*. Comm. Math. Phys. **171** (1995), no. 2, 279–322.
46. L. Rozansky, *Reshetikhin's formula for the Jones polynomial of a link: Feynman diagrams and Milnor's linking numbers. Topology and physics*. J. Math. Phys. **35** (1994), no. 10, 5219–5246.
47. G. Segal, "The definition of conormal field theory." In: *Topology, Geometry and Quantum Field Theory, London Mathematical Society Lecture Note Series (No. 308)*, Cambridge University Press, (2002).
48. K. Symanzik, *Schrödinger representation and Casimir effect in renormalizable quantum field theory*. Nuclear Phys. B **190** (1981), no. 1, FS 3, 1–44.
49. P. Teichner S. Stoltz, *Super symmetric Euclidean field theories and generalized cohomology, a survey*. Preprint, (2008).
50. L. Takhtajan, *Quantum Mechanics for Mathematicians. Graduate Studies in Mathematics, 95*. American Mathematical Society, Providence, RI (2008).
51. D. Thurston, Integral expressions for the Vassiliev knot invariants, arXive:math. QA/9901110.
52. V.G. Turaev, O. Viro, Ya. *State sum invariants of 3-manifolds and quantum 6j-symbols*. Topology **31**(4) (1992) 865–902.
53. V.G. Turaev, *Quantum Invariants of Knots and 3-manifolds. de Gruyter Studies in Mathematics*, 18. Walter de Gruyter and Co., Berlin, (1994).
54. I.V. Tyutin, *Gauge Invariance in Field Theory and Statistical Physics in Operator Formalizam*, preprint LEBEDEV-75-39, arXiv:0812.0580[hep-th]
55. S. Weinberg, *The Quantum Theory of Fields. Vol. I- III. Supersymmetry*. Cambridge University Press, Cambridge, 2005
56. E. Witten, *Quantum field theory and the Jones polynomial*. Comm. Math. Phys. **121** (1989), no. 3, 351–399.
57. E. Witten, *Topological quantum field theory*. Comm. Math. Phys. **117** (1988), no. 3, 353–386.

Chapter 4
Mathematical Tools for Calculation of the Effective Action in Quantum Gravity

I.G. Avramidi

Abstract We review the status of covariant methods in quantum field theory and quantum gravity, in particular, some recent progress in the calculation of the effective action via the heat kernel method. We study the heat kernel associated with an elliptic second-order partial differential operator of Laplace type acting on smooth sections of a vector bundle over a Riemannian manifold without boundary. We develop a manifestly covariant method for computation of the heat kernel asymptotic expansion as well as new algebraic methods for calculation of the heat kernel for covariantly constant background, in particular, on homogeneous bundles over symmetric spaces, which enables one to compute the low-energy non-perturbative effective action.

4.1 Introduction

One of the most important problems of modern fundamental physics is the problem of reconciling classical general relativity, the theory of macroscopic gravitational phenomena, with quantum theory, so-called quantum gravity problem. This is a really difficult task since one has to answer the very basic questions concerning the local and the global structure of the spacetime itself as well as deep questions about the nature of quantum mechanics.

Although, over the last several decades many competing approaches (Euclidean path integrals, string theory, loop gravity, noncommutative geometry, asymptotic safety, various lattice approaches and others) has been put forward and despite some real progress in some of these approaches in the last two decades, we still do not have a complete consistent theory of quantum gravitational phenomena. It looks like we are missing an important piece of the puzzle which prevents us to find the solution.

I.G. Avramidi (✉)
New Mexico Institute of Mining and Technology, Socorro, NM 87801, USA, `iavramid@nmt.edu`

Avramidi, I.G.: *Mathematical Tools for Calculation of the Effective Action in Quantum Gravity*. Lect. Notes Phys. **807**, 193–259 (2010)
DOI 10.1007/978-3-642-11897-5_4

In this situation it seems to be wise to go back and to recall some pioneering works in quantum gravity. This review will concentrate on so-called *covariant methods* in quantum gravity. Some other approaches are reviewed by other lecturers of this school. The basis of the covariant methods in quantum gravity is the background field method. This method was developed mainly by De Witt in his classical papers [20, 21] and reviews [22, 23] (for the latest update see the book [24]). It is a generalization of the method of generating functionals in quantum field theory developed and successfully used by Schwinger [30, 31]. For a detailed review see, for example, [18, 16, 26].

The basic object in the background field method is the *effective action*. The effective action is a functional of the background fields that encodes, in principle, all the information of quantum field theory. It determines the full one-point propagator and the full vertex functions and, hence, the whole S-matrix. Moreover, the variation of the effective action gives the effective equations for the background fields, which makes it possible to study the back-reaction of quantum processes on the classical background. In particular, the low-energy effective action (called the effective potential) is the most appropriate tool for investigating the structure of the physical vacuum in quantum field theory.

The only practical method for the calculation of the effective action is the semi-classical perturbative expansion of the path integral in the number of loops. All fields are split in background classical parts and quantum perturbations propagating on this background and the classical action is expanded in quantum fields. Then the quadratic part determines the propagators of the quantum fields and the higher order terms reproduce the vertex functions of the perturbation theory.

In the perturbation theory the effective action is expressed in terms of the propagators and the vertex functions. One of the most powerful methods to study the propagators is the *proper time method* (also called the *heat kernel method*, in particular, by mathematicians), which was originally proposed by Fock [25] and later generalized by Schwinger [30, 31] who also applied it to the calculation of the one-loop effective action in quantum electrodynamics. It was De Witt [20, 22, 23] who perfected the proper time method; he reformulated it in the geometrical language and applied it to the case of gravitational field.

At one-loop level, the contribution of the gravitational loop is of the same order as the contributions of matter fields. At low energies (lower than the Planckian energy, $\hbar c^5/G$) the contribution of higher gravitational loops should be highly suppressed. Therefore, a semi-classical concept applies when the quantum matter fields together with the linearized perturbations of the gravitational field interact with the background gravitational field (and, probably, with the background matter fields). This is what is usually called the *one-loop quantum gravity*. The main difficulty of quantum gravity is the fact that there is no consistent way to eliminate the ultraviolet divergences arising in perturbation theory, even at one-loop level.

The present review is devoted to the development of the covariant methods for calculation of the effective action in quantum field theory and quantum gravity. The outline of the chapter is as follows. In Sect. 4.2 we review the formal structure of quantum gauge field theory and quantum gravity and the construction of the effective action following [10, 11]. In Sect. 4.3 we describe the heat kernel method and develop the asymptotic expansion of the heat kernel following [2, 3, 9–11]. In Sect. 4.4 we describe the local structure of the Green function following [8]. In Sect. 4.5 we develop a method for the calculation of the heat kernel coefficients and describe their general structure following [3, 9–11]. In Sect. 4.6 we compute the heat trace in the high-energy approximation following [1, 3, 10]. In Sect. 4.7 we describe our results for the calculation of the low-energy heat trace following our recent work [4–7, 12, 13]. In Sect. 4.8 we apply the obtained results to compute the low-energy one-loop effective action in quantum gravity.

Although it might seem to become classical and a bit old-fashioned, this field of research is pretty active even today. There are more than 100 papers on arXiv with the "heat kernel" in the title, among them 53 just since 1999. Moreover, there are about 600 papers on arXiv with the word "effective action" in the title, among which almost 400 since 1999. So, this field is far from being dead. We would like to stress that no attempts have been made to give a fully comprehensive list of references; this is beyond the intent and the scope of this chapter. We apologize in advance for not citing the work of many authors who contributed to the subject. This chapter should not be thought of as a comprehensive survey but rather an introduction aimed at non-specialists and based primarily on our own work. Once "seduced" into the field, an interested reader will easily find more recent works. Besides our own work we only cite some classical papers that laid the foundation of the field and some books and reviews that summarized the development of the subject at different stages. We did not want to expand the bibliography just for the sake of it.

4.2 Effective Action in Quantum Field Theory and Quantum Gravity

In this section we briefly describe the standard formal construction of the generating functional and the effective action in gauge theories. The basic object of any physical theory is the spacetime M, which we will assume to be a n-dimensional manifold with the topological structure of a cylinder

$$M = I \times \Sigma, \tag{1}$$

where I is an open interval of the real line (or the whole real line) and Σ is some $(n-1)$-dimensional manifold. The spacetime manifold is here assumed to be globally hyperbolic and equipped with a (pseudo)-Riemannian metric g of signature $(- + \cdots +)$; thus, a foliation of spacetime exists into spacelike sections identical to Σ. Usually one also assumes the existence of a spin structure on M. A point $x = (x^\mu)$ in the spacetime is described locally by the time x^0 and the space coordinates $(x^1, \ldots, x^{n-1})$. We label the spacetime coordinates by Greek indices, which run from 0 to $(n-1)$, and sum over-repeated indices.

Let us consider a vector bundle $\mathcal{V}$ over the spacetime M each fiber of which is isomorphic to a vector space, V, on which the spin group $\mathrm{Spin}(1, n-1)$, i.e., the covering group of Lorentz group, acts. The vector bundle $\mathcal{V}$ can also have an additional structure on which a gauge group acts. The sections of the vector bundle $\mathcal{V}$ are called fields. The tensor fields describe the particles with integer spin (bosons), while the spin-tensor fields describe particles with half-integer spin (fermions). Although the whole scheme can be developed for superfields (a combination of boson and fermion fields), we restrict ourselves in the present lecture to boson fields (which, without loss of generality, can be considered real). A field φ is represented locally by a set of real-valued functions $\varphi = (\varphi^A(x))$, where $A = 1, \ldots, \dim V$. Capital Latin indices will be used to label the local components of the fields. To construct invariant functionals we need to introduce an invariant fiber inner product and an L^2 inner product

$$(\psi, \varphi) = \int\limits_M d\mathrm{vol}\,(x)\, \psi^A(x) E_{AB}(x) \varphi^B(x), \tag{2}$$

where $d\mathrm{vol}\,(x) = dx g^{1/2}$, $g = |\det g_{\mu\nu}|$, is the natural Riemannian volume element defined by some background metric g, and E^{AB} is a nondegenerate symmetric matrix (a fiber metric). As usual, we assume that a summation over repeated indices is performed. This metric (and its inverse E^{-1AB}) can be used to naturally identify the bundle $\mathcal{V}$ with its dual $\mathcal{V}^*$ (that is to raise and lower the field indices). The sections of the dual bundle are called currents and are represented locally by a set of functions, e.g.,

$$J_A = E_{AB} \varphi^B. \tag{3}$$

We will also use the condensed DeWitt notation, where the discrete index A and the spacetime point x are combined in one lower case Latin index $i \equiv (A, x)$. Then the components of a field φ are $(\varphi^i) \equiv (\varphi^A(x))$. There is a natural pairing between the bundles $\mathcal{V}$ and $\mathcal{V}^*$ defined by

$$\langle J, \varphi \rangle \equiv J_i \varphi^i \equiv \int\limits_M d\mathrm{vol}\,(x)\, J_A(x) \varphi^A(x). \tag{4}$$

It is assumed that a summation over repeated lower case Latin indices, i.e., a combined summation – integration, is performed.

The set of all sections of the vector bundle $\mathcal{V}$ is called the configuration space, which one assumes to be an infinite-dimensional manifold $\mathcal{M}$. The fields φ^i are the coordinates on this manifold, the variational derivative $\delta/\delta\varphi$ is a tangent vector, a small disturbance $\delta\varphi$ is a one-form and so on. If $S(\varphi)$ is a scalar field on the configuration space, then its variational derivative $\delta S/\delta\varphi$ is a one-form on $\mathcal{M}$ defined by

$$\frac{d}{d\varepsilon} S(\varphi + \varepsilon h)\Big|_{\varepsilon=0} = \left\langle \frac{\delta S}{\delta \varphi}, h \right\rangle = \frac{\delta S}{\delta \varphi^i} h^i. \tag{5}$$

By using the functional differentiation one can define formally the concept of tangent space, the tangent vectors, Lie derivative, one-forms, metric, connection, geodesics, and so on (for more details, see [24]).

4.2.1 Non-gauge Field Theories

The dynamics of quantum field theory is determined by an action functional $S(\varphi)$, which is a differentiable real-valued scalar field on the configuration space. The dynamical field configurations are defined as the field configurations satisfying the stationary action principle, i.e., they must satisfy the dynamical equations of motion

$$\frac{\delta S}{\delta \varphi} = 0 \tag{6}$$

with given boundary (and initial) conditions. The set of all dynamical field configurations, i.e., those that satisfy the dynamical equations of motion, $\mathcal{M}_0$, is a subspace of the configuration space called the dynamical subspace (or the mass shell in the high-energy physics jargon).

Quantum field theory is basically a theory of small disturbances on the dynamical subspace. Most of the problems of standard quantum field theory deal with scattering processes, which are described by the transition amplitudes between some well-defined initial and final states in the remote past and the remote future. The collection of all these amplitudes is called the scattering matrix, or shortly S-matrix.

Let us single out in the spacetime two causally connected in- and out- regions, that lie in the past and in the future, respectively, relative to the region Ω, which is of interest from the dynamical standpoint. Let $|\text{in}\rangle$ and $|\text{out}\rangle$ be some initial and final states of the quantum field system in these regions. Let us consider the transition amplitude $\langle\text{out}|\text{in}\rangle$ and ask the question: how does this amplitude change under a variation of the interaction with a compact support in the region Ω. The answer to this question gives the *Schwinger variational principle* which states that

$$\delta \langle\text{out}|\text{in}\rangle = \frac{i}{\hbar} \langle\text{out}|\, \delta S \,|\text{in}\rangle, \tag{7}$$

where δS is the corresponding change of the action. This principle gives a very powerful tool to study the transition amplitudes. The Schwinger variational principle can be called the *quantization postulate*, because all the information about quantum fields can be derived from it.

Let us change the external conditions by adding a linear interaction with some external classical sources J in the dynamical region Ω, i.e.,

$$\delta S = \langle J, \varphi \rangle. \tag{8}$$

The amplitude $\langle \text{out}|\text{in}\rangle$ becomes a functional of the sources that we denote by $Z(J)$. The primary objects of interest in quantum field theory are the chronological mean values

$$\Psi_n^{i_n \dots i_1} \equiv \frac{\langle \text{out}|T(\varphi^{i_n} \cdots \varphi^{i_1})|\text{in}\rangle}{\langle \text{out}|\text{in}\rangle}, \tag{9}$$

where T denotes the operator of chronological ordering that orders the (non-commuting) operators in order of their time variables from right to left. Of course, in the presence of the sources they become functionals of J. By using the Schwinger variational principle one can obtain the chronological mean values in terms of the functional derivatives of the functional $Z(J)$, that is,

$$\begin{aligned} Z(J+\eta) = Z(J)\Bigg\{ & 1 + \left(\frac{i}{\hbar}\right) \langle \eta, \Psi_1(J)\rangle + \frac{1}{2}\left(\frac{i}{\hbar}\right)^2 \langle \eta, \Psi_2(J)\eta\rangle \\ & + \sum_{n=3}^{\infty} \frac{1}{n!}\left(\frac{i}{\hbar}\right)^n \Psi_n^{i_n \dots i_1}(J)\eta_{i_1}\cdots\eta_{i_n}\Bigg\}. \end{aligned} \tag{10}$$

In other words, the functional $Z(J)$ is the generating functional for the chronological amplitudes Ψ_n.

Let us now define another functional $W(J)$ by

$$Z = \exp\left(\frac{i}{\hbar} W\right). \tag{11}$$

Its functional derivatives define so-called full *connected Green functions*, $\mathcal{G}_n^{i_i \dots i_n}$, (or the *correlation functions*) by

$$W(J+\eta) = W(J) + \langle \eta, \mathcal{G}_1(J)\rangle + \frac{1}{2}\langle \eta, \mathcal{G}_2(J)\eta\rangle + \sum_{n=3}^{\infty} \frac{1}{n!}\mathcal{G}_n^{i_i \dots i_n}(J)\eta_{i_1}\cdots\eta_{i_n}. \tag{12}$$

The functional $\phi = \mathcal{G}_1$ is called the background (or the mean) field, and the operator $\mathcal{G} = \mathcal{G}_2$ is called the full propagator. Then, it is easy to see that all chronological mean amplitudes can be expressed in terms of connected Green functions. In particular, we have

$$\Psi_1 = \phi, \tag{13}$$

$$\Psi_2^{jk} = \phi^j \phi^k + \frac{\hbar}{i}\mathcal{G}^{jk}. \tag{14}$$

Thus, while $Z(J)$ is the generating functional for chronological amplitudes, the functional $W(J)$ is the generating functional for the connected Green functions. The Green functions satisfy the boundary conditions which are determined by the states $|\mathrm{in}\rangle$ and $|\mathrm{out}\rangle$.

The mean field itself is a functional of the sources, $\phi = \phi(J)$. It is easy to see that the functional derivative of the mean field is equal to the full propagator, that is,

$$\frac{d}{d\varepsilon}\phi(J + \varepsilon\eta)\Big|_{\varepsilon=0} = \mathcal{G}\eta. \tag{15}$$

In the non-gauge theories the full propagator $\mathcal{G}$, which plays the role of the (infinite-dimensional) Jacobian, is nondegenerate. Therefore, one can change variables and consider ϕ as independent variable and $J = J(\phi)$ (as well as all other functionals) as a functional of ϕ.

There are many different ways to show that there is a functional $\Gamma(\phi)$ such that

$$\left\langle \frac{\delta S(\varphi)}{\delta\varphi} \right\rangle = \frac{\delta\Gamma(\phi)}{\delta\phi}. \tag{16}$$

This functional is defined by

$$\langle \mathrm{out}|\mathrm{in}\rangle = \exp\left\{\frac{i}{\hbar}\left[\Gamma + \langle J, \phi\rangle\right]\right\}, \tag{17}$$

or by the functional Legendre transform

$$\Gamma(\phi) = W(J(\phi)) - \langle J(\phi), \phi\rangle. \tag{18}$$

This is the most important object in quantum field theory. It contains all the information about quantized fields. The functional expansion of this functional reads

$$\Gamma(\phi+h) = \Gamma(\phi) - \langle J(\phi), h\rangle - \frac{1}{2}\langle h, \mathcal{G}(J(\phi))h\rangle + \sum_{n=3}^{\infty}\frac{1}{n!}\Gamma_{,i_1\dots i_n}(\phi)h^{i_n}\cdots h^{i_1}. \tag{19}$$

Therefore, the first variation of Γ gives the effective equations for the background fields

$$\frac{\delta\Gamma}{\delta\phi} = -J. \tag{20}$$

These equations replace the classical equations of motion and describe the effective dynamics of the background field with regard to all quantum corrections. That is why Γ is called the *effective action*.

Furthermore, the second derivative of $\Gamma(\phi)$ determines the full propagator

$$\mathcal{G} = \left(-\frac{\delta^2\Gamma}{\delta\phi^2}\right)^{-1}.$$

The higher derivatives, $\Gamma_{,i_1\cdots i_k}$, determine the so-called full *vertex functions* (also called *strongly connected*, or *one-particle irreducible*, functions). In other words, $\Gamma(\phi)$ is the generating functional for the full vertex functions. The full vertex functions together with the full propagator determine the full-connected Green functions and, therefore, all chronological amplitudes and, hence, the S-matrix. Thus, the entire quantum field theory is summed up in the functional structure of the effective action.

One can obtain a very useful formal representation for the effective action in terms of functional integrals (also called *path integrals* or Feynman integrals). A functional integral is an integral over the (infinite-dimensional) configuration space $\mathcal{M}$. Although a rigorous mathematical definition of functional integrals is absent, they can be used in perturbation theory of quantum field theory as an effective tool, especially in gauge theories, for manipulating the whole series of perturbation theory. The point is that in perturbation theory one encounters only functional integrals of Gaussian type, which can be well defined effectively in terms of classical propagators and vertex functions. The Gaussian integrals do not depend much on the dimension and, therefore (after a proper normalization), all formulas from the finite-dimensional case, like Fourier transform, integration by parts, delta-function, change of variables etc., are valid in the infinite-dimensional case as well. One has to note that functional integrals are formally divergent—if one tries to evaluate the integrals, one encounters meaningless divergent expressions. This difficulty can be overcome in the framework of the renormalization theory (in so-called renormalizable field theories). In non-renormalizable theories (like quantum general relativity) this issue becomes the main difficulty of the theory.

Integrating the Schwinger variational principle one can obtain the following functional integral:

$$\langle \text{out}|\text{in}\rangle = \int_{\mathcal{M}} \mathcal{D}\varphi \, \exp\left\{\frac{i}{\hbar}\left[S(\varphi) + \langle J, \varphi\rangle\right]\right\}. \tag{21}$$

Here $\mathcal{D}\varphi$ represents the functional measure; however, it should not be taken too seriously—it will just provide a formal device for manipulations of Gaussian integrals. Correspondigly, for the effective action one obtains the functional equation

$$\exp\left\{\frac{i}{\hbar}\Gamma(\phi)\right\} = \int_{\mathcal{M}} \mathcal{D}\varphi \, \exp\left\{\frac{i}{\hbar}\left[S(\varphi) - \left\langle\frac{\delta\Gamma(\phi)}{\delta\phi}, (\varphi - \phi)\right\rangle\right]\right\}. \tag{22}$$

The only way to get numbers from this formal expression is to take advantage of the semi-classical approximation within a formal (asymptotic) expansion in powers of the Planck constant $\hbar$:

$$\Gamma \sim S + \sum_{k=1}^{\infty} \hbar^k \Gamma_{(k)}.$$

Next, we substitute this expansion in the functional equation for the effective action, shift the integration variable in the functional integral

$$\varphi = \phi + \sqrt{\hbar}\, h, \tag{23}$$

and expand the action $S(\varphi)$ in functional Taylor series in quantum fields h

$$\begin{aligned} S(\phi + \sqrt{\hbar}\, h) &= S(\phi) + \hbar^{1/2} \left\langle \frac{\delta S(\phi)}{\delta \phi}, h \right\rangle - \hbar \frac{1}{2} \langle h, L(\phi) h \rangle \\ &\quad + \sum_{n=3}^{\infty} \frac{1}{n!} \hbar^{n/2} S_{,i_1 \dots i_n}(\phi) h^{i_n} \cdots h^{i_1}, \end{aligned} \tag{24}$$

where L is a (usually, partial differential) operator defined by the second variation of the action

$$L = -\frac{\delta^2 S}{\delta \varphi^2}\,. \tag{25}$$

Notice that the operator L maps sections of the vector bundle $\mathcal{V}$ to sections of the dual bundle $\mathcal{V}^*$, that is,

$$L : C^\infty(\mathcal{V}) \to C^\infty(\mathcal{V}^*). \tag{26}$$

In order to have a well-defined operator, which is self-adjoint with respect to the L^2 inner product on the bundle $\mathcal{V}$, we define another operator

$$\hat{L} : C^\infty(\mathcal{V}) \to C^\infty(\mathcal{V}), \tag{27}$$

such that

$$(\varphi, \hat{L} h) = \int\limits_M d\text{vol}\ \varphi^A E_{AB} \hat{L}^B{}_C h^C = \int\limits_M d\text{vol}\ \varphi^A L_{AC} h^C = \langle \varphi, L h \rangle\,, \tag{28}$$

that is,

$$E_{AB} \hat{L}^B{}_C = L_{AC} h^C. \tag{29}$$

Now, by expanding both sides of the functional equation for the effective action in powers on $\hbar$ and equating the coefficients of equal powers of $\hbar$, we get the recurrence relations that uniquely define all coefficients $\Gamma_{(k)}$. The measure formally transforms as $\mathcal{D}\varphi = \mathcal{D}h$. All functional integrals appearing in this expansion are Gaussian and can be calculated in terms of the functional determinant, $\mathrm{Det}\,\hat{L}$, of the operator $\hat{L}$ and the bare propagator $G = L^{-1}$, i.e., the Green function of the operator L with Feynman boundary conditions. More precisely, with the proper normalization of the measure one can define

$$\int\limits_{\mathcal{M}} \mathcal{D}h\ \exp\left(-\frac{i}{2}\left(h, \hat{L}h\right)\right) = (\mathrm{Det}\,\hat{L})^{-1/2}, \tag{30}$$

$$\int\limits_{\mathcal{M}} \mathcal{D}h\ \exp\left(-\frac{i}{2}\left(h, \hat{L}h\right)\right) h^{i_1}\cdots h^{i_{2m+1}} = 0, \tag{31}$$

$$\int\limits_{\mathcal{M}} \mathcal{D}h\ \exp\left(-\frac{i}{2}\left(h, \hat{L}h\right)\right) h^{i_1}\cdots h^{i_{2m}} = \frac{(2m)!}{2^m m! i^m}(\mathrm{Det}\,\hat{L})^{-1/2} G^{(i_1 i_2}\cdots G^{i_{2m-1}i_{2m})}, \tag{32}$$

where parenthesis denote the complete symmetrization over all indices included. Of course, the Green functions of the operators L and $\hat{L}$ are related by

$$\hat{G}^A{}_B(x, y) = G^{AC}(x, y) E_{CB}(y). \tag{33}$$

In particular, the one-loop effective action is determined by the functional determinant of the operator L

$$\Gamma_{(1)} = -\frac{1}{2i}\log \mathrm{Det}\ \hat{L}, \tag{34}$$

and the two-loop effective action is given by

$$\Gamma_{(2)} = -\frac{1}{8} S_{,ijkl} G^{ij} G^{kl} - \frac{1}{12} S_{,ijk} G^{il} G^{jm} G^{kn} S_{,lmn}. \tag{35}$$

Strictly speaking, the Gaussian integrals are well defined for elliptic partial differential operators in terms of the functional determinants and their Green functions. Although the Gaussian integrals of quantum field theory are determined by hyperbolic partial differential operators with Feynman boundary conditions they can be well defined by means of the analytic continuation from the Euclidean sector of the theory where the operators become elliptic. This is done by so-called Wick rotation – one replaces the real-time coordinate by a purely imaginary one $x^0 \to i\tau$ and singles out the imaginary factor also from the action $S \to iS$ and the effective action $\Gamma \to i\Gamma$. Then the metric of the spacetime manifold becomes positive definite and the classical action in all 'nice' field theories becomes a positive-definite functional. Then the fast oscillating Gaussian functional integrals become exponentially decreasing and can be given a rigorous mathematical meaning.

4.2.2 Gauge Field Theories

Let us try to apply the formalism described above to a gauge field theory. A characteristic feature of a gauge field theory is the fact that the dynamical equations

$$\frac{\delta S}{\delta \varphi} = 0 \tag{36}$$

are not independent—there are certain identities, called Nöther identities, between them. This means that there are some nowhere vanishing vector fields

$$\mathbb{R}_\alpha = R^i{}_\alpha \frac{\delta}{\delta \varphi^i} \tag{37}$$

on the configuration space $\mathcal{M}$ that annihilate the action,

$$\mathbb{R}_\alpha S = 0, \tag{38}$$

and, hence, define invariance flows on $\mathcal{M}$. The transformations of the fields

$$\delta_\xi \varphi^i = R^i{}_\alpha \xi^\alpha \tag{39}$$

are called the invariance transformations and $\mathbb{R}_\alpha$ are called the generators of invariance transformations. The infinitesimal parameters of these transformations ξ are sections of another vector bundle (usually the tangent bundle TG of a compact Lie group G) that are respresented locally by a set of functions $(\xi^\alpha) = (\xi^a(x))$, $a = 1, \dots, \dim G$, over spacetime with compact support. To distinguish between the components of the gauge fields and the components of the gauge parameters we introduce lower case Latin indices from the beginning of the alphabet; the Greek indices from the beginning of the alphabet are used as condensed labels $\alpha = (a, x)$ that include the spacetime point.

We assume that the vector fields $\mathbb{R}_\alpha$ are linearly independent and complete, which means that they form a complete basis in the tangent space of the invariant subspace of configuration space. The vector fields $\mathbb{R}_\alpha$ form the gauge algebra. We restrict ourselves to the simplest case when the gauge algebra is the Lie algebra of an infinite-dimensional gauge Lie group $\mathcal{G}$. This is the case in Yang – Mills theory and gravity. Then the flow vectors $\mathbb{R}_\alpha$ decompose the configuration space into the invariant subspaces of $\mathcal{M}$ (called the orbits) consisting of the points connected by the gauge transformations. The space of orbits is then $\mathcal{M}/\mathcal{G}$. The linear independence of the vectors $\mathbb{R}_\alpha$ at each point implies that each orbit is a copy of the group manifold. One can show that the vector fields $\mathbb{R}_\alpha$ are tangent to the dynamical subspace $\mathcal{M}_0$, which means that the orbits do not intersect $\mathcal{M}_0$ and the invariance flow maps the dynamical subspace $\mathcal{M}_0$ into itself. Since all field configurations connected by a gauge transformation, i.e., the points on an orbit, are physically equivalent, the physical dynamical variables are the classes of gauge equivalent field configurations, i.e., the orbits. The physical configuration space is, hence, the space of orbits $\mathcal{M}/\mathcal{G}$. In other words the physical observables must be the invariants of the gauge group.

To quantize a gauge theory by means of the functional integral, we consider the in- and out- regions, define some $|\text{in}\rangle$ and $|\text{out}\rangle$ states in these regions and study the amplitude $\langle\text{out}|\text{in}\rangle$. Since all field configurations along an orbit are physically equivalent we have to integrate over the orbit space $\mathcal{M}/\mathcal{G}$. To deal with such situations one has to choose a representative field in each orbit. This can be done by choosing special coordinates $(I^A(\varphi), \chi^\alpha(\varphi))$ on the configuration space $\mathcal{M}$, where I^A label the orbits and χ^α the points in the orbit. Computing the Jacobian of the field transformation and introducing a delta functional $\delta(\chi-\zeta)$ we can fix the coordinates on the orbits and obtain the measure on the orbit space $\mathcal{M}/\mathcal{G}$

$$\mathcal{D}I = \mathcal{D}\varphi \,\mathrm{Det}\, F(\varphi)\delta(\chi(\varphi)-\zeta), \tag{40}$$

where

$$F^\beta{}_\alpha = \mathbb{R}_\alpha \chi^\beta \tag{41}$$

is a nondegenerate operator. Thus we obtain a functional integral for the transition amplitude

$$\langle\text{out}|\text{in}\rangle = \int\limits_{\mathcal{M}} \mathcal{D}\varphi \,\mathrm{Det}\, F(\varphi)\delta(\chi(\varphi)-\zeta)\exp\left\{\frac{i}{\hbar}S(\varphi)\right\}. \tag{42}$$

Now one can go further and integrate this equation over parameters ζ with a Gaussian measure determined by a symmetric nondegenerate matrix $\gamma = (\gamma_{\alpha\beta})$, which most naturally can be chosen as the metric on the orbit (gauge group metric). As a result we get

$$\langle\text{out}|\text{in}\rangle = \int\limits_{\mathcal{M}} \mathcal{D}\varphi \;(\mathrm{Det}\,\gamma)^{1/2}\;\mathrm{Det}\, F(\varphi)\exp\left\{\frac{i}{\hbar}\left[S(\varphi)+\frac{1}{2}\langle\chi(\varphi),\gamma\chi(\varphi)\rangle\right]\right\}. \tag{43}$$

The functional equation for the effective action takes the form

$$\exp\left\{\frac{i}{\hbar}\Gamma(\phi)\right\} = \int\limits_{\mathcal{M}} \mathcal{D}\varphi \;(\mathrm{Det}\,\gamma(\phi))^{1/2}\;\mathrm{Det}\, F(\varphi) \tag{44}$$
$$\times\exp\left\{\frac{i}{\hbar}\left[S(\varphi)+\frac{1}{2}\langle\chi(\varphi),\gamma(\phi)\chi(\varphi)\rangle - \left\langle\frac{\delta\Gamma(\phi)}{\delta\phi},(\varphi-\phi)\right\rangle\right]\right\}.$$

The determinants of the operators F and γ are usually represented as a result of the integration over some auxiliary Grassmannian variables, so-called ghost fields.

This equation can be used to construct the semi-classical perturbation theory in powers of the Planck constant (loop expansion), which gives the effective action in terms of the bare propagators and the vertex functions. In particular, one finds the one-loop effective action

$$\Gamma_{(1)} = -\frac{1}{2i} \log \mathrm{Det}\ \hat{L} + \frac{1}{i} \log \mathrm{Det}\, F + \frac{1}{2i} \log \mathrm{Det}\, \gamma, \tag{45}$$

where $\hat{L}$ is an operator defined by

$$\frac{d^2}{d\varepsilon^2} \left\{ S(\varphi + \varepsilon h) + \frac{1}{2} \langle \chi(\varphi + \varepsilon h), \gamma \chi(\varphi + \varepsilon h) \rangle \right\} \Big|_{\varepsilon=0} = -(h, \hat{L}h). \tag{46}$$

In DeWitt notation it reads

$$\hat{L}^k{}_j = E^{-1ki} L_{ij}, \tag{47}$$

where

$$L_{ij} = -\frac{\delta^2 S}{\delta\varphi^i \delta\varphi^j} - \frac{\delta\chi^\alpha}{\delta\varphi^i} \gamma_{\alpha\beta} \frac{\delta\chi^\beta}{\delta\varphi^j}. \tag{48}$$

4.2.3 *Quantum General Relativity*

Einstein's theory of general relativity is an example of a gauge theory with the gauge group $\mathcal{G}$ being the group of all diffeomorphisms of the spacetime manifold M and the configuration space $\mathcal{M}$ being the space of all pseudo-Riemannian metrics on M. The physical configuration space $\mathcal{M}/\mathcal{G}$ of all orbits of the gauge group is then the space of all geometries on the spacetime.

The gravitational field can be parametrized by the metric tensor of the spacetime

$$\varphi^i = g_{\mu\nu}(x), \qquad i \equiv (\mu\nu, x). \tag{49}$$

An invariant fiber metric is defined by

$$E^{\mu\nu\alpha\beta} = g^{\mu(\alpha} g^{\beta)\nu} - \varkappa g^{\mu\nu} g^{\alpha\beta}, \tag{50}$$

where $\varkappa \neq 1/n$ is a real parameter. The inverse metric is then

$$E^{-1}_{\mu\nu\alpha\beta} = g_{\mu(\alpha} g_{\beta)\nu} - \frac{\varkappa}{n\varkappa - 1} g_{\mu\nu} g_{\alpha\beta}. \tag{51}$$

The parameters of gauge transformations are the components of the vector of the infinitesimal diffeomorphism

$$\xi^\mu = \xi^\mu(x), \qquad \mu \equiv (\mu, x). \tag{52}$$

An invariant metric in the gauge group can be chosen to be just a background metric $g_{\mu\nu}$.

The local generators of the gauge transformations in this parametrization are defined by their action as follows:

$$R^i{}_\alpha \xi^\alpha = 2\nabla_{(\mu}\xi_{\nu)}, \qquad i \equiv (\mu\nu, x), \tag{53}$$

$$J_i R^i{}_\alpha = -2\nabla_\mu J^\mu{}_\alpha, \qquad \alpha \equiv (\alpha, x). \tag{54}$$

The Hilbert – Einstein action of general relativity has the form

$$S = \frac{1}{k^2}\int\limits_M dx\; g^{1/2}\,(R - 2\Lambda), \tag{55}$$

where R is the scalar curvature, $k^2 = 16\pi G$ is the Einstein coupling constant, G is the Newtonian gravitational constant, and Λ is the cosmological constant. Here we neglect the boundary term for simplicity; it will not affect our calculations.

The first variation of the action gives the classical equations of motion

$$g^{-1/2}\frac{\delta S}{\delta g_{\mu\nu}} = -\frac{1}{k^2}\left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R + \Lambda g^{\mu\nu}\right), \tag{56}$$

which satisfy, of course, the Nöther identities

$$\nabla_\mu\left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R + \Lambda g^{\mu\nu}\right) = 0. \tag{57}$$

Here $R_{\mu\nu}$ is the Ricci tensor defined in terms of the Riemann tensor by $R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu}$.

The second variation of the action defines a second-order partial differential operator by

$$g^{-1/2}\frac{\delta^2 S}{\delta g_{\mu\nu}\delta g_{\alpha\beta}}h_{\alpha\beta} = P^{\mu\nu\alpha\beta}h_{\alpha\beta}, \tag{58}$$

where

$$P^{\mu\nu,\alpha\beta} = -\frac{1}{2k^2}\Bigg\{-\left(g^{\alpha(\mu}g^{\nu)\beta} - g^{\alpha\beta}g^{\mu\nu}\right)\Delta$$

$$-g^{\mu\nu}\nabla^{(\alpha}\nabla^{\beta)} - g^{\alpha\beta}\nabla^{(\mu}\nabla^{\nu)} + 2\nabla^{(\mu}g^{\nu)(\alpha}\nabla^{\beta)}$$

$$-2R^{(\mu|\alpha|\nu)\beta} - g^{\alpha(\mu}R^{\nu)\beta} - g^{\beta(\mu}R^{\nu)\alpha} + R^{\mu\nu}g^{\alpha\beta} + R^{\alpha\beta}g^{\mu\nu}$$
$$+\left(g^{\mu(\alpha}g^{\beta)\nu} - \frac{1}{2}g^{\mu\nu}g^{\alpha\beta}\right)(R - 2\Lambda)\Bigg\}. \tag{59}$$

Here, of course, $\Delta = g^{\mu\nu}\nabla_\mu\nabla_\nu$ denotes the Laplacian.

Next, we choose the DeWitt gauge condition

$$\chi^\alpha = -E^{\alpha\beta\mu\nu}\nabla_\beta h_{\mu\nu} = -\left(g^{\alpha(\nu}\nabla^{\mu)} - \varkappa g^{\mu\nu}\nabla^\alpha\right)h_{\mu\nu}. \tag{60}$$

The ghost operator in this gauge is a second-order differential operator defined by

$$F^\mu{}_\nu = -2E^{\mu\alpha\beta}{}_\nu\nabla_\alpha\nabla_\beta = -\delta^\mu_\nu\Delta + (2\varkappa - 1)\nabla^\mu\nabla_\nu - R^\mu_\nu. \tag{61}$$

For this operator to be non-singular, the gauge parameter should satisfy the condition $\varkappa \neq 1$.

For the graviton operator L to be nondegenerate it is necessary to choose the operator γ as a zero-order differential operator defined by

$$\gamma_{\mu\nu} = \frac{\alpha}{k^2}g_{\mu\nu}, \tag{62}$$

where $\alpha \neq 0$ is a real parameter. Thus we obtain a two-parameter class of gauges involving two arbitrary parameters, $\varkappa$ and α.

The graviton operator L now reads

$$L^{\mu\nu,\alpha\beta} = \frac{1}{2k^2}\Bigg\{-\left(g^{\alpha(\mu}g^{\nu)\beta} - (1 + 2\alpha\varkappa^2)g^{\alpha\beta}g^{\mu\nu}\right)\Delta$$

$$-(1 + 2\alpha\varkappa)g^{\mu\nu}\nabla^{(\alpha}\nabla^{\beta)} - (1 + 2\alpha\varkappa)g^{\alpha\beta}\nabla^{(\mu}\nabla^{\nu)} + 2(1 + \alpha)\nabla^{(\mu}g^{\nu)(\alpha}\nabla^{\beta)}$$

$$-2R^{(\mu|\alpha|\nu)\beta} - g^{\alpha(\mu}R^{\nu)\beta} - g^{\beta(\mu}R^{\nu)\alpha)} + R^{\mu\nu}g^{\alpha\beta} + g^{\mu\nu}R^{\alpha\beta}$$
$$+\left(g^{\mu(\alpha}g^{\beta)\nu} - \frac{1}{2}g^{\mu\nu}g^{\alpha\beta}\right)(R - 2\Lambda)\Bigg\} \tag{63}$$

The most convenient choice is the so-called minimal gauge

$$\varkappa = \frac{1}{2}, \qquad \alpha = -1. \tag{64}$$

In this gauge the non-diagonal derivatives in both the graviton operator and the ghost operator vanish

$$L^{\mu\nu,\alpha\beta} = \frac{1}{2k^2}\Biggl\{\left(g^{\alpha(\mu}g^{\nu)\beta} - \frac{1}{2}g^{\alpha\beta}g^{\mu\nu}\right)(-\Delta + R - 2\Lambda)$$
$$-2R^{(\mu|\alpha|\nu)\beta} - g^{\alpha(\mu}R^{\nu)\beta} - g^{\beta(\mu}R^{\nu)\alpha)} + R^{\mu\nu}g^{\alpha\beta} + g^{\mu\nu}R^{\alpha\beta}\Biggr\}, \qquad (65)$$

$$F^{\mu}{}_{\nu} = -\delta^{\mu}{}_{\nu}\Delta - R^{\mu}{}_{\nu}. \qquad (66)$$

Finally, we define the graviton operator in the canonical Laplace-type form, $\hat{L}$, by factoring out the configuration space metric (in the minimal gauge $\varkappa = 1/2$)

$$\hat{L}_{\mu\nu}{}^{\alpha\beta} = 2k^2 E^{-1}_{\mu\nu\rho\sigma}L^{\rho\sigma\alpha\beta}\,. \qquad (67)$$

We obtain

$$\hat{L}_{\mu\nu}{}^{\alpha\beta} = -\delta^{\alpha}_{(\mu}\delta^{\beta}_{\nu)}\Delta + Q_{\mu\nu}{}^{\alpha\beta}\,, \qquad (68)$$

where

$$Q_{\mu\nu}{}^{\alpha\beta} = -2R_{\mu}{}^{(\alpha}{}_{\nu}{}^{\beta)} - 2\delta^{(\alpha}_{(\mu}R^{\beta)}_{\nu)} + R_{\mu\nu}g^{\alpha\beta} + \frac{2}{n-2}g_{\mu\nu}R^{\alpha\beta}$$
$$+\left(\delta^{\alpha}_{(\mu}\delta^{\beta}_{\nu)} - \frac{1}{(n-2)}g_{\mu\nu}g^{\alpha\beta}\right)R - 2\Lambda\delta^{\alpha}_{(\mu}\delta^{\beta}_{\nu)}\,. \qquad (69)$$

One can show that the contribution of the determinant of the operator γ can be neglected (more precisely, it can be absorbed in the measure of the path integral) since it is of zero order. Thus, with this choice of gauge parameters the one-loop effective action of quantum general relativity is given by

$$\Gamma_{(1)} = -\frac{1}{2i}\log\mathrm{Det}\,\hat{L} + \frac{1}{i}\log\mathrm{Det}\,F\,. \qquad (70)$$

Therefore, in order to compute the effective action we need to compute the determinants of Laplace-type partial differential operators acting on symmetric two tensors and vectors.

4.3 Heat Kernel Method

As we described in the previous section the effective action in quantum field theory can be computed within the semi-classical perturbation theory. It is determined by the functional determinants of second-order hyperbolic partial differential operators with Feynman boundary conditions and the higher-loop approximations are determined in terms of the Feynman propagators and the classical vertex functions. As we noted above these expressions are purely formal and need to be regularized and renormalized, which can be done in a consistent way in renormalizable field theories. One should stress, of course, that many physically interesting theories (including Einstein's general relativity) are perturbatively non-renormalizable. Since we only need Feynman propagators we can do the Wick rotation and consider instead of hyperbolic operators the elliptic ones. The Green functions of elliptic operators and their functional determinants can be expressed in terms of the heat kernel. That is why we concentrate below on the calculation of the heat kernel.

The heat kernel is one of the most powerful tools in mathematical physics and geometric analysis (see, for example, the books [27, 15, 10, 28] and reviews [19, 9, 11, 34, 14]). The short-time asymptotic expansion of the trace of the heat kernel determines the spectral asymptotics of the differential operator. The coefficients of this asymptotic expansion, called the heat invariants, are extensively used in geometric analysis, in particular, in spectral geometry and index theorems proofs [27, 15].

The gauge invariance (or covariance) in quantum gauge field theory and quantum gravity is of fundamental importance. That is why, manifestly covariant methods present inestimable advantage. A manifestly covariant calculus is such that every step is expressed in terms of geometric objects; it does not have some intermediate non-covariant steps that lead to an invariant result. Below we describe a manifestly covariant method for calculation of the heat kernel following mainly our papers [2, 3, 9–11].

4.3.1 Laplace-Type Operators

Let (M, g) be a smooth compact Riemannian manifold of dimension n without boundary, equipped with a positive-definite Riemannian metric g. We assume that it is complete, simply connected, orientable, and spin. We denote the local coordinates on M by x^μ, with Greek indices running over $1, \ldots, n$. Let $e_a{}^\mu$ be a local orthonormal frame defining a basis for the tangent space T_xM. We denote the frame indices by lower case Latin indices from the beginning of the alphabet, which also run over $1, \ldots, n$. The frame indices are raised and lowered by the metric δ_{ab}. Let $e^a{}_\mu$ be the matrix inverse to $e_a{}^\mu$, defining the dual basis in the cotangent space T^*_xM. As usual, the orthonormal frame, $e^a{}_\mu$ and $e_a{}^\mu$, will be used to transform the coordinate (Greek) indices to the orthonormal (Latin) indices. The Riemannian volume element is defined as usual by $d\text{vol} = dx g^{1/2}$, where $g = \det g_{\mu\nu} = (\det e_a{}^\mu)^2$. The spin connection $\omega^{ab}{}_\mu$ is defined in terms of the covariant derivatives of the orthonormal frame with the Levi-Civita connection. The curvature of the spin connection defines the Riemann tensor $R^a{}_{b\mu\nu}$, the Ricci tensor $R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu}$, and the scalar curvature $R = R^\mu{}_\mu$, as usual.

Let $\mathcal{T}$ be a spin-tensor bundle realizing a representation Σ of the spin group spin(n), the double covering of the group $SO(n)$, with the fiber Λ. Let Σ_{ab} be the generators of the orthogonal algebra $\mathcal{SO}(n)$, the Lie algebra of the orthogonal group $SO(n)$. The spin connection induces a connection on the bundle $\mathcal{T}$ defining the covariant derivative of smooth sections φ of the bundle $\mathcal{T}$ by

$$\nabla^{\rm spin}_\mu \varphi = \left(\partial_\mu + \frac{1}{2}\omega^{ab}{}_\mu \Sigma_{ab}\right)\varphi\,. \tag{71}$$

The commutator of covariant derivatives defines the curvature of this connection via

$$[\nabla^{\rm spin}_\mu, \nabla^{\rm spin}_\nu]\varphi = \frac{1}{2}R^{ab}{}_{\mu\nu}\Sigma_{ab}\varphi\,. \tag{72}$$

The covariant derivative along the frame vectors is defined by $\nabla_a = e_a{}^\mu \nabla_\mu$. For example, with our notation, $\nabla_a\nabla_b T_{cd} = e_a{}^\mu e_b{}^\nu e_c{}^\alpha e_d{}^\beta \nabla_\mu\nabla_\nu T_{\alpha\beta}$. The metric δ_{ab} induces a positive-definite fiber metric on tensor bundles.

Let G_{YM} be a compact Lie group (called a gauge group). It naturally defines the principal fiber bundle over the manifold M with the structure group G_{YM}. We consider a representation of the structure group G_{YM} and the associated vector bundle through this representation with the same structure group G_{YM} whose typical fiber is a k-dimensional vector space W. Then for any spin-tensor bundle $\mathcal{T}$, we define the *twisted spin-tensor bundle* $\mathcal{V}$ via the twisted product of the bundles $\mathcal{W}$ and $\mathcal{T}$. The fiber of the bundle $\mathcal{V}$ is $V = \Lambda \otimes W$ so that the sections of the bundle $\mathcal{V}$ are represented locally by k-tuples of spin-tensors.

Let $\mathcal{A}^{YM}$ be a connection one-form on the bundle $\mathcal{W}$ (called Yang – Mills or gauge connection) taking values in the Lie algebra $\mathcal{G}_{YM}$ of the gauge group G_{YM}. Then the total connection on the bundle $\mathcal{V}$ is defined by

$$\nabla_\mu\varphi = \left(\partial_\mu + \mathcal{A}_\mu\right)\varphi\,, \tag{73}$$

where

$$\mathcal{A}_\mu = \frac{1}{2}\omega^{ab}{}_\mu \Sigma_{ab}\otimes \mathbb{I}_W + \mathbb{I}_\Lambda \otimes \mathcal{A}^{YM}_\mu\,, \tag{74}$$

and the total curvature $\mathcal{R}$ of the bundle $\mathcal{V}$ is defined by

$$[\nabla_\mu, \nabla_\nu]\varphi = \mathcal{R}_{\mu\nu}\varphi\,, \tag{75}$$

where

$$\mathcal{R}_{\mu\nu} = \frac{1}{2}R^{ab}{}_{\mu\nu}\Sigma_{ab} + \mathcal{R}^{YM}_{\mu\nu} \tag{76}$$

and

$$\mathcal{R}^{YM}_{\mu\nu} = \partial_\mu \mathcal{A}^{YM}_\nu - \partial_\nu \mathcal{A}^{YM}_\mu + [\mathcal{A}^{YM}_\mu, \mathcal{A}^{YM}_\mu] \tag{77}$$

is the curvature of the Yang – Mills connection.

We also consider the bundle End$(\mathcal{V})$ of endomorphisms of the bundle $\mathcal{V}$. The covariant derivative of sections of this bundle is defined by

$$\nabla_\mu Q = \partial_\mu Q + [\mathcal{A}_\mu, Q] \tag{78}$$

and the commutator of covariant derivatives is equal to

$$[\nabla_\mu, \nabla_\nu]Q = [\mathcal{R}_{\mu\nu}, Q]\,. \tag{79}$$

We assume that the vector bundle $\mathcal{V}$ is equipped with a Hermitian metric. This naturally identifies the dual vector bundle $\mathcal{V}^*$ with $\mathcal{V}$. We assume that the connection ∇ is compatible with the Hermitian metric on the vector bundle $\mathcal{V}$. The connection is given its unique natural extension to bundles in the tensor algebra over $\mathcal{V}$ and $\mathcal{V}^*$. In fact, using the Levi-Civita connection of the metric g together with the connection on the bundle $\mathcal{V}$, we naturally obtain connections on all bundles in the tensor algebra over $\mathcal{V}$, $\mathcal{V}^*$, TM, and T^*M; the resulting connection will usually be denoted just by ∇. It is usually clear which bundle's connection is being referred to, from the nature of the section being acted upon.

We denote by $C^\infty(\mathcal{V})$ the space of smooth sections of the bundle $\mathcal{V}$. The fiber inner product on the bundle $\mathcal{V}$ defines a natural L^2 inner product and the L^2-trace Tr using the invariant Riemannian measure on the manifold M. The completion of $C^\infty(\mathcal{V})$ in this norm defines the Hilbert space $L^2(\mathcal{V})$ of square integrable sections. Let ∇^* be the formal adjoint to ∇ defined using the Riemannian metric and the Hermitian structure on $\mathcal{V}$ and let Q be a smooth Hermitian section of the endomorphism bundle End$(\mathcal{V})$.

A Laplace-type operator $L : C^\infty(V) \to C^\infty(V)$ is a partial differential operator of the form

$$L = \nabla^*\nabla + Q = -\Delta + Q\,. \tag{80}$$

In local coordinates the Laplacian is defined by

$$\Delta = g^{\mu\nu}\nabla_\mu\nabla_\nu = g^{-1/2}(\partial_\mu + \mathcal{A}_\mu)g^{1/2}g^{\mu\nu}(\partial_\nu + \mathcal{A}_\nu). \tag{81}$$

and, therefore,

$$\begin{aligned} L &= -g^{-1/2}(\partial_\mu + \mathcal{A}_\mu)g^{1/2}g^{\mu\nu}(\partial_\nu + \mathcal{A}_\nu) + Q \\ &= -g^{\mu\nu}\partial_\mu\partial_\nu - 2a^\mu\partial_\mu + q, \end{aligned} \tag{82}$$

where

$$a^\mu = g^{\mu\nu}\mathcal{A}_\nu + \frac{1}{2}g^{-1/2}\partial_\nu(g^{1/2}g^{\nu\mu}), \tag{83}$$

$$q = Q - g^{\mu\nu}\mathcal{A}_\mu\mathcal{A}_\nu - g^{-1/2}\partial_\mu(g^{1/2}g^{\mu\nu}\mathcal{A}_\nu). \tag{84}$$

Thus, a Laplace-type operator is constructed from the following three pieces of geometric data: (i) a Riemannian metric g on M, which determines the second-order part, (ii) a connection one-form $\mathcal{A}$ on the vector bundle $\mathcal{V}$, which determines the first-order part, and (iii) an endomorphism Q of the vector bundle $\mathcal{V}$, which determines the zeroth-order part. It is worth noting that every second-order differential operator with a scalar leading symbol given by the metric tensor is of Laplace type and can be put in this form by choosing the appropriate connection and the endomorphism Q.

It is easy to show that the Laplacian, Δ, and, therefore, the operator L is an elliptic symmetric partial differential operator satisfying

$$(L\varphi, \psi) = (\varphi, L\psi), \tag{85}$$

with a positive principal symbol. Moreover, the operator L is essentially self-adjoint, i.e., it has a unique self-adjoint extension. We will not be very careful about distinguishing between the operator L and its closure, and will simply say that the operator L is elliptic and self-adjoint.

It is well known [27] that

(i) the operator L has a discrete real spectrum, $\{\lambda_n\}_{n=1}^{\infty}$, bounded from below:

$$\lambda_0 < \lambda_1 < \lambda_2 < \cdots < \lambda_n < \cdots \tag{86}$$

with some real constant λ_0,

(ii) the eigenvalues grow as $k \to \infty$ as $\lambda_k \sim Ck^{2/n}$, where $n = \dim M$,

(iii) all eigenspaces of the operator L are finite dimensional, and

(iv) the eigenvectors, $\{\varphi_n\}_{n=1}^{\infty}$, of the operator L are smooth sections of the vector bundle $\mathcal{V}$ that form a complete orthonormal basis in $L^2(\mathcal{V})$.

4.3.2 Spectral Functions

The spectrum of the operator L can be described by cerain spectral invariants called spectral functions. First of all, we define the *heat trace*

$$\Theta(t) = \sum_{n=1}^{\infty} e^{-t\lambda_n}\,, \tag{87}$$

where each eigenvalue is counted with multiplicities. The heat trace is well defined for real positive t. Note that it can be analytically continued to an analytic function of t in the right half-plane (for $\operatorname{Re} t > 0$).

The heat trace determines other spectral functions by integral transforms: the *distribution function* (also called counting function), defined as the number of eigenvalues below the level λ,

$$N(\lambda) = \sum_{n=1}^{\infty} \theta(\lambda - \lambda_n) = \frac{1}{2\pi i} \int_{\varepsilon - i\infty}^{\varepsilon + i\infty} \frac{dt}{t} \, e^{t\lambda} \, \Theta(t), \tag{88}$$

where ε is a positive constant, the *density function*

$$\rho(\lambda) = \sum_{n=1}^{\infty} \delta(\lambda - \lambda_n) = \frac{1}{2\pi i} \int_{\varepsilon - i\infty}^{\varepsilon + i\infty} dt \, e^{t\lambda} \, \Theta(t) \tag{89}$$

and the *zeta-function*

$$\zeta(s, \lambda) = \sum_{n=1}^{\infty} \frac{1}{(\lambda_n - \lambda)^s} = \frac{1}{\Gamma(s)} \int_0^{\infty} dt \, t^{s-1} \, e^{t\lambda} \Theta(t), \tag{90}$$

where λ is a large negative constant such that $\mathrm{Re}\,\lambda < \lambda_0$ and s is a complex parameter with $\mathrm{Re}\, s > n/2$.

In principle, if known *exactly*, they determine the spectrum. Of course, this is not valid for asymptotic expansions of the spectral functions. There are examples of operators that have the same asymptotic series of the spectral functions but different spectrum.

One can show (see, for example, [32, 27]) that the zeta-function admits an analytic continuation to a meromorphic function of s with isolated simple poles at the points $s = [n/2]+1/2-k$ $(k = 0, 1, 2, \ldots)$ and $s = 1, 2, \ldots, [n/2]$. We will derive this property below in Sect. 4.3.5. In particular, the zeta-function is analytic at the origin. This enables one to define, in particular, the *zeta-regularized determinant* of the operator $(L - \lambda)$,

$$\zeta'(0, \lambda) \equiv \frac{\partial}{\partial s} \zeta(s, \lambda) \Big|_{s=0} = -\log \mathrm{Det}\,(L - \lambda), \tag{91}$$

which determines the one-loop effective action in quantum field theory.

4.3.3 Heat Kernel

For $t > 0$ the operators

$$U(t) = \exp(-tL) \tag{92}$$

form a semigroup of bounded operators on $L^2(\mathcal{V})$, the so-called heat semigroup. The kernel of this operator is defined by

$$U(t|x,x') = \sum_{n=1}^{\infty} e^{-t\lambda_n}\varphi_n(x)\otimes\varphi_n^*(x'), \tag{93}$$

where each eigenvalue is counted with multiplicities. It is a section of the external tensor product of vector bundles $\mathcal{V}\boxtimes\mathcal{V}^*$ over $M\times M$, which can also be regarded as an endomorphism from the fiber of $\mathcal{V}$ over x' to the fiber of $\mathcal{V}$ over x. This kernel satisfies the heat equation

$$(\partial_t + L)\,U(t) = 0 \tag{94}$$

with the initial condition

$$U(0^+|x,x') = \delta(x,x') \tag{95}$$

and is called the *heat kernel*.

Moreover, the heat semigroup $U(t)$ is a trace-class operator with a well-defined L^2-trace,

$$\mathrm{Tr}\,\exp(-tL) = \int\limits_M d\mathrm{vol}\;\;\mathrm{tr}_V U^{\mathrm{diag}}(t)\,. \tag{96}$$

Hereafter tr_V denotes the fiber trace and the label 'diag' means the diagonal value of a two-point quantity, e.g.,

$$U^{\mathrm{diag}}(t|x) = U(t|x,x')\Big|_{x=x'}\,. \tag{97}$$

It is easy to see that

$$\int\limits_M d\mathrm{vol}\;\;\mathrm{tr}_V U^{\mathrm{diag}}(t) = \sum_{n=1}^{\infty} e^{-t\lambda_n}, \tag{98}$$

which is a remarkable fact that equates the global and the local quantities. This means that the trace of the heat semigroup is equal to the heat trace defined above, that is,

$$\mathrm{Tr}\,\exp(-tL) = \Theta(t)\,. \tag{99}$$

4.3.4 Asymptotic Expansion of the Heat Kernel

In the following we are going to study the heat kernel only locally, i.e., in the neighborhood of the diagonal of $M\times M$, when the points x and x' are close to each other. The exposition will follow mainly our papers [3, 10, 9, 11]. We will keep a point x'

of the manifold fixed and consider a small geodesic ball, i.e., a small neighborhood of the point x': $B_\varepsilon(x') = \{x \in M | r(x, x') < \varepsilon\}$, $r(x, x')$ being the geodesic distance between the points x and x'. We will take the radius of the ball sufficiently small, so that each point x of the ball of this neighborhood can be connected by a unique geodesic with the point x'. This can be always done if the size of the ball is smaller than the injectivity radius of the manifold, $\varepsilon < r_{\rm inj}$.

Let $\sigma(x, x')$ be the geodetic interval, also called world function, defined as one half the square of the length of the geodesic connecting the points x and x'

$$\sigma(x, x') = \frac{1}{2} r^2(x, x'). \tag{100}$$

The first derivatives of this function with respect to x and x' define tangent vector fields to the geodesic at the points x and x', respectively, pointing in opposite directions

$$u^\mu = g^{\mu\nu} \nabla_\nu \sigma, \tag{101}$$
$$u^{\mu'} = g^{\mu'\nu'} \nabla'_{\nu'} \sigma, \tag{102}$$

and the determinant of the mixed second derivatives defines a so-called Van Vleck – Morette determinant

$$\Delta(x, x') = g^{-1/2}(x) \det\left[-\nabla_\mu \nabla'_{\nu'} \sigma(x, x')\right] g^{-1/2}(x'). \tag{103}$$

This object should not be confused with the Laplacian, which is also denoted by Δ.

Let, finally, $\mathcal{P}(x, x')$ denote the parallel transport operator of sections of the vector bundle $\mathcal{V}$ along the geodesic from the point x' to the point x. It is a section of the external tensor product of the vector bundle $\mathcal{V} \boxtimes \mathcal{V}^*$ over $M \times M$, or, in other words, it is an endomorphism from the fiber of $\mathcal{V}$ over x' to the fiber of $\mathcal{V}$ over x. Here and everywhere below the coordinate indices of the tangent space at the point x' are denoted by primed Greek letters. They are raised and lowered by the metric tensor $g_{\mu'\nu'}(x')$ at the point x'. The derivatives with respect to x' will be denoted by primed Greek indices as well.

We extend the local orthonormal frame $e_a{}^{\mu'}(x')$ at the point x' to a local orthonormal frame $e_a{}^\mu(x)$ at the point x by parallel transport. The parameters of the geodesic connecting the points x and x', namely the unit tangent vector at the point x' and the length of the geodesic (or, equivalently, the tangent vector at the point x' with the norm equal to the length of the geodesic), provide a normal coordinate system for $B_\varepsilon(x')$. Now, let us define the following geometric parameters:

$$y^a = e^a{}_\mu u^\mu = -e^a{}_{\mu'} u^{\mu'}, \tag{104}$$

so that

$$u^\mu = e_a{}^\mu y^a \qquad \text{and} \qquad u^{\mu'} = -e_a{}^{\mu'} y^a. \tag{105}$$

Note that $y^a = 0$ at $x = x'$. The geometric parameters y^a are nothing but the normal coordinates.

Near the diagonal of $M \times M$ all these two-point functions are smooth single-valued functions of the coordinates of the points x and x'. Let us note from the beginning that we will construct the heat kernel in form of covariant Taylor series in coordinates. In the smooth case these series do not necessarily converge. However, if one assumes additionally that the two-point functions are analytic, then the Taylor series converge in a sufficiently small neighborhood of the diagonal.

Further, one can easily prove that the function

$$U_0(t|x, x') = (4\pi t)^{-n/2} \Delta^{1/2}(x, x') \exp\left(-\frac{1}{2t}\sigma(x, x')\right) \mathcal{P}(x, x') \tag{106}$$

satisfies the initial condition

$$U_0(0^+|x, x') = \delta(x, x'). \tag{107}$$

Moreover, locally it also satisfies the heat equation in the free case, when the Riemannian curvature of the manifold Riem, the curvature of the bundle connection $\mathcal{R}$, and the endomorphism Q vanish:

$$\text{Riem} = \mathcal{R} = Q = 0\,. \tag{108}$$

Therefore, $U_0(t|x, x')$ is the exact heat kernel for a pure Laplacian in flat Euclidean space with a flat trivial bundle connection and without the endomorphism Q.

4.3.4.1 Transport Function

This function gives a good framework for the approximate solution in the general case. Namely, by factorizing out this free factor we get an ansatz

$$U(t|x, x') = (4\pi t)^{-n/2} \Delta^{1/2}(x, x') \exp\left(-\frac{1}{2t}\sigma(x, x')\right) \mathcal{P}(x, x')\Omega(t|x, x'). \tag{109}$$

The function $\Omega(t|x, x')$, called the *transport function*, is a section of the endomorphism vector bundle End (V) over the point x'. Using the definition of the functions $\sigma(x, x')$, $\Delta(x, x')$, and $\mathcal{P}(x, x')$ it is not difficult to find that the transport function satisfies a transport equation

$$\left(\partial_t + \frac{1}{t}D + \tilde{L}\right)\Omega(t) = 0, \tag{110}$$

where D is the radial vector field, i.e., operator of differentiation along the geodesic defined by

$$D = u^\mu \nabla_\mu, \tag{111}$$

and $\tilde{L}$ is a second-order differential operator defined by

$$\tilde{L} = \mathcal{P}^{-1}\Delta^{-1/2} L \Delta^{1/2}\mathcal{P}. \tag{112}$$

The initial condition for the transport function is obviously

$$\Omega(t|x, x') = \mathbb{I}_V, \tag{113}$$

where $\mathbb{I}_V$ is the identity endomorphism of the vector bundle $\mathcal{V}$ over x'.

It is obvious that if we replace the operator L by $(L-\lambda)$, with $\mathrm{Re}\,\lambda < \lambda_0$, then the heat kernel and the transport function are simply multiplied by $e^{t\lambda}$, i.e., the transport function for the operator $(L-\lambda)$ is $e^{t\lambda}\Omega(t)$. Further, for $\lambda < \lambda_0$ the operator $(L-\lambda)$ becomes a positive operator. Therefore, the function $e^{t\lambda}\Omega(t)$ satisfies the following asymptotic conditions:

$$\lim_{t\to\infty,0} t^\alpha \partial_t^N \left[e^{t\lambda}\Omega(t)\right] = 0 \qquad \text{for } \lambda < \lambda_1,\ \alpha > 0,\ N \geq 0. \tag{114}$$

In other words, as $t \to \infty$ the function $e^{t\lambda}\Omega(t)$ and all its derivatives decreases faster than any power of t, actually it decreases exponentially, and as $t \to 0$ the product of $e^{t\lambda}\Omega(t)$ with any positive power of t vanishes.

Hereafter we fix $\lambda < \lambda_0$, so that $(L - \lambda)$ is a positive operator. Now, let us consider a slightly modified version of the Mellin transform of the function $e^{t\lambda}\Omega(t)$ introduced in [3]

$$b_q(\lambda) = \frac{1}{\Gamma(-q)} \int_0^\infty dt\ t^{-q-1} e^{t\lambda}\Omega(t). \tag{115}$$

Note that for fixed λ this is a Mellin transform of $e^{t\lambda}\Omega(t)$ and for a fixed q this is a Laplace transform of the function $t^{-q-1}\Omega(t)$. The integral (115) converges for $\mathrm{Re} q < 0$. By integrating by parts N times and using the asymptotic conditions (114) we also get

$$b_q(\lambda) = \frac{1}{\Gamma(-q+N)} \int_0^\infty dt\ t^{-q-1+N} (-\partial_t)^N \left[e^{t\lambda}\Omega(t)\right]. \tag{116}$$

This integral converges for $\mathrm{Re}\, q < N - 1$. Using this representation one can prove that [3] the function $b_q(\lambda)$ is an entire function of q (analytic everywhere) satisfying the asymptotic condition:

$$\lim_{|q|\to\infty,\ \mathrm{Re}\, q<N} \Gamma(-q+N) b_q(\lambda) = 0, \qquad \text{for any } N > 0. \tag{117}$$

Moreover, the values of the function $b_q(\lambda)$ at the integer positive points $q = k$ are given by

$$b_k(\lambda) = (-\partial_t)^k \left[e^{t\lambda}\Omega(t)\right]\Big|_{t=0} = \sum_{n=0}^{k} \binom{k}{n} a_n, \tag{118}$$

where

$$a_k = (-\partial_t)^k \Omega(t)\Big|_{t=0}. \tag{119}$$

By inverting the Mellin transform we obtain a new ansatz for the transport function and, hence, for the heat kernel

$$\Omega(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} dq\; e^{-t\lambda} t^q\, \Gamma(-q) b_q(\lambda), \tag{120}$$

where $c < 0$ and $\operatorname{Re}\lambda < \lambda_0$. Clearly, since the left-hand side of this equation does not depend on λ, neither does the right-hand side. Thus, λ serves as an auxiliary parameter that regularizes the behavior at $t \to \infty$. If we invert instead the Laplace transform, we obtain another representation

$$\Omega(t) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} d\lambda\; e^{-t\lambda} t^{q+1} \Gamma(-q) b_q(\lambda), \tag{121}$$

where $\gamma < \lambda_0$ and $\operatorname{Re} q < 0$.

Substituting this ansatz into the transport equation we get a functional equation for the function b_q

$$\left(1 + \frac{1}{q} D\right) b_q(\lambda) = (\tilde{L} - \lambda)\, b_{q-1}(\lambda). \tag{122}$$

The initial condition for the transport function is translated into

$$b_0(\lambda) = \mathbb{I}_V. \tag{123}$$

Thus, we have reduced the problem of solving the heat equation to the following problem: one has to find an entire function of q, $b_q(\lambda|x, x')$ that satisfies the functional equation (122) with the initial condition (123) and the asymptotic condition (117).

Although the variables q and λ seem to be independent they are very closely related to each other. In particular, by differentiating with respect to λ we obtain an important result

$$\frac{\partial}{\partial\lambda} b_q(\lambda) = -q b_{q-1}(\lambda). \tag{124}$$

Also, by differentiating (122) with respect to q one obtains another recursion

$$\left(1 + \frac{1}{q} D\right) b_q'(\lambda) = \tilde{L}\, b_{q-1}'(\lambda) + \frac{1}{q^2} D b_q(\lambda), \tag{125}$$

where

$$b_q'(\lambda) = \frac{\partial}{\partial q} b_q(\lambda), \tag{126}$$

which enables one to compute the derivatives of the function $b_q(\lambda)$ at positive integer points, if one fixes its value $b_0'(\lambda)$. This turns out to be useful when computing the determinant of the operator $(L - \lambda)$.

Moreover, one can actually manifest the dependence of $b_q(\lambda)$ on λ. It is not difficult to prove that [3] the integral

$$b_q(\lambda) = \frac{1}{2\pi i} \int\limits_{c_1 - i\infty}^{c_1 + i\infty} dp\, \frac{\Gamma(-p)\Gamma(p-q)}{\Gamma(-q)} (-\lambda)^{q-p} a_p, \tag{127}$$

with $\mathrm{Re}\, q < c_1 < 0$, satisfies (122) if a_p satisfies this equation for $\lambda = 0$, i.e.,

$$\left(1 + \frac{1}{q} D\right) a_q = \tilde{L}\, a_{q-1} \tag{128}$$

with the initial condition

$$a_0 = \mathbb{I}_V. \tag{129}$$

For integer $q = k = 1, 2, \ldots$ the functional equation (128) becomes a recursion system that, together with the initial condition (129), determines all coefficients a_k.

Now, from (127) we also obtain the asymptotic expansion of $b_q(\lambda)$ as $\lambda \to -\infty$

$$b_q(\lambda) \sim \sum_{n=0}^{\infty} \frac{\Gamma(q+1)}{n!\,\Gamma(q-n+1)} (-\lambda)^{q-n} a_n. \tag{130}$$

For integer q this coincides with (118).

The function $b_q(\lambda)$ turns out to be extremely useful in computing the heat kernel, the resolvent kernel, the zeta-function, and the determinant of the operator L. It contains the same information about the operator L as the heat kernel. In some cases the function $b_q(\lambda)$ can be constructed just by analytical continuation from the integer positive values b_k [3].

4.3.4.2 Asymptotic Expansion of the Transport Function

Now we are going to do the usual trick, namely, to move the contour of integration over q to the right. Due to the presence of the gamma-function $\Gamma(-q)$ the integrand has simple poles at the non-negative integer points $q = 0, 1, 2 \ldots$, which contribute to the integral while moving the contour. So, we get

$$\Omega(t) = e^{-t\lambda} \left\{ \sum_{k=0}^{N-1} \frac{(-t)^k}{k!} b_k(\lambda) + R_N(t) \right\}, \tag{131}$$

where

$$R_N(t) = \frac{1}{2\pi i} \int_{c_N - i\infty}^{c_N + i\infty} dq\, t^q\, \Gamma(-q) b_q(\lambda) \tag{132}$$

with c_N a constant satisfying the condition $N - 1 < c_N < N$. As $t \to 0$ the rest term $R_N(t)$ behaves like $O(t^N)$, so we obtain an asymptotic expansion as $t \to 0$

$$\Omega(t) \sim e^{-t\lambda} \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} b_k(\lambda) = \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} a_k. \tag{133}$$

Using our ansatz (109) we find immediately the heat trace

$$\Theta(t) = (4\pi t)^{-n/2} e^{-t\lambda} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} dq\, t^q\, \Gamma(-q) B_q(\lambda), \tag{134}$$

where

$$B_q(\lambda) = \int_M d\mathrm{vol}\ \mathrm{tr}_V\, b_q^{\mathrm{diag}}(\lambda). \tag{135}$$

The heat trace has an analogous asymptotic expansion as $t \to 0$

$$\Theta(t) \sim (4\pi t)^{-n/2} e^{-t\lambda} \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} B_k(\lambda) = \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} A_k\,, \tag{136}$$

where

$$A_k = \int_M d\mathrm{vol}\ \mathrm{tr}_V\, a_k^{\mathrm{diag}}. \tag{137}$$

This is the famous Minakshisundaram – Pleijel asymptotic expansion [29]. The physicists call it the Schwinger – De Witt expansion [14]. Its coefficients A_k are also called sometimes Hadamard – Minakshisundaram – De Witt – Seeley (HMDS) coefficients. This expansion is of great importance in differential geometry, spectral geometry, quantum field theory, and other areas of mathematical physics, such as theory of Huygens' principle, heat kernel proofs of the index theorems, Korteveg – De Vries hierarchy, Brownian motion.

One should stress, however, that this series does not converge, in general. In that sense our ansatz (120) or (131) in form of a Mellin transform of an entire function is much better since it is exact and gives an explicit formula for the rest term. Of course, this pushes the problem of evaluating the heat trace into the problem of finding the function $B_q(\lambda)$ [or the local function $b_q(\lambda)$]. The beauty of the function B_q is that for positive integer values $q = k$ it is equal to the standard locally computable heat kernel coefficients $B_k(\lambda)$, and, in some cases, $B_q(\lambda)$ can be evaluated by just analytically continuing the heat kernel coefficients $B_k(\lambda)$ to the whole complex plane of q (see, for example, [1, 3]). This will be illustrated briefly in Sect. 4.6 below.

4.3.5 Zeta-Function and Determinant

Let us apply our ansatz for computation of the complex power of the operator $(L-\lambda)$ (with $\lambda < \lambda_0$ so that the operator $(L - \lambda)$ is positive) defined by

$$G_s(\lambda) = (L - \lambda)^{-s} = \frac{1}{\Gamma(s)} \int_0^\infty dt\, t^{s-1}\, e^{t\lambda}\, U(t). \tag{138}$$

Using our ansatz for the heat kernel one can obtain [3]

$$G_s(\lambda) = (4\pi)^{-n/2} \Delta^{1/2} \mathcal{P} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} dq\, \frac{\Gamma(-q)\Gamma(-q-s+n/2)}{\Gamma(s)} \left(\frac{\sigma}{2}\right)^{q+s-n/2} b_q(\lambda), \tag{139}$$

where $c < -\mathrm{Re}\, p + n/2$.

Outside the diagonal, i.e., for $\sigma \neq 0$, this integral converges for any s and defines an entire function of s. The integrand in this formula is a meromorphic function of q with some simple and maybe some double poles. If we move the contour of integration to the right, we get contributions from the simple poles in form of powers of σ and a logarithmic part due to the double poles (if any). This gives the complete structure of diagonal singularities of $G_s(x,x')$. Thus the function $b_q(\lambda)$ turns out to be very useful to study the diagonal singularities.

Now, let us consider the diagonal limit of G_s. By taking the limit $\sigma \to 0$ we obtain a very simple formula in terms of the function b_q

$$G_s^{\rm diag}(\lambda) = (4\pi)^{-n/2}\frac{\Gamma(s-n/2)}{\Gamma(s)} b^{\rm diag}_{n/2-s}(\lambda). \tag{140}$$

This gives automatically the zeta-function

$$\zeta(s,\lambda) = (4\pi)^{-n/2}\frac{\Gamma(s-n/2)}{\Gamma(s)} B_{n/2-s}(\lambda). \tag{141}$$

This formula expresses the zeta-function in terms of the function $B_q(\lambda)$. The main advantage of this formula is that the function $B_q(\lambda)$ is an *entire* function of q. Because the analytical properties of the gamma-function are well known (it is a meromorphic function with simple isolated poles at non-positive integers), this enables one to immediately see the analytical structure of the zeta-function. We see that both $G_s^{\rm diag}(\lambda)$ and $\zeta(s,\lambda)$ are meromorphic functions of s with simple poles at the points $s = [n/2] + 1/2 - k$, $(k = 0, 1, 2, \dots)$ and $s = 1, 2, \dots, [n/2]$. In particular, the zeta-function is analytic at the origin. Its value at the origin is given by

$$\zeta(0,\lambda) = \begin{cases} 0 & \text{for odd } n\,, \\ (4\pi)^{-n/2}\dfrac{(-1)^{n/2}}{\Gamma(n/2+1)} B_{n/2}(\lambda) & \text{for even } n\,. \end{cases} \tag{142}$$

This gives the regularized number of all modes of the operator L.

Moreover, the derivative of the zeta-function at the origin is also well defined. As we already mentioned above it determines the regularized determinant of the operator $(L-\lambda)$

$$\log \mathrm{Det}\,(L-\lambda) = -(4\pi)^{-n/2}\frac{\pi(-1)^{(n+1)/2}}{\Gamma(n/2+1)} B_{n/2}(\lambda) \tag{143}$$

for odd n, and

$$\log \mathrm{Det}\,(L-\lambda) = (4\pi)^{-n/2}\frac{(-1)^{n/2}}{\Gamma(n/2+1)} \left\{ B'_{n/2}(\lambda) - [\Psi(n/2+1) + \mathbf{C}] B_{n/2}(\lambda) \right\} \tag{144}$$

for even n. Here $\Psi(z) = (d/dz)\log\Gamma(z)$ is the psi-function, $\mathbf{C} = -\Psi(1) = 0.577\ldots$ is the Euler constant, and

$$B'_{n/2}(\lambda) = \frac{\partial}{\partial q} B_q(\lambda)\bigg|_{q=n/2} . \tag{145}$$

4.4 Green Function

In this section we closely follow our paper [8]. Let λ be a sufficiently large negative parameter, such that $\lambda < \lambda_0$ and, therefore, $(L-\lambda)$ be a positive operator. The Green function of the operator $(L-\lambda)$ reads

$$G(\lambda|x,x') = \sum_{n=1}^{\infty} \frac{1}{\lambda_n - \lambda} \varphi_n(x) \otimes \varphi_n^*(x') . \tag{146}$$

It is not difficult to see that the Green function can be represented as the Laplace transform of the heat kernel

$$G(\lambda) = \int_0^\infty dt\ e^{t\lambda} U(t) . \tag{147}$$

Using our ansatz for the heat kernel we obtain

$$G(\lambda) = (4\pi)^{-n/2} \Delta^{1/2} \mathcal{P} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} dq\ \Gamma(-q)\Gamma(-q-1+n/2) \left(\frac{\sigma}{2}\right)^{q+1-n/2} b_q(\lambda), \tag{148}$$

where $c < n/2 - 1$.

This ansatz is especially useful for studying the singularities of the Green function, or, more generally, for constructing the Green function as a power series in σ. The integrand in (148) is a meromorphic function with poles at the points $q = k$ and $q = k - 1 + n/2$, where $(k = 0, 1, 2, \ldots)$. Here one has to distinguish between odd and even dimensions. In odd dimensions, the poles are at the points $q = k$ and $q = k + [n/2] - 1/2$ and are simple, whereas in even dimension there are simple poles at $q = 0, 1, 2, \ldots, n/2 - 2$ and double poles at the points $q = k + n/2 - 1$.

Moving the contour of integration in (148) to the right one can obtain an expansion of the Green function in powers of σ (Hadamard series). Generally, we obtain

$$G(\lambda) = G^{\text{sing}}(\lambda) + G^{\text{non-anal}}(\lambda) + G^{\text{reg}}(\lambda) . \tag{149}$$

Here $G^{\text{sing}}(\lambda)$ is the singular part which is polynomial in the inverse powers of $\sqrt{\sigma}$

$$G^{\mathrm{sing}}(\lambda) = (4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\sum_{k=0}^{[(n+1)/2]-2}\frac{(-1)^k}{k!}\Gamma(n/2-k-1)\left(\frac{2}{\sigma}\right)^{n/2-k-1}b_k(\lambda). \tag{150}$$

Let us fix an integer N such that $N > (n-1)/2$.

For the rest we get in *odd* dimensions

$$\begin{aligned}
&G^{\text{non-anal}}(\lambda) + G^{\mathrm{reg}}(\lambda)\\
=&(-1)^{(n-1)/2}(4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\sum_{k=0}^{N-(n+1)/2}\frac{\pi}{\Gamma\left(k+\frac{n+1}{2}\right)\Gamma\left(k+\frac{3}{2}\right)}\left(\frac{\sigma}{2}\right)^{k+1/2}b_{k+\frac{n-1}{2}}(\lambda)\\
&+(-1)^{(n+1)/2}(4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\sum_{k=0}^{N-(n+1)/2}\frac{\pi}{k!\Gamma(k+n/2)}\left(\frac{\sigma}{2}\right)^k b_{k-1+n/2}(\lambda)\\
&+(4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\frac{1}{2\pi i}\int\limits_{c_N-i\infty}^{c_N+i\infty}dq\,\left(\frac{\sigma}{2}\right)^{q+1-n/2}\Gamma(-q)\Gamma(-q-1+n/2)b_q(\lambda),
\end{aligned} \tag{151}$$

where $N-1 < c_N < N-1/2$. Thus, by putting $N\to\infty$ we recover the Hadamard power series in σ for *odd* dimension n

$$\begin{aligned}
G^{\text{non-anal}}(\lambda) \sim{}& (-1)^{(n-1)/2}(4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\\
&\sum_{k=0}^{\infty}\frac{\pi}{\Gamma\left(k+\frac{n+1}{2}\right)\Gamma\left(k+\frac{3}{2}\right)}\left(\frac{\sigma}{2}\right)^{k+1/2}b_{k+\frac{n-1}{2}}(\lambda)
\end{aligned} \tag{152}$$

$$\begin{aligned}
G^{\mathrm{reg}}(\lambda) \sim{}& (-1)^{(n+1)/2}(4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\\
&\sum_{k=0}^{\infty}\frac{\pi}{k!\Gamma(k+n/2)}\left(\frac{\sigma}{2}\right)^k b_{k-1+n/2}(\lambda).
\end{aligned} \tag{153}$$

In *even* dimensions, the point is more subtle due to the presence of double poles. Moving the contour in (148) to the right and calculating the contribution of the residues at the simple and double poles we obtain

$$
\begin{aligned}
&G^{\text{non-anal}}(\lambda) + G^{\text{reg}}(\lambda)\\
=\ & (-1)^{n/2-1}(4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\log\left(\frac{\mu^2\sigma}{2}\right)\sum_{k=0}^{N-1}\frac{1}{k!\Gamma(k+n/2)}\left(\frac{\sigma}{2}\right)^k b_{k-1+n/2}(\lambda)\\
&+(-1)^{n/2-1}(4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\sum_{k=0}^{N-1}\frac{1}{k!\Gamma(k+n/2)}\left(\frac{\sigma}{2}\right)^k\\
&\times\left\{b'_{k-1+n/2}(\lambda) - \left[\log\mu^2 + \Psi(k+1) + \Psi(k+n/2)\right]b_{k-1+\frac{n}{2}}(\lambda)\right\}\\
&+(4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\frac{1}{2\pi i}\int\limits_{c_N-i\infty}^{c_N+i\infty} dq\,\left(\frac{\sigma}{2}\right)^{q+1-n/2}\Gamma(-q)\Gamma(-q-1+n/2)b_q(\lambda),
\end{aligned}
\tag{154}
$$

where μ is an arbitrary mass parameter introduced to preserve dimensions, $N-1 < c_N < N$ and $\Psi(z) = (d/dz)\log\Gamma(z)$. If we let $N \to \infty$ we obtain the Hadamard expansion of the Green function for *even* dimension $n \geq 2$

$$
\begin{aligned}
&G^{\text{non-anal}}(\lambda) \sim (-1)^{n/2-1}(4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\log\left(\frac{\mu^2\sigma}{2}\right)\\
&\sum_{k=0}^{\infty}\frac{1}{k!\Gamma(k+n/2)}\left(\frac{\sigma}{2}\right)^k b_{k-1+n/2}(\lambda)
\end{aligned}
\tag{155}
$$

$$
\begin{aligned}
&G^{\text{reg}}(\lambda) \sim (-1)^{n/2-1}(4\pi)^{-n/2}\Delta^{1/2}\mathcal{P}\sum_{k=0}^{\infty}\frac{1}{k!\Gamma(k+n/2)}\left(\frac{\sigma}{2}\right)^k\\
&\times\left\{b'_{k-1+n/2}(\lambda) - \left[\log\mu^2 + \Psi(k+1) + \Psi(k+n/2)\right]b_{k-1+n/2}(\lambda)\right\}
\end{aligned}
\tag{156}
$$

Note that the singular part (which is a polynomial in inverse powers of $\sqrt{\sigma}$) and the *non-analytical* parts (proportional to $\sqrt{\sigma}$ and $\log\sigma$) are expressed in terms of the values of the function $b_q(\lambda)$ at the integer points q, which are uniquely locally computable from the recursion relation, whereas the regular analytical part contains the values of the function $b_q(\lambda)$ at half-integer positive points q and the derivatives of the function $b_q(\lambda)$ with respect to q at integer positive points q, which are *not* expressible in terms of the local information. These objects are global and cannot be expressed further in terms of the local heat kernel coefficients. However, they can be computed from (122) and (125) in terms of the value of the function $b(q)$ at some fixed point q_0 (see [3]).

The regular part of the Green function has a well-defined diagonal value and the functional trace. It reads in odd dimensions n as follows:

$$\begin{aligned}\mathrm{Tr}\ G^{\mathrm{reg}}(\lambda) &= (-1)^{(n+1)/2}(4\pi)^{-n/2}\frac{\pi}{\Gamma(n/2)}B_{n/2-1}(\lambda)\\ &\overset{\lambda\to-\infty}{\sim} (-1)^{(n+1)/2}(4\pi)^{-n/2}\pi\sum_{k=0}^{\infty}\frac{(-\lambda)^{n/2-1-k}}{k!\Gamma(n/2-k)}A_k \qquad (157)\end{aligned}$$

and in even dimensions n

$$\begin{aligned}&\mathrm{Tr}\ G^{\mathrm{reg}}(\lambda) = (-1)^{n/2-1}\frac{(4\pi)^{-n/2}}{\Gamma(n/2)}\\ &\left\{B'_{n/2-1}(\lambda) - \left[\log\mu^2 + \Psi(n/2) - \mathbb{C}\right]B_{n/2-1}(\lambda)\right\} \overset{\lambda\to-\infty}{\sim} (-1)^{n/2-1}(4\pi)^{-n/2}\\ &\left\{\sum_{k=0}^{n/2-1}\frac{(-\lambda)^{n/2-1-k}}{k!\Gamma(n/2-k)}\left[\mathbb{C} - \Psi(n/2-k) + \log\left(\frac{-\lambda}{\mu^2}\right)\right]A_k\right.\\ &\left. + \sum_{k=n/2}^{\infty}\frac{(-1)^{k-n/2}}{k!}(-\lambda)^{n/2-1-k}\Gamma(k+1-n/2)A_k\right\}. \qquad (158)\end{aligned}$$

This trace determines the regularized vacuum expectation values like $\langle\varphi^2\rangle$ in quantum field theory.

Thus, we see that

(i) all the singularities of the Green function and the non-analytical parts thereof (proportional to $\sqrt{\sigma}$ in odd dimensions and to $\log\sigma$ in even dimensions) are determined by the values of the function $b_q(\lambda)$ at integer points q, which are determined, in turn, by the heat kernel coefficients a_k;
(ii) there are no power singularities, i.e., $G^{\mathrm{sing}}(\lambda) = 0$, in lower dimensions $n = 1, 2$;
(iii) there is no logarithmic singularity (more generally, no logarithmic part at all) in odd dimensions;
(iv) the regular part depends on the values of the function $b_q(\lambda)$ at half-integer points q and its derivative $b'_q(\lambda)$ at integer points q and is a global object that cannot be reduced to purely local information like the heat kernel coefficients a_k.

The logarithmic part of the Green function is very important. On the one hand it determines, as usual, the renormalization properties of the regular part of the Green function, i.e., the derivative $\mu(\partial/\partial\mu)G^{\mathrm{reg}}(\lambda)$. In particular,

$$\mu\frac{\partial}{\partial\mu}\mathrm{Tr}\, G^{\mathrm{reg}}(\lambda) = \begin{cases} 0 & \text{for odd } n \\ \dfrac{(4\pi)^{-n/2}}{\Gamma(n/2)} B_{n/2-1}(\lambda) & \text{for even } n. \end{cases} \tag{159}$$

On the other hand, it is of crucial importance in studying the Huygens principle. Namely, the absence of the logarithmic part of the Green function is a necessary and sufficient condition for the validity of the Huygens principle for hyperbolic operators. The heat kernel coefficients and, therefore, the logarithmic part of the Green function are defined for the hyperbolic operators just by analytic continuation from the elliptic case. Thus, the condition of the validity of Huygens' principle reads

$$\sum_{k=0}^{\infty} \frac{\Gamma(n/2)}{k!\Gamma(k+n/2)} \left(\frac{\sigma}{2}\right)^k b_{k-1+n/2}(\lambda) = 0, \tag{160}$$

or, by using (118),

$$\sum_{k=0}^{\infty} \sum_{j=0}^{k-1+n/2} \frac{\Gamma(n/2)}{k!j!\Gamma(k-j+n/2)} \left(\frac{\sigma}{2}\right)^k (-\lambda)^{k-j} a_j = 0\,. \tag{161}$$

By expanding this equation in covariant Taylor series using the methods of [3] one can obtain an infinite set of local conditions for validity of the Huygens principle, see [8]. In particular,

$$[b_{n/2-1}(\lambda)]^{\mathrm{diag}} = 0, \tag{162}$$

$$[\nabla_\mu b_{n/2-1}(\lambda)]^{\mathrm{diag}} = 0, \tag{163}$$

$$[\nabla_{(\mu}\nabla_{\nu)} b_{n/2-1}(\lambda)]^{\mathrm{diag}} + \frac{1}{2n} g_{\mu\nu}[b_{n/2}(\lambda)]^{\mathrm{diag}} = 0. \tag{164}$$

4.5 Heat Kernel Coefficients

As we have shown above the calculation of the effective action and the Green function reduces to the calculation of the heat kernel. An important part of that calculation is the calculation of the coefficients of the asymptotic expansion of the heat kernel. They are determined by a recursion system which is obtained simply by restricting the complex variable q in (128) to positive integer values $q = 1, 2, \ldots$.

4.5.1 Non-recursive Solution of the Recursion Relations

This problem was solved in [2, 3, 10] where a systematic technique for calculation of a_k was developed. The formal solution of this recursion system is

$$a_k = \left(1 + \frac{1}{k}D\right)^{-1} \tilde{L} \left(1 + \frac{1}{k-1}D\right)^{-1} \tilde{L} \cdots \left(1 + \frac{1}{1}D\right)^{-1} \tilde{L} \cdot I. \tag{165}$$

Now, the problem is to give a precise practical meaning to this formal operator solution. To do this one has, first of all, to define the inverse operator $(1 + D/k)^{-1}$. This can be done by constructing the complete set of eigenvectors of the operator D. However, first we introduce some auxiliary notions from the theory of symmetric tensors.

Let S^n_m be the bundle of symmetric tensors of type (m, n). First of all, we define the exterior symmetric tensor product

$$\vee : \; S^n_m \times S^i_j \to S^{n+i}_{m+j} \tag{166}$$

of symmetric tensors by

$$(A \vee B)^{\beta_1 \dots \beta_{n+i}}_{\alpha_1 \dots \alpha_{m+j}} = A^{(\beta_1 \dots \beta_n}_{(\alpha_1 \dots \alpha_m} B^{\beta_{n+1} \dots \beta_{n+i})}_{\alpha_{m+1} \dots \alpha_{m+j})} . \tag{167}$$

Next, we define the inner product

$$\star : \; S^n_m \times S^i_n \to S^i_m \tag{168}$$

by

$$(A \star B)^{\beta_1 \dots \beta_i}_{\alpha_1 \dots \alpha_m} = A^{\gamma_1 \dots \gamma_n}_{\alpha_1 \dots \alpha_m} B^{\beta_1 \dots \beta_i}_{\gamma_1 \dots \gamma_n} . \tag{169}$$

Finally, let $\mathbb{I}_{(n)}$ be the identity endomorphism on the space of symmetric tensors of type $(n, 0)$; it is a section of the bundle S^n_n, that is,

$$\mathbb{I}_{(n)}{}^{\mu_1 \dots \mu_n}_{\nu_1 \dots \nu_n} = \delta^{(\mu_1}_{(\nu_1} \cdots \delta^{\mu_n)}_{\nu_n)} . \tag{170}$$

We also define the exterior symmetric covariant derivative

$$\nabla^S : \; S^m_n \to S^m_{n+1} \tag{171}$$

by

$$(\nabla^S A)^{\beta_1 \dots \beta_m}_{\alpha_1 \dots \alpha_{n+1}} = \nabla_{\left(\alpha_1 A^{\beta_1 \dots \beta_m}_{\alpha_2 \dots \alpha_{n+1}}\right)} . \tag{172}$$

These definitions are naturally extended to $\mathrm{End}\,(\mathcal{V})$-valued symmetric tensors, i.e., to the sections of the bundle $S^m_n \otimes \mathrm{End}\,(\mathcal{V})$.

4.5.2 Covariant Taylor Basis

Let us consider the space of smooth two-point functions in a small neighborhood of the diagonal $x = x'$; we will denote such functions by $|f\rangle$. Let us define a special set of such functions $\{|n\rangle\}_{n=0}^{\infty}$, labeled by a non-negative integer n, by

$$|0\rangle = 1, \tag{173}$$

$$|n\rangle = \frac{1}{n!} y^{a_1} \cdots y^{a_n} , \tag{174}$$

where y^a are the geometric parameters (normal coordinates) defined by (104). These functions are scalars at the point x and symmetric tensors of type $(0, n)$ at the point x'. Recall the definition (111) of the operator D introduced in Sect. 4.3.4.1. It is easy to show that these functions satisfy the equation

$$D|n\rangle = n|n\rangle , \tag{175}$$

and, hence, are the eigenfunctions of the operator D with positive integer eigenvalues.

Let $\langle n|$ denote the dual linear functionals defined by

$$\langle n|f\rangle = (\nabla^S)^n f \Big|_{x=x'} , \tag{176}$$

so that

$$\langle n|m\rangle = \delta_{mn} \mathbb{I}_{(n)} . \tag{177}$$

Using this notation the covariant Taylor series for an analytic function $|f\rangle$ can be written in the form

$$|f\rangle = \sum_{n=0}^{\infty} |n\rangle \star \langle n|f\rangle , \tag{178}$$

For smooth functions the Taylor series is only an asymptotic series, which does not necessarily converge. For analytic functions, however, the Taylor series converges in a sufficiently small neighborhood of the fixed point x'. Therefore, the functions $|n\rangle$ form a complete orthonormal basis in the subspace of analytic functions. This is a reflection of the fact that an analytic function that is orthogonal to all functions $|n\rangle$, that is, whose all symmetrized derivatives vanish at the point x', is, in fact, identically equal to zero in this neighborhood. Note, however, that the space of functions we are talking about is not a Hilbert space since there are many analytic functions $|f\rangle$ such that the norm $\langle f|f\rangle$ defined above diverges. If we restrict ourselves to polynomials of some order, then this problem does not appear, and, hence, the space of polynomials is a Hilbert space with the inner product defined above.

4.5.3 Matrix Algorithm

The complete set of eigenfunctions $|n\rangle$ can be employed to present the action of the operator $\tilde{L}$ on a function $|f\rangle$ in the form

$$\tilde{L}|f\rangle = \sum_{m,n\geq 0} |m\rangle \star \langle m|\tilde{L}|n\rangle \star \langle n|f\rangle, \tag{179}$$

where $\langle m|\tilde{L}|n\rangle$ are the "matrix elements" of the operator $\tilde{L}$ that are just End $(\mathcal{V})$-valued symmetric tensors, i.e., sections of the vector bundle $S^n_m \otimes \text{End}\,(\mathcal{V})$. When acting on an analytic function this series is nothing but the Taylor series of the result and converges in a sufficiently small neighborhood of the point x'; for smooth functions it gives an asymptotic expansion.

Now it should be clear that the inverse of the operator $\left(1+\frac{1}{k}D\right)^{-1}$ can be defined by

$$\left(1+\frac{1}{k}D\right)^{-1}|f\rangle = \sum_{n=0}^{\infty}\frac{k}{k+n}|n\rangle \star \langle n|f\rangle. \tag{180}$$

Using such representations for the operators $\left(1+\frac{1}{k}D\right)^{-1}$ and $\tilde{L}$ we obtain a covariant Taylor series for the coefficients a_k

$$a_k = \sum_{n=0}^{\infty} |n\rangle \star \langle n|a_k\rangle, \tag{181}$$

where

$$\langle n|a_k\rangle = \sum_{n_1,\dots,n_{k-1}\geq 0} \frac{k}{k+n}\cdot\frac{k-1}{k-1+n_{k-1}}\cdots\frac{2}{2+n_2}\cdot\frac{1}{1+n_1}$$

$$\times \langle n|\tilde{L}|n_{k-1}\rangle \star \langle n_{k-1}|\tilde{L}|n_{k-2}\rangle \star \cdots \star \langle n_1|\tilde{L}|0\rangle\,, \tag{182}$$

where the summation is over all non-negative integers $n_1,\dots,n_{k-1}$. It is not difficult to show that for a differential operator of second order the matrix elements $\langle m|\tilde{L}|n\rangle$ do not vanish only for $n \leq m+2$. Therefore, the summation over n_i here is limited from above by

$$0 \leq n_1\,, \qquad n_i \leq n_{i+1}+2\,, \qquad (i=1,2,\dots,k-1)\,, \tag{183}$$

where $n_k \equiv n$. Thus, the sum (182) contains only a finite number of terms.

Thus, we have reduced the problem of computation of the heat kernel coefficients a_k to the evaluation of the matrix elements $\langle m|\tilde{L}|n\rangle$ of the operator $\tilde{L}$. For a differential operator $\tilde{L}$ of second order, the matrix elements $\langle m|\tilde{L}|n\rangle$ vanish for $n > m + 2$. Therefore, the summation over n_i in (182) is limited from above: $n_1 \geq 0$, and $n_i \leq n_{i+1} + 2$, for $i = 1, 2, \ldots, k-1$, and, hence, the sum (182) always contains only a finite number of terms.

The matrix elements $\left\langle n|\tilde{L}|m\right\rangle$ of a Laplace-type operator have been computed in our papers [3, 1]. They have the following general form:

$$\langle m|L|m+2\rangle = -g^* \vee \mathbb{I}_{(m)}, \tag{184}$$

$$\langle m|L|m+1\rangle = 0, \tag{185}$$

$$\begin{aligned}\langle m|L|n\rangle = &\binom{m}{n}\mathbb{I}_{(n)} \vee Z_{(m-n)} + \binom{m}{n-1}\mathbb{I}_{(n-1)} \vee Y_{(m-n+1)} \\ &+\binom{m}{n-2}\mathbb{I}_{(n-2)} \vee X_{(m-n+2)},\end{aligned} \tag{186}$$

where g^* is the metric on the cotangent bundle, $Z_{(n)}$ is a section of the vector bundle $S_n \otimes \mathrm{End}\,(\mathcal{V})$, $Y_{(n)}$ is a section of the vector bundle $S^1_n \otimes \mathrm{End}\,(\mathcal{V})$, and $X_{(n)}$ is a section of the vector bundle S^2_n (a symmetric tensor of type $(2, n)$). Here it is also meant that the binomial coefficient $\binom{n}{k}$ is equal to zero if $k < 0$ or $n < k$.

We will not present here explicit formulas (they have been computed explicitly for arbitrary m, n in [3, 10]) but note that all these quantities are expressed polynomially in terms of three sorts of geometric data as follows:

(i) symmetric tensors of type $(2, n)$, i.e., sections of the bundle S^2_n obtained by symmetric derivatives

$$K_{(n)} = (\nabla^S)^{n-2}\mathrm{Riem} \tag{187}$$

of the symmetrized Riemann tensor Riem taken as a section of the bundle S^2_2,

(ii) sections

$$\mathcal{R}_{(n)} = (\nabla^S)^{n-1}\mathcal{R} \tag{188}$$

of the vector bundle $S^1_n \otimes \mathrm{End}\,(\mathcal{V})$ obtained by symmetrized derivatives of the curvature $\mathcal{R}$ of the connection $\nabla^{\mathcal{V}}$ taken as a section of the bundle $S^1_1 \otimes \mathrm{End}\,(\mathcal{V})$,

(iii) $\mathrm{End}\,(\mathcal{V})$-valued symmetric forms, i.e., sections of the vector bundle $S^0_n \otimes \mathrm{End}\,(\mathcal{V})$, constructed from the symmetrized covariant derivatives

$$Q_{(n)} = (\nabla^S)^n Q \tag{189}$$

of the endomorphism Q.

From dimensional arguments it is obvious that the matrix elements $\langle n|L|n\rangle$ are expressed in terms of the Riemann curvature tensor Riem, the bundle curvature $\mathcal{R}$, and the endomorphism Q; the matrix elements $\langle n+1|L|n\rangle$ in terms of the quantities ∇Riem, $\nabla\mathcal{R}$, and ∇Q; the elements $\langle n+2|L|n\rangle$ in terms of the quantities of the form $\nabla\nabla$Riem, Riem $\cdot$ Riem, etc.

4.5.4 Diagramatic Technique

In the computation of the heat kernel coefficients by means of the matrix algorithm a "diagrammatic" technique, i.e., a graphic method for enumerating the different terms of the sum (182), turns out to be very convenient and pictorial [3, 10].

The matrix elements $\langle m|L|n\rangle$ are presented by some blocks with m lines coming in from the left and n lines going out to the right (Fig. 4.1),
and the product of the matrix elements $\langle m|L|k\rangle \star \langle k|L|n\rangle$ – by two blocks connected by k intermediate lines (Fig. 4.2)
that represents the contractions of the corresponding tensor indices (the inner product).

To obtain the coefficient $\langle n|a_k\rangle$ one should draw, first, all possible diagrams which have n lines incoming from the left and which are constructed from k blocks connected in all possible ways by any number of intermediate lines. When doing this, one should keep in mind that the number of the lines going out of any block cannot be greater than the number of the lines coming in by more than two and by exactly one. Then one should sum up all diagrams with the weight determined for each diagram by the number of intermediate lines from the analytical formula (182). Drawing of such diagrams is of no difficulties. This helps to keep under control the whole variety of different terms. Therefore, the main problem is reduced to the computation of some standard blocks, which can be computed once and for all.

For example, the diagrams for the diagonal values of the HMDS coefficients $a_k^{\text{diag}} = \langle 0|\mathrm{a_k}\rangle$ have the form

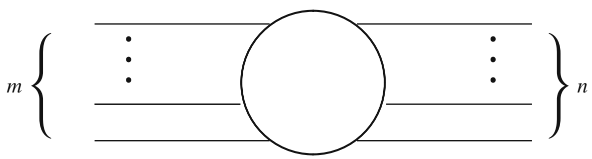


Fig. 4.1 Matrix elements

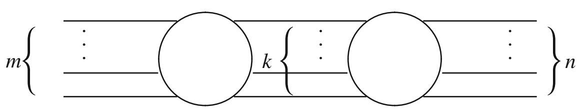
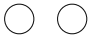
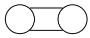
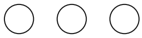
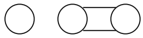
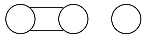
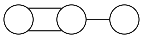
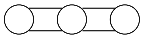
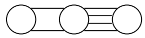


Fig. 4.2 Product of matrix elements

$a_1^{\mathrm{diag}} =$  (190)

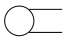
$a_2^{\mathrm{diag}} =$  $+\frac{1}{3}$  (191)

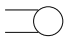
$a_3^{\mathrm{diag}} =$  $+\frac{1}{3}$  $+\frac{2}{4}$  (192)

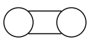
$+\frac{2}{4}\cdot\frac{1}{2}$  $+\frac{2}{4}\cdot\frac{1}{3}$  $+\frac{2}{4}\cdot\frac{1}{5}$  .

As an illustration let us compute the coefficients a_1^{diag} and a_2^{diag}. We have [3]

 $= \langle 0|L|0\rangle = Z_{(0)}$ (193)

 $= \langle 0|L|2\rangle = -g^{ab}$ (194)

 $= \langle 2|L|0\rangle = Z_{(2)ab}$ (195)

 $= \langle 0|L|2\rangle \star \langle 2|L|0\rangle = -g^{ab} Z_{(2)ab},$ (196)

where

$$Z_{(0)} = Q - \frac{1}{6} R \; \mathbb{I}_V, \tag{197}$$

$$Z_{(2)ab} = \nabla_{(a}\nabla_{b)} Q - \frac{1}{2}\mathcal{R}_{c(a}\mathcal{R}^{c}{}_{b)} + \frac{1}{2}\nabla_{(a}\nabla_{|c|}\mathcal{R}^{c}{}_{b)} \tag{198}$$
$$+\mathbb{I}_V\left(-\frac{3}{20}\nabla_a\nabla_b R - \frac{1}{20}\Delta R_{ab} + \frac{1}{15}R_{ac}R^{c}{}_{b} - \frac{1}{30}R_{acde}R_{b}{}^{cde} - \frac{1}{30}R_{cd}R^{c}{}_{a}{}^{d}{}_{b}\right).$$

Here, as usual, the parenthesis denote the complete symmetrization over all indices included and the vertical lines indicate the indices excluded from the symmetrization. Hence, we immediately get

$$a_1^{\text{diag}} = Q - \frac{1}{6} R\, \mathbb{I}_V, \tag{199}$$

and, by taking the trace of $Z_{(2)}$ and using the identity $\nabla_a \nabla_b \mathcal{R}^{ab} = 0$, we obtain the well-known result [3]

$$a_2^{\text{diag}} = \left(Q - \frac{1}{6} R \mathbb{I}_V\right)^2 - \frac{1}{3}\Delta Q + \frac{1}{6}\mathcal{R}_{ab}\mathcal{R}^{ab} + \mathbb{I}_V \left(\frac{1}{15}\Delta R - \frac{1}{90} R_{ab} R^{ab} + \frac{1}{90} R_{abcd} R^{abcd}\right). \tag{200}$$

The technique described above is manifestly covariant and is applicable for any Riemannian (or pseudo-Riemannian) manifold M and for any vector bundle V. It is also valid for local analysis on noncompact manifolds and manifolds with boundary (at finite distance from the boundary). This method gives not only the diagonal values of the heat kernel coefficients but also the diagonal values of all their derivatives, that is, it gives also the off-diagonal coefficients in form of a covariant Taylor series. Due to the use of symmetric forms and symmetric covariant derivatives the famous "combinatorial explosion" in the complexity of the heat kernel coefficients is avoided. This technique is very algorithmic and well suited to automated computation. The developed method is very powerful; it enabled us to compute for the first time the diagonal value of the fourth HMDS coefficient a_4^{diag} [2, 3]. It was used in [35, 33] to compute the coefficient a_5^{diag}. Lastly, this technique enables one not only to carry out explicit computations but also to analyze the general structure of the heat kernel coefficients for all orders k.

4.5.5 General Structure of Heat Kernel Coefficients

Now we are going to investigate the general structure of the heat kernel coefficients. We will follow mainly our papers [10, 9, 11].

Our analysis will be again purely local. Since locally one can always expand the metric, the connection and the endomorphism Q in the covariant Taylor series, they are completely characterized by their Taylor coefficients, i.e., the covariant derivatives of the curvatures, more precisely by the objects $K_{(n)}$, $\mathcal{R}_{(n)}$, and $Q_{(n)}$ introduced above. We introduce the following notation for all of them:

$$\Re_{(n)} = \{K_{(n+2)}, \mathcal{R}_{(n+1)}, Q_{(n)}\}, \qquad (n = 0, 1, 2, \dots) \tag{201}$$

and call these objects *covariant jets*; n will be called the order of a jet $\Re_{(n)}$. It is worth noting that the jets are defined by *symmetrized* covariant derivatives. This makes them well defined as the order of the derivatives becomes not important. It is only the *number* of derivatives that plays a role.

The low-order coefficients A_0 and A_1 have been described above. As far as the higher order coefficients A_k, $(k \geq 2)$, are concerned, they are integrals of local invariants which are polynomial in the jets. One can classify all the terms in them according to the number of the jets and their order. The terms linear in the jets in higher order coefficients A_k, ($k \geq 2$), are given by integrals of total derivatives, symbolically $\int_M d\mathrm{vol}\ \mathrm{tr}_V\ \Delta^{k-1}\Re$. They are calculated explicitly in [3, 10, 1]. Since the total derivatives do not contribute to an integral over a complete compact manifold, it is clear that the linear terms vanish. Thus A_k, ($k = 2, 3, \dots$), begin with the terms quadratic in the jets. These terms contain the jets of highest order (or the *leading derivatives* of the curvatures) and can be shown to be of the form $\int_M d\mathrm{vol}\ \ \mathrm{tr}_V\ \Re\Delta^{k-2}\Re$. Then it follows a class of terms cubic in the jets, etc. The last class of terms does not contain any covariant derivatives at all but only the powers of the curvatures. In other words, the higher order HMDS coefficients have a general structure, which can be presented symbolically in the following form. For $k \geq 2$ one can classify the terms in A_{2k} according to the number of the jets and their order

$$A_k = \sum_{j=2}^{k} A_{k,(j)}\,, \tag{202}$$

where $A_{k,(j)}$ is the contribution of order j in the jets; they can be presented symbolically in the form

$$\begin{aligned}
A_{k,(2)} &= \int\limits_M d\mathrm{vol}\ \ \mathrm{tr}_V \sum \Re_{(0)}\Re_{(2k-4)}\,, &(203)\\
A_{k,(3)} &= \int\limits_M d\mathrm{vol}\ \ \mathrm{tr}_V \sum_{i=0}^{2k-6}\sum \Re_{(0)}\Re_{(i)}\Re_{(2k-6-i)}\,, &(204)\\
&\dots\\
A_{k,(k-1)} &= \int\limits_M d\mathrm{vol}\ \ \mathrm{tr}_V \left[\sum \Re_{(2)}\left(\Re_{(0)}\right)^{k-2} + \sum\left(\Re_{(1)}\right)^2\left(\Re_{(0)}\right)^{k-3}\right], &(205)\\
A_{k,(k)} &= \int\limits_M d\mathrm{vol}\ \ \mathrm{tr}_V \sum\left(\Re_{(0)}\right)^k\,. &(206)
\end{aligned}$$

More precisely, the functionals $A_{k,(j)}$ transform under the rescaling of the jets

$$\Re_{(k)} \mapsto \varepsilon\alpha^k\Re_{(k)} \tag{207}$$

as follows

$$A_{k,(j)} \mapsto \varepsilon^j\alpha^{2(k-j)}A_{k,(j)}\,. \tag{208}$$

4.6 High-Energy Approximation

One can show that all quadratic terms can be reduced to five independent invariants, viz. [1, 3, 10]

$$A_{k,(2)} = \frac{k!(k-2)!}{2(2k-3)!}\int\limits_M d\mathrm{vol}\ \mathrm{tr}_V \left\{ f_k^{(1)} Q\Delta^{k-2}Q + 2f_k^{(2)}\mathcal{R}^{bc}\nabla_b\Delta^{k-3}\nabla_a\mathcal{R}^a{}_c \right.$$

$$\left. +f_k^{(3)} Q\Delta^{k-2}R + f_k^{(4)} R_{ab}\Delta^{k-2}R^{ab} + f_k^{(5)} R\Delta^{k-2}R \right\}, \tag{209}$$

where $f_k^{(i)}$ are some numerical coefficients. These numerical coefficients can be computed by the technique developed in the previous section. From the formula (182) we have for the diagonal coefficients a_k^{diag} up to cubic terms in the jets

$$\begin{aligned} a_k^{\mathrm{diag}} =\ & \langle 0|a_k\rangle = \frac{(-1)^{k-1}}{\binom{2k-1}{k}}\langle 0; k-1|L|0\rangle \\ & +(-1)^k \sum_{i=1}^{k-1}\sum_{n_i=0}^{2(k-i-1)} \frac{\binom{2k-1}{i}}{\binom{2k-1}{k}\binom{2i+n_i-1}{i}} \langle 0; k-i-1|L|n_i\rangle \star \langle n_i; i-1|L|0\rangle \\ & +O(\Re^3), \end{aligned} \tag{210}$$

where

$$\langle n; k|L|m\rangle = (\vee^k g^*) \star \langle n|L|m\rangle \tag{211}$$

and $O(\Re^3)$ denote terms of third order in the jets.

By computing the matrix elements in the second order in the jets and integrating over M one obtains [1, 3]

$$f_k^{(1)} = 1\,, \tag{212}$$

$$f_k^{(2)} = \frac{1}{2(2k-1)}\,, \tag{213}$$

$$f_k^{(3)} = \frac{k-1}{2(2k-1)}\,, \tag{214}$$

$$f_k^{(4)} = \frac{1}{2(4k^2-1)}\,, \tag{215}$$

$$f_k^{(5)} = \frac{k^2-k-1}{4(4k^2-1)}\,. \tag{216}$$

One should note that the same results were obtained by a completely different method in [17].

It is now easy to see that the expression for A_k can be analytically continued to an entire function A_q of q, which ultimately gives the function $B_q(\lambda)$ introduced in Sect. 4.3.4. This gives an example of how the function $B_q(\lambda)$ can be computed without solving the differential equation (for more details, see [1, 3]).

Let us consider the situation when the curvatures are small but rapidly varying (*high-energy approximation* in quantum field theory), i.e., the derivatives of the curvatures are more important than the powers of them. This corresponds to an asymptotic expansion in the deformation parameter ε as $\varepsilon \to 0$. Then the leading derivative terms in the heat kernel are the largest ones. Thus the heat trace has the form

$$\Theta(t) \sim (4\pi t)^{-n/2}\left\{A_0 - tA_1 + \frac{t^2}{2}H_2(t)\right\} + O(\Re^3), \tag{217}$$

where $H_2(t)$ is some complicated *nonlocal* functional that has the following asymptotic expansion as $t \to 0$

$$H_2(t) \sim 2\sum_{k=2}^{\infty}\frac{(-t)^{k-2}}{k!}A_{k,(2)}. \tag{218}$$

Using the results for $A_{k,(2)}$ one can easily construct such a functional H_2 just by a formal summation of the leading derivatives

$$\begin{aligned} H_2(t) = \int\limits_M d\text{vol}\ \operatorname{tr}_V \Big\{ & Q\gamma^{(1)}(-t\Delta)Q + 2\mathcal{R}^a{}_c\nabla_a\frac{1}{\Delta}\gamma^{(2)}(-t\Delta)\nabla_b\mathcal{R}^{bc} \\ & -2Q\gamma^{(3)}(-t\Delta)R + R_{ab}\gamma^{(4)}(-t\Delta)R^{ab} + R\gamma^{(5)}(-t\Delta)R\Big\}, \end{aligned} \tag{219}$$

where $\gamma^{(i)}(z)$ are entire functions defined by [1, 3]

$$\gamma^{(i)}(z) = \sum_{k=0}^{\infty}\frac{k!}{(2k+1)!}f_k^{(i)}z^k = \int\limits_0^1 d\xi\ f^{(i)}(\xi)\exp\left(-\frac{1-\xi^2}{4}z\right), \tag{220}$$

where

$$f^{(1)}(\xi) = 1\,, \tag{221}$$

$$f^{(2)}(\xi) = \frac{1}{2}\xi^2\,, \tag{222}$$

$$f^{(3)}(\xi) = \frac{1}{4}(1-\xi^2)\,, \tag{223}$$

$$f^{(4)}(\xi) = \frac{1}{6}\xi^4\,, \tag{224}$$

$$f^{(5)}(\xi) = \frac{1}{48}(3-6\xi^2-\xi^4). \tag{225}$$

Therefore, $H_2(t)$ can be regarded as generating functional for quadratic terms $A_{k,(2)}$ (leading derivative terms) in all coefficients A_k. It also plays a very important role in investigating the nonlocal structure of the effective action in quantum field theory in high-energy approximation [1, 3].

4.7 Low-Energy Approximation

Let us consider now the opposite case, when the curvatures are strong but slowly varying (low-energy approximation in quantum field theory), i.e., the powers of the curvatures are more important than their derivatives. This corresponds to the asynptotic expansion in the deformation parameter α as $\alpha \to 0$. The main terms in this approximation are the terms without any covariant derivatives of the curvatures, i.e., the lowest order jets. We will consider mostly the zeroth order of this approximation, which corresponds simply to covariantly constant background curvatures

$$\nabla \text{Riem} = 0, \qquad \nabla \mathcal{R} = 0, \qquad \nabla Q = 0. \tag{226}$$

The asymptotic expansion of the heat trace

$$\Theta(t) \sim (4\pi t)^{-n/2} \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} A_{k,(k)} \tag{227}$$

determines then *all* the terms without covariant derivatives (highest order terms in the jets), $A_{k,(k)}$, in all heat kernel coefficients A_k. These terms do not contain any covariant derivatives and are just polynomials in the curvatures and the endomorphism Q. Thus the heat trace is a generating functional for all heat kernel coefficients for a covariantly constant background, in particular, for all symmetric spaces. Thus the problem is to calculate the heat trace for covariantly constant background.

4.7.1 Algebraic Approach

There exists a very elegant indirect way to construct the heat kernel without solving the heat equation but using only the commutation relations of some covariant first-order differential operators. Below we will follow our papers [4–7, 12, 13]. The main idea is to employ a generalization of the usual Fourier transform to the case of operators; it consists in the following. We are going to use the following representation of the heat trace:

$$\Theta(t) = \int\limits_M d\mathrm{vol}\ \mathrm{tr}_V \left[\exp(-tL)\delta(x,x')\right]^{\mathrm{diag}} . \tag{228}$$

Let us consider for a moment a trivial case, where the curvatures vanish but the potential term does not

$$\mathrm{Riem} = 0, \qquad \mathcal{R} = 0, \qquad \nabla Q = 0. \tag{229}$$

In this case the operators of covariant derivatives obviously commute and form, together with the potential term, an Abelian Lie algebra

$$[\nabla_\mu, \nabla_\nu] = 0, \qquad [\nabla_\mu, Q] = 0. \tag{230}$$

It is easy to show that the heat semigroup can be presented in the form

$$\exp(-tL) = (4\pi t)^{-n/2} \exp(-tQ) \int\limits_{\mathbb{R}^n} dk\ g^{1/2} \exp\left(-\frac{1}{4t}\langle k, gk\rangle + k\cdot\nabla\right), \tag{231}$$

where $\langle k, gk\rangle = k^\mu g_{\mu\nu} k^\nu$ and $k \cdot \nabla = k^\mu \nabla_\mu$. Here, of course, it is assumed that the covariant derivatives also commute with the metric

$$[\nabla, g] = 0. \tag{232}$$

Acting with this operator on the Dirac distribution and using the obvious relation

$$\left[\exp(k\cdot\nabla)\delta(x,x')\right]^{\mathrm{diag}} = \delta(k), \tag{233}$$

one can integrate easily over k to obtain the heat trace

$$\Theta(t) = (4\pi t)^{-n/2} \int\limits_M d\mathrm{vol}\ \mathrm{tr}_V \exp(-tQ). \tag{234}$$

Of course, on curved manifolds the covariant differential operators ∇ do not commute – their commutators are determined by the curvatures $\Re$. The commutators of covariant derivatives ∇ with the curvatures $\Re$ give the first derivatives of the curvatures, i.e., the jets $\Re_{(1)}$, the commutators of covariant derivatives with $\Re_{(1)}$ give the second jets $\Re_{(2)}$, etc. Thus the operators ∇ together with the whole set of the jets $\mathcal{J}$ form an infinite-dimensional Lie algebra $\mathcal{G} = \{\nabla, \Re_{(i)}; (i = 1, 2, \dots)\}$.

Now, let us remember that the heat trace is a functional of the jets, with the jets being defined by symmetrized covariant derivatives. This makes the order of a jet well defined. For example, the structures involving commutators of covariant derivatives (like $[\nabla_a, \nabla_b]R^e{}_c{}^f{}_d$, which involve two-jets of the Riemann tensor on the left but, after using the Ricci identity, only zero-jets on the right) are not allowed. After symmetrizing over *abcd* this jet vanishes. So, if we express the final answer for the heat kernel diagonal or for the heat kernel coefficients in terms of the symmetrized jets, then there is a natural filtration with respect to the order of the jets involved. In other words, one can always say, what is the maximal order of symmetrized covariant derivative of the curvature involved in the result. This is especially true for the heat kernel coefficients A_k since they are polynomial in the jets.

If we identify a small deformation parameter α with each derivative then a jet of order n is, actually, of order α^n. Thus, we get a perturbation theory in this small parameter. Since the derivatives are naturally identified with the momentum (or energy), the physicists call a situation when the derivatives are small the *low-energy approximation*. To evaluate the heat kernel in the low-energy approximation one can take into account only a finite number of low-order jets, i.e., the low-order covariant derivatives of the background fields, $\{\Re_{(i)}; (i \le N)\}$ with some fixed N, and neglect all higher order jets, i.e., the covariant derivatives of higher orders, i.e., put $\Re_{(i)} = 0$ for $i > N$. Then one can show that there exist a set of covariant differential operators that, together with the low-order jets, generate a finite-dimensional Lie algebra $\mathcal{G}_N = \{\nabla, \Re_{(i)}; (i = 1, 2, \dots, N)\}$. One should stress here what problem one can solve this way. We try to answer the following concrete question: How do the heat kernel coefficients look if we throw away all the (symmetrized) jets of order higher than N?

Thus one can try to generalize the above idea in such a way that (231) would be the zeroth approximation in the commutators of the covariant derivatives, i.e., in the curvatures. Roughly speaking, we would like to find a representation of the heat semigroup in the form

$$\exp(-tL) = (4\pi t)^{-D/2} \int\limits_{\mathbb{R}^D} dk\, \Phi(t,k) \exp\left(-\frac{1}{4t}\, \langle k, \Psi(t)k\rangle + T(k)\right), \qquad (235)$$

where $\langle k, \Psi(t)k\rangle = k^A \Psi_{AB}(t) k^B$, $T(k) = k^A T_A$ $(A = 1, 2, \dots, D)$ T_A are some first-order differential operators and the functions $\Psi(t)$ and $\Phi(t,k)$ are expressed in terms of commutators of these operators, i.e., in terms of the curvatures; that is, these functions are analytic functions of t. In general, the operators T_A do not form a closed finite-dimensional algebra because at each step, by taking more commutators, there appear more and more derivatives of the curvatures. It is the low-energy reduction $\mathcal{G} \mapsto \mathcal{G}_N$, i.e., the restriction to the low-order jets, that actually closes the algebra $\mathcal{G}$ of the operators T_A and the background jets, i.e., makes it finite-dimensional.

Using this representation one can, as above, act with $\exp[T(k)]$ on the Dirac distribution to get the heat kernel. The main point of this idea is that it is much easier to calculate the action of the exponential of the first-order operator $T(k)$ on the Dirac distribution than that of the exponential of the second-order operator L.

4.7.2 Covariantly Constant Background in Flat Space

Let us consider now the more complicated case of nontrivial covariantly constant curvature of the connection on the vector bundle V in flat space as follows:

$$\mathrm{Riem} = 0, \qquad \nabla\mathcal{R} = 0, \qquad \nabla Q = 0. \tag{236}$$

Using the condition of covariant constancy of the curvatures one can show that in this case the covariant derivatives form a *nilpotent* Lie algebra [4]

$$[\nabla_\mu, \nabla_\nu] = \mathcal{R}_{\mu\nu}, \tag{237}$$
$$[\nabla_\mu, \mathcal{R}_{\alpha\beta}] = [\nabla_\mu, Q] = 0, \tag{238}$$
$$[\mathcal{R}_{\mu\nu}, \mathcal{R}_{\alpha\beta}] = [\mathcal{R}_{\mu\nu}, Q] = 0. \tag{239}$$

For this algebra one can prove a theorem expressing the heat semigroup operator in terms of an average over the corresponding Lie group [4]

$$\exp(-tL) = (4\pi t)^{-n/2}\exp(-tQ)\left[\det{}_{TM}\left(\frac{t\mathcal{R}}{\sinh(t\mathcal{R})}\right)\right]^{1/2} \tag{240}$$
$$\times\int\limits_{\mathbb{R}^n} dk\; g^{1/2}\exp\left(-\frac{1}{4t}\langle k, gt\mathcal{R}\coth(t\mathcal{R})k\rangle + k\cdot\nabla\right), \tag{241}$$

where $k\cdot\nabla = k^\mu\nabla_\mu$. Here functions of the curvatures $\mathcal{R}$ are understood as functions of sections of the bundle $\mathrm{End}\,(TM)\otimes\mathrm{End}\,(\mathcal{V})$, and the determinant $\det_{TM}$ is taken with respect to the tangent space indices; the fiber indices of the bundle $\mathcal{V}$ being intact.

It is not difficult to show that in this case we also have

$$\left[\exp(k\cdot\nabla)\,\delta(x,x')\right]^{\mathrm{diag}} = \delta(k)\,. \tag{242}$$

Subsequently, the integral over k becomes trivial and we obtain immediately the trace of the heat kernel [4]

$$\Theta(t) = (4\pi t)^{-n/2}\int\limits_M d\mathrm{vol}\;\mathrm{tr}_V\exp(-tQ)\left[\det{}_{TM}\left(\frac{t\mathcal{R}}{\sinh(t\mathcal{R})}\right)\right]^{1/2}. \tag{243}$$

Expanding it in a power series in t one can find *all* covariantly constant terms in *all* heat kernel coefficients A_k.

As we have seen the contribution of the curvature $\mathcal{R}_{\mu\nu}$ is not as trivial as that of the potential term. However, the algebraic approach does work in this case too. It is a good example of how one can get the heat kernel without solving any differential equations but using only the algebraic properties of the covariant derivatives.

4.7.2.1 Quadratic Potential in Flat Space

In fact, in flat space it is possible to do a bit more, i.e., to calculate the contribution of the first and the second derivatives of the potential term Q [6]. That is, we consider the case when the derivatives of the endomorphism Q vanish only starting from the *third* order, i.e.,

$$\text{Riem} = 0, \qquad \nabla\mathcal{R} = 0, \qquad \nabla\nabla\nabla Q = 0. \tag{244}$$

Besides we assume the background to be *Abelian*, i.e., all the nonvanishing background quantities, $\mathcal{R}_{\alpha\beta}$, Q, $Q_{;\mu} \equiv \nabla_\mu Q$ and $Q_{;\nu\mu} \equiv \nabla_\mu \nabla_\nu Q$, commute with each other. Thus we have again a nilpotent Lie algebra

$$[\nabla_\mu, \nabla_\nu] = \mathcal{R}_{\mu\nu}\,, \tag{245}$$

$$[\nabla_\mu, Q] = Q_{;\mu}\,, \tag{246}$$

$$[\nabla_\mu, Q_{;\nu}] = Q_{;\nu\mu}\,, \tag{247}$$

all other commutators being zero.

Now, let us represent the endomorphism Q in the form

$$Q = Q_0 - \alpha^{ik} N_i N_k, \tag{248}$$

where $(i = 1, \dots, q; q \le n)$, α^{ik} is some constant symmetric nondegenerate $q \times q$ matrix, Q_0 is a covariantly constant endomorphism, and N_i are some endomorphisms with vanishing second covariant derivative as follows:

$$\nabla Q_0 = 0, \qquad \nabla\nabla N_i = 0. \tag{249}$$

Next, let us introduce the operators $X_A = (\nabla_\mu, N_i)$ $(A = 1, \dots, n+q)$ and the matrix

$$(\mathcal{F}_{AB}) = \begin{pmatrix} \mathcal{R}_{\mu\nu} & N_{i;\mu} \\ -N_{k;\nu} & 0 \end{pmatrix}, \tag{250}$$

with $N_{i;\mu} \equiv \nabla_\mu N_i$.

The operator L can now be written in the form

$$L = -G^{AB} X_A X_B + Q_0\,, \tag{251}$$

where

$$(G^{AB}) = \begin{pmatrix} g^{\mu\nu} & 0 \\ 0 & \alpha^{ik} \end{pmatrix}. \tag{252}$$

and the commutation relations (247) take a more compact form

$$[X_A, X_B] = \mathcal{F}_{AB}, \tag{253}$$

all other commutators being zero.

This algebra is again a nilpotent Lie algebra. Thus one can apply the previous results in this case too to get [6]

$$\begin{aligned}\exp(-tL) = (4\pi t)^{-(n+q)/2} \exp(-tQ_0) \left[\det\left(\frac{t\mathcal{F}}{\sinh(t\mathcal{F})}\right)\right]^{1/2} \\ \times \int_{\mathbb{R}^{n+q}} dk\, G^{1/2} \exp\left(-\frac{1}{4t}\langle k, Gt\mathcal{F}\coth(t\mathcal{F})k\rangle + X(k)\right),\end{aligned} \tag{254}$$

where $G = \det G_{AB}$ and $X(k) = k^A X_A$.

Thus we have expressed the heat semigroup operator in terms of the operator $\exp[X(k)]$. The integration over k is Gaussian except for the noncommutative part. Splitting the integration variables $(k^A) = (q^\mu, \omega^i)$ and using the Campbell–Hausdorff formula we obtain [6]

$$\left[\exp[X(k)]\,\delta(x, x')\right]^{\text{diag}} = \exp[N(\omega)]\,\delta(q), \tag{255}$$

where $N(\omega) = \omega^i N_i$. Further, after taking off the trivial integration over q and a Gaussian integral over ω, we obtain the heat trace [6].

To describe the result let us introduce a matrix determined by second derivatives of the potential term as follows

$$P = \left(P^\mu{}_\nu\right), \qquad P^\mu{}_\nu = \frac{1}{2}\nabla^\mu\nabla_\nu Q, \tag{256}$$

and the matrices $C(t) = (C^\mu{}_\nu(t))$, $K(t) = (K^\mu{}_\nu(t))$ $S(t) = (S^\mu{}_\nu(t))$, and $E(t) = (E^\mu{}_\nu(t))$ by

$$C(t) = \oint\limits_C \frac{dz}{2\pi i} F(z)\, t \coth\left(\frac{t}{z}\right), \tag{257}$$

$$K(t) = \oint\limits_C \frac{dz}{2\pi i} F(z) \frac{t}{z^2} \sinh\left(\frac{t}{z}\right), \tag{258}$$

$$S(t) = \oint\limits_C \frac{dz}{2\pi i} F(z) \frac{t}{z} \sinh\left(\frac{t}{z}\right), \tag{259}$$

$$E(t) = \oint\limits_C \frac{dz}{2\pi i} F(z)\, t \sinh\left(\frac{t}{z}\right), \tag{260}$$

where

$$F(z) = (1 - z\mathcal{R} - z^2 P)^{-1} \tag{261}$$

and the integral is taken along a sufficiently small closed contour C that encircles the origin counter-clockwise, so that $F(z)$ is analytic inside this contour.

Then the heat trace has the form

$$\Theta(t) = (4\pi t)^{-n/2} \int\limits_M d\text{vol } \operatorname{tr}_V [\Phi(t)]^{-1/2} \exp\left[-tQ + \frac{1}{4} t^3 \langle \nabla Q, \Psi(t) \nabla Q \rangle\right], \tag{262}$$

where $\langle \nabla Q, \Psi(t) \nabla Q \rangle = \nabla_\mu Q \Psi^\mu{}_\nu(t) \nabla^\nu Q$,

$$\begin{aligned} \Phi(t) = \det{}_{TM} K(t) \det{}_{TM} \Big[1 + t^2 C(t) P\Big], \\ \times \det{}_{TM} \Big\{1 + t^2 [E(t) - S(t) K^{-1}(t) S(t)] P\Big\}, \end{aligned} \tag{263}$$

$$\Psi(t) = \left(\Psi^\mu_\nu(t)\right) = \Big[1 + t^2 C(t) P\Big]^{-1} C(t). \tag{264}$$

The formula (262) exhibits the general structure of the heat trace. One sees immediately how the endomorphism Q and its first derivatives ∇Q enter the result. The nontrivial information is contained only in a scalar $\Phi(t)$, and a tensor $\Psi_{\mu\nu}(t)$. These objects are constructed purely from the curvature $\mathcal{R}_{\mu\nu}$ and the second derivatives of the endomorphism Q, $\nabla\nabla Q$. Thus, the heat kernel coefficients A_k are constructed from three different types of scalar (connected) blocks Q, $\Phi_{(n)}(\mathcal{R}, \nabla\nabla Q)$, and $\nabla_\mu Q \Psi^{\mu\nu}_{(n)}(\mathcal{R}, \nabla\nabla Q) \nabla_\nu Q$. They are listed explicitly up to A_8 in [6].

4.7.3 Homogeneous Bundles over Symmetric Spaces

The exposition in this section closely follows our papers [5, 7, 12, 13]. Our goal is to compute the heat kernel of the Laplace-type operator $L = -\Delta + Q$ in the zero order of the low-energy approximation. The difference with the previous sections is that now we are going to do it on the most general covariantly constant background, that is, bundles with parallel curvature (that are called *homogeneous bundles*) on Riemannian manifolds with parallel curvature (that are called *symmetric spaces*).

It is well known that heat invariants are determined essentially by local geometry. They are polynomial invariants in the curvature with universal constants that do not depend on the global properties of the manifold [27]. It is in this universal structure that we are interested in this chapter. Our goal is to compute the heat kernel asymptotics of the Laplacian acting on homogeneous vector bundles over symmetric spaces.

In this section we will further assume that M is a *locally symmetric space* with a Riemannian metric with the parallel curvature

$$\nabla_\mu R_{\alpha\beta\gamma\delta} = 0, \tag{265}$$

which means, in particular, that the curvature satisfies the integrability constraints

$$R^{fg}{}_{ea} R^e{}_{bcd} - R^{fg}{}_{eb} R^e{}_{acd} + R^{fg}{}_{ec} R^e{}_{dab} - R^{fg}{}_{ed} R^e{}_{cab} = 0. \tag{266}$$

In the following we will also consider *homogeneous vector bundles* with parallel bundle curvature

$$\nabla_\mu \mathcal{R}_{\alpha\beta} = 0, \tag{267}$$

which means that the curvature satisfies the integrability constraints

$$[\mathcal{R}_{cd}, \mathcal{R}_{ab}] - R^f{}_{acd} \mathcal{R}_{fb} - R^f{}_{bcd} \mathcal{R}_{af} = 0. \tag{268}$$

Finally, we consider a parallel section Q of the endomorphism bundle End $(\mathcal{V})$, that is,

$$\nabla_\mu Q = 0, \tag{269}$$

which means that

$$[\mathcal{R}_{cd}, Q] = 0. \tag{270}$$

We will use normal coordinates defined above. Note that for symmetric spaces normal coordinates cover the whole manifold except for a set of measure zero where they become singular [19]. This set is precisely the set of points conjugate to the fixed point x' [where $\Delta^{-1}(x, x') = 0$] and of points that can be connected to the point x' by multiple geodesics. In any case, this set is a set of measure zero and, as we will show below, it can be dealt with by some regularization technique. Thus, we will use the normal coordinates defined above for the whole manifold.

4.7.3.1 Curvature Group of a Symmetric Space

We assumed that the manifold M is locally symmetric. Since we also assume that it is simply connected and complete, it is a globally symmetric space (or simply symmetric space). A symmetric space is said to be compact, noncompact, or Euclidean if all sectional curvatures are positive, negative, or zero. A generic symmetric space has the structure $M = M_0 \times M_s$, where $M_0 = \mathbb{R}^{n_0}$ and M_s is a semi-simple symmetric space; it is a product of a compact symmetric space M_+ and a noncompact symmetric space M_-, $M_s = M_+ \times M_-$. Of course, the dimensions must satisfy the relation $n_0 + n_s = n$, where $n_s = \dim M_s$.

Let Λ_2 be the vector space of two-forms on M at a fixed point x'. It has the dimension $\dim \Lambda_2 = n(n-1)/2$, and the inner product in Λ_2 is defined by

$$\langle X, Y\rangle = \frac{1}{2} X_{ab} Y^{ab}. \tag{271}$$

The Riemann curvature tensor naturally defines the curvature operator

$$\text{Riem} : \Lambda_2 \to \Lambda_2 \tag{272}$$

by

$$(\text{Riem}\, X)_{ab} = \frac{1}{2} R_{ab}{}^{cd} X_{cd}. \tag{273}$$

This operator is symmetric and has real eigenvalues which determine the principal sectional curvatures. Now, let Ker (Riem) and Im (Riem) be the kernel and the range of this operator, and

$$p = \dim \text{Im}(\text{Riem}) = \frac{n(n-1)}{2} - \dim \text{Ker}\,(\text{Riem}). \tag{274}$$

Further, let λ_i $(i = 1, \dots, p)$ be the non-zero eigenvalues, and $E^i{}_{ab}$ be the corresponding orthonormal eigen-two-forms. Then the components of the curvature tensor can be presented in the form [7]

$$R_{abcd} = \beta_{ik} E^i{}_{ab} E^k{}_{cd}, \tag{275}$$

where β_{ik} is a symmetric, in fact, diagonal, nondegenerate $p \times p$ matrix. Of course, the zero eigenvalues of the curvature operator correspond to the flat subspace M_0, the positive ones correspond to the compact submanifold M_+ and the negative ones to the noncompact submanifold M_-. Therefore, Im (Riem) $= T_x M_s$.

In the following the Latin indices from the middle of the alphabet will be used to denote tensors in Im(Riem); they should not be confused with the Latin indices from the beginning of the alphabet which denote tensors in M. They will be raised and lowered with the matrix β_{ik} and its inverse β^{ik}.

Next, we define the traceless $n \times n$ matrices $D_i = (D^a{}_{ib})$, where

$$D^a{}_{ib} = -\beta_{ik} E^k{}_{cb} \delta^{ca}. \tag{276}$$

The matrices D_i are known to be the generators of the holonomy algebra, $\mathcal{H}$, i.e., the Lie algebra of the restricted holonomy group, H,

$$[D_i, D_k] = F^j{}_{ik} D_j, \tag{277}$$

where $F^j{}_{ik}$ are the structure constants of the holonomy group. The structure constants of the holonomy group define the $p \times p$ matrices F_i, by $(F_i)^j{}_k = F^j{}_{ik}$, which generate the adjoint representation of the holonomy algebra

$$[F_i, F_k] = F^j{}_{ik} F_j. \tag{278}$$

For symmetric spaces the introduced quantities satisfy additional algebraic constraints. The most important consequence of (266) is the equation [7]

$$E^i{}_{ac} D^c{}_{kb} - E^i{}_{bc} D^c{}_{ka} = F^i{}_{kj} E^j{}_{ab}. \tag{279}$$

Now, we introduce a new type of indices, the capital Latin indices, $A, B, C, \ldots,$ which split according to $A = (a, i)$ and run from 1 to $N = p + n$. We define new quantities $C^A{}_{BC}$ by

$$C^i{}_{ab} = E^i{}_{ab}, \qquad C^a{}_{ib} = -C^a{}_{bi} = D^a{}_{ib}, \qquad C^i{}_{kl} = F^i{}_{kl}, \tag{280}$$

all other components being zero. Let us also introduce rectangular $p \times n$ matrices T_a by $(T_a)^j{}_c = E^j{}_{ac}$ and the $n \times p$ matrices $\bar{T}_a$ by $(\bar{T}_a)^b{}_i = -D^b{}_{ia}$. Then we can define $N \times N$ matrices $C_A = (C_a, C_i)$

$$C_a = \begin{pmatrix} 0 & \bar{T}_a \\ T_a & 0 \end{pmatrix}, \qquad C_i = \begin{pmatrix} D_i & 0 \\ 0 & F_i \end{pmatrix}, \tag{281}$$

so that $(C_A)^B{}_C = C^B{}_{AC}$.

Then one can prove the following [7]:

Theorem 1 *The matrices C_A generate the adjoint representation of a Lie algebra $\mathcal{G}$ with the structure constants $C^A{}_{BC}$, that is,*

$$[C_A, C_B] = C^C{}_{AB} C_C, \tag{282}$$

For the lack of a better name we call the algebra $\mathcal{G}$ the *curvature algebra*. As it will be clear from the next section it is a subalgebra of the total isometry algebra of the symmetric space. It should be clear that the holonomy algebra $\mathcal{H}$ is the subalgebra of the curvature algebra $\mathcal{G}$. The curvature algebra $\mathcal{G}$ is compact; it is a direct sum of two ideals, $\mathcal{G} = \mathcal{G}_0 \oplus \mathcal{G}_s$, an Abelian center $\mathcal{G}_0$ of dimension n_0 and a semi-simple algebra $\mathcal{G}_s$ of dimension $p + n_s$.

Next, we define a symmetric nondegenerate $N \times N$ matrix

$$(\gamma_{AB}) = \begin{pmatrix} \delta_{ab} & 0 \\ 0 & \beta_{ik} \end{pmatrix} . \tag{283}$$

This matrix and its inverse γ^{AB} will be used to lower and to raise the capital Latin indices.

4.7.3.2 Killing Vectors Fields

We will use extensively the isometries of the symmetric space M. We follow the approach developed in [7, 3, 10, 13]. The generators of isometries are the Killing vector fields ξ. The set of all Killing vector fields forms a representation of the isometry algebra, the Lie algebra of the isometry group of the manifold M. We define two subspaces of the isometry algebra. One subspace is formed by Killing vectors (called translations) satisfying the initial conditions $\nabla_\mu \xi^\nu\big|_{x=x'} = 0$ and another subspace is formed by Killing vectors (called rotations) satisfying the initial conditions $\xi^\nu\big|_{x=x'} = 0$.

One can easily show that a basis of translations can be chosen as

$$P_a = \left(\sqrt{K}\cot\sqrt{K}\right)^b{}_a \frac{\partial}{\partial y^b}, \tag{284}$$

where $K = (K^a{}_b)$ is a matrix defined by

$$K^a{}_b = R^a{}_{cbd} y^c y^d . \tag{285}$$

We can also show that the vector fields

$$L_i = -D^b{}_{ia} y^a \frac{\partial}{\partial y^b}, \tag{286}$$

define p linearly independent rotations. By adding the trivial Killing vectors for flat subspaces we find that the number of independent rotations is $p + n_0 n_s + n_0(n_0 - 1)/2$. We introduce the following notation $(\xi_A) = (P_a, L_i)$.

By using the explicit form of the Killing vector fields obtained above [7] one can prove the following theorem.

Theorem 2 *The Killing vector fields ξ_A form a representation of the curvature algebra $\mathcal{G}$*

$$[\xi_A, \xi_B] = C^C{}_{AB}\xi_C. \tag{287}$$

Note that they *do not* generate the complete isometry algebra of the symmetric space M. The curvature algebra $\mathcal{G}$ introduced in the previous section is a subalgebra of the total isometry algebra. It is clear that the Killing vector fields L_i form a representation of the holonomy algebra $\mathcal{H}$, which is the isotropy algebra of the semi-simple submanifold M_s, and a subalgebra of the total isotropy algebra of the symmetric space M.

4.7.3.3 Homogeneous Vector Bundles

Let $h^a{}_b$ be the projection to the subspace $T_x M_s$ of the tangent space and

$$q^a{}_b = \delta^a{}_b - h^a{}_b \tag{288}$$

be the projection tensor to the flat subspace $\mathbb{R}^{n_0}$. Since the curvature exists only in the semi-simple submanifold M_s, the components of the curvature tensor R_{abcd}, as well as the tensors $E^i{}_{ab}$, are non-zero only in the semi-simple subspace $T_x M_s$. Then

$$R_{abcd}q^a{}_e = R_{ab}q^a{}_e = E^i{}_{ab}q^a{}_e = D^a{}_{ib}q^b{}_e = D^a{}_{ib}q_a{}^e = 0. \tag{289}$$

Equation (268) imposes strong constraints on the curvature of the homogeneous bundle $\mathcal{W}$. We define

$$\mathcal{B}_{ab} = \mathcal{R}^{YM}_{cd}q^c{}_a q^d{}_b\,, \qquad \mathcal{E}_{ab} = \mathcal{R}^{YM}_{cd}h^c{}_b h^d{}_b, \tag{290}$$

so that

$$\mathcal{R}^{YM}_{ab} = \mathcal{E}_{ab} + \mathcal{B}_{ab}. \tag{291}$$

Then, from (268) we obtain

$$[\mathcal{B}_{ab}, \mathcal{B}_{cd}] = [\mathcal{B}_{ab}, \mathcal{E}_{cd}] = 0 \tag{292}$$

and

$$[\mathcal{E}_{cd}, \mathcal{E}_{ab}] - R^f{}_{acd}\mathcal{E}_{fb} - R^f{}_{bcd}\mathcal{E}_{af} = 0. \tag{293}$$

This means that $\mathcal{B}_{ab}$ takes values in an Abelian ideal of the gauge algebra $\mathcal{G}_{YM}$ and $\mathcal{E}_{ab}$ takes values in the holonomy algebra. More precisely, (293) is only possible if the holonomy algebra $\mathcal{H}$ is an ideal of the gauge algebra $\mathcal{G}_{YM}$. Thus, the gauge group G_{YM} must have a subgroup $Z \times H$, where Z is an Abelian group and H is the holonomy group.

Let X_{ab} be the generators of the orthogonal algebra $\mathcal{SO}(n)$ in some representation X. Then the matrices $T_i = -\frac{1}{2}D^a{}_{ib}X^b{}_a$ are the generators of the gauge algebra

$\mathcal{G}_{YM}$ realizing a representation T of the holonomy algebra $\mathcal{H}$. Next, we can show that the curvature of the homogeneous bundle $\mathcal{W}$ is given by

$$\mathcal{R}^{YM}_{ab} = -E^i{}_{ab}T_i + \mathcal{B}_{ab} = \frac{1}{2}R^{cd}{}_{ab}X_{cd} + \mathcal{B}_{ab}. \tag{294}$$

Now, we consider the representation Σ of the orthogonal algebra $\mathcal{SO}(n)$ defining the spin-tensor bundle $\mathcal{T}$ and define the matrices

$$G_{ab} = \Sigma_{ab} \otimes \mathbb{I}_X + \mathbb{I}_\Sigma \otimes X_{ab}. \tag{295}$$

Obviously, these matrices are the generators of the orthogonal algebra $\mathcal{SO}(n)$ in the product representation $\Sigma \otimes X$. Next, the matrices $Y_i = -\frac{1}{2}D^a{}_{ib}\Sigma^b{}_a$ form a representation Y of the holonomy algebra $\mathcal{H}$ and the matrices

$$\mathcal{R}_i = -\frac{1}{2}D^a{}_{ib}G^b{}_a \tag{296}$$

are the generators of the holonomy algebra in the product representation $\mathcal{R} = Y \otimes T$.

Then the total curvature, that is, the commutator of covariant derivatives (76) of a twisted spin-tensor bundle $\mathcal{V}$ is

$$\mathcal{R}_{ab} = -E^i{}_{ab}\mathcal{R}_i + \mathcal{B}_{ab} = \frac{1}{2}R^{cd}{}_{ab}G_{cd} + \mathcal{B}_{ab}. \tag{297}$$

4.7.3.4 Twisted Lie Derivatives

Let φ be a section of a twisted homogeneous spin-tensor bundle $\mathcal{T}$. Let ξ_A be the basis of Killing vector fields. Then the covariant (or generalized, or twisted) Lie derivative of φ along ξ_A is defined by

$$\mathcal{L}_A\varphi = \left(\xi_A{}^\mu\nabla_\mu + \frac{1}{2}\xi_A{}^a{}_{;b}G^b{}_a\right)\varphi\,. \tag{298}$$

One can prove the theorem [12, 13].

Theorem 3 *The operators $\mathcal{L}_A$ satisfy the commutation relations*

$$[\mathcal{L}_A, \mathcal{L}_B] = C^C{}_{AB}\mathcal{L}_C + \mathcal{B}_{AB}, \tag{299}$$

where

$$\mathcal{B}_{AB} = \begin{pmatrix} \mathcal{B}_{ab} & 0 \\ 0 & 0 \end{pmatrix}, \tag{300}$$

The details of the proof of this theorem (too long and technical to be presented in this chapter) are explained in the papers [12, 13] cited above. It uses extensively the properties of the Killing vector fields on symmetric spaces (in particular, Theorem 2) and the properties of the parallel curvature of the homogeneous vector bundle (some of them described above in Sect. 4.7.3.3).

The operators $\mathcal{L}_A$ form an algebra that is a direct sum of a nilpotent ideal and a semi-simple algebra. For the lack of a better name we call this algebra *gauged curvature algebra* and denote it by $\mathcal{G}_{\text{gauge}}$.

Now, let us define the operator

$$\mathcal{L}^2 = \gamma^{AB}\mathcal{L}_A\mathcal{L}_B \tag{301}$$

and the Casimir operator of the holonomy group

$$\mathcal{R}^2 = \frac{1}{4}R^{abcd}G_{ab}G_{cd}\,. \tag{302}$$

Then one can prove that [13]

Theorem 4 *The Laplacian Δ acting on sections of a twisted spin-tensor bundle $\mathcal{V}$ over a symmetric space has the form*

$$\Delta = \mathcal{L}^2 - \mathcal{R}^2. \tag{303}$$

This theorem is not new. It is a very well-known property of the Laplacian on homogeneous vector bundles. For more details and references see the paper [13] cited above.

4.7.3.5 Geometry of the Curvature Group

Let G_{gauge} be the gauged curvature group and H be its holonomy subgroup. Both these groups have compact algebras. However, while the holonomy group is always compact, the curvature group is, in general, a product of a nilpotent group G_0 and a semi-simple group G_s, $G_{\text{gauge}} = G_0 \times G_s$. The semi-simple group G_s is a product $G_s = G_+ \times G_-$ of a compact G_+ and a noncompact G_- subgroups.

Let ξ_A be the basis Killing vectors, k^A be the canonical coordinates on the curvature group G and $\xi(k) = k^A\xi_A$. The canonical coordinates are exactly the normal coordinates on the group defined above. Let C_A be the generators of the curvature group in adjoint representation and $C(k) = k^A C_A$.

Let $X = (X_A{}^M)$ be the matrix defined by

$$X = \frac{C(k)}{1-\exp[-C(k)]}\,, \tag{304}$$

and X_A be the vector fields on the group G defined by

$$X_A = X_A{}^M\frac{\partial}{\partial k^M}\,. \tag{305}$$

Then one can show that [13] the vector fields X_A form a representation of the curvature algebra $\mathcal{G}$

$$[X_A, X_B] = C^C{}_{AB} X_C\,. \tag{306}$$

The vector fields X_A are nothing but the right-invariant vector fields.

Since we will actually be working with the gauged curvature group, we introduce now the operators (*covariant right-invariant vector fields*) J_A by

$$J_A = X_A - \frac{1}{2}\mathcal{B}_{AB} k^B, \tag{307}$$

Then we show [13] that the operators J_A form the following algebra

$$[J_A, J_B] = C^C{}_{AB} J_C + \mathcal{B}_{AB}\,. \tag{308}$$

Thus, the operators J_A form a representation of the gauged curvature algebra $\mathcal{G}_{\text{gauge}}$.

Now, let $\mathcal{L}_A$ be the Lie derivatives and $\mathcal{L}(k) = k^A \mathcal{L}_A$. Then we find [13]

$$J_A \exp[\mathcal{L}(k)] = \exp[\mathcal{L}(k)]\mathcal{L}_A\,. \tag{309}$$

Note that J_A are first-order differential operators with respect to k^A, whereas $\mathcal{L}_A$ are first-order partial differential operators with respect to the coordinates x acting on sections of the bundle $\mathcal{V}$.

4.7.3.6 Heat Kernel on the Curvature Group

Now, let us define the operator

$$J^2 = \gamma^{AB} J_A J_B \tag{310}$$

and the invariant (scalar curvature of the curvature group)

$$R_G = -\frac{1}{4}\gamma^{AB} C^C{}_{AD} C^D{}_{BC}. \tag{311}$$

Then by using the properties of the right-invariant vector fields J_A one can find the heat kernel of the operator J^2 on the curvature group $\mathcal{G}$ [13].

Theorem 5 *Let $\Phi(t;k)$ be a function on the curvature group defined in canonical coordinates k^A by*

$$\Phi(t;k) = (4\pi t)^{-N/2}\left[\det{}_{TM}\left(\frac{\sinh[t\mathcal{B}]}{t\mathcal{B}}\right)\right]^{-1/2}\left[\det{}_{\mathcal{G}}\left(\frac{\sinh\left[C(k)/2\right]}{C(k)/2}\right)\right]^{-1/2}$$
$$\times\exp\left(-\frac{1}{4t}\langle k, \gamma t\mathcal{B}\coth(t\mathcal{B})k\rangle + \frac{1}{6}R_G t\right), \tag{312}$$

where $\langle u, \gamma v\rangle = \gamma_{AB}u^A v^B$ *is the inner product on the algebra* $\mathcal{G}$. *Then* $\Phi(t;k)$ *satisfies the heat equation*

$$\partial_t \Phi = J^2 \Phi, \tag{313}$$

and the initial condition

$$\Phi(0;k) = \gamma^{-1/2}\delta(k), \tag{314}$$

where $\gamma = \det \gamma_{AB}$.

In the following we will complexify the gauged curvature group in the following sense. We extend the canonical coordinates $(k^A) = (p^a, \omega^i)$ to the whole complex Euclidean space $\mathbb{C}^N$. Then all group-theoretic functions introduced above become analytic functions of k^A possibly with some poles on the real section $\mathbb{R}^N$ for compact groups. In fact, we replace the actual real slice $\mathbb{R}^N$ of $\mathbb{C}^N$ with an N-dimensional subspace $\mathbb{R}^N_{\rm reg}$ in $\mathbb{C}^N$ obtained by rotating the real section $\mathbb{R}^N$ counter-clockwise in $\mathbb{C}^N$ by $\pi/4$. That is, we replace each coordinate k^A by $e^{i\pi/4}k^A$. In the complex domain the group becomes noncompact. We call this procedure the *decompactification*. If the group is compact, or has a compact subgroup, then this plane will cover the original group infinitely many times.

Since the metric $(\gamma_{AB}) = \operatorname{diag}(\delta_{ab}, \beta_{ij})$ is not necessarily positive definite (actually, only the metric of the holonomy group β_{ij} is non-definite), we analytically continue the function $\Phi(t;k)$ in the complex plane of t with a cut along the negative imaginary axis so that $-\pi/2 < \arg t < 3\pi/2$. Thus, the function $\Phi(t;k)$ defines an analytic function of t and k^A. For the purpose of the following exposition we shall consider t to be *real negative*, $t < 0$. This is needed in order to make all integrals convergent and well defined and to be able to do the analytical continuation.

As we will show below, the singularities occur only in the holonomy group. This means that there is no need to complexify the coordinates p^a. Thus, in the following we assume the coordinates p^a to be real and the coordinates ω^i to be complex, more precisely, to take values in the p-dimensional subspace $\mathbb{R}^p_{\rm reg}$ of $\mathbb{C}^p$ obtained by rotating $\mathbb{R}^p$ counter-clockwise by $\pi/4$ in $\mathbb{C}^p$. That is, we have $\mathbb{R}^N_{\rm reg} = \mathbb{R}^n \times \mathbb{R}^p_{\rm reg}$.

This procedure (that we call a regularization) with the nonstandard contour of integration is necessary for the convergence of the integrals below since we are treating both the compact and the noncompact symmetric spaces simultaneously. Remember, that, in general, the nondegenerate diagonal matrix β_{ij} is not positive-definite. The space $\mathbb{R}^p_{\rm reg}$ is chosen in such a way to make the Gaussian exponent purely imaginary. Then the indefiniteness of the matrix β does not cause any problems. Moreover, the integrand does not have any singularities on these contours. The convergence of the integral is guaranteed by the exponential growth of the sine for imaginary argument. These integrals can be computed then in the following way. The coordinates ω^j corresponding to the compact directions are rotated further by another $\pi/4$ to imaginary axis and the coordinates ω^j corresponding to the non-compact directions are rotated back to the real axis. Then, for $t < 0$ all the integrals below are well defined and convergent and define an analytic function of t in a complex plane with a cut along the negative imaginary axis.

4.7.3.7 Heat Trace

Now, by using the heat kernel (312) of the operator J^2 on the curvature group obtained above, the relation (303) of the Laplacian and the operator $\mathcal{L}^2$, and the property (309) one can find the following integral representation of the heat semigroup of the Laplace-type operator [13].

Theorem 6 *Let $L = -\Delta + Q$ be the Laplace-type operator acting on sections of a homogeneous twisted spin-tensor vector bundle over a symmetric space. Then the heat semigroup* $\exp(-tL)$ *can be represented in form of an integral*

$$\begin{aligned}\exp(-tL) &= (4\pi t)^{-N/2}\left[\det{}_{TM}\left(\frac{\sinh(t\mathcal{B})}{t\mathcal{B}}\right)\right]^{-1/2}\exp\left(-tQ - t\mathcal{R}^2 + \frac{1}{6}R_G t\right)\\ &\quad\times\int\limits_{\mathbb{R}^N_{\text{reg}}} dk\ \gamma^{1/2}\left[\det{}_{\mathcal{G}}\left(\frac{\sinh[C(k)/2]}{C(k)/2}\right)\right]^{1/2}\\ &\quad\times\exp\left\{-\frac{1}{4t}\langle k, \gamma t\mathcal{B}\coth(t\mathcal{B})k\rangle\right\}\exp[\mathcal{L}(k)]. \end{aligned}\tag{315}$$

The heat trace can be obtained by acting by the heat semigroup $\exp(-tL)$ on the delta-function. To be able to use this integral representation we need to compute the action of the isometries $\exp[\mathcal{L}(k)]$ on the delta-function.

Let ω^i be the canonical coordinates on the holonomy group H and $(k^A) = (p^a, \omega^i)$ be the natural splitting of the canonical coordinates on the curvature group G. Then we can prove that [13]

$$\left[\exp[\mathcal{L}(k)]\delta(x,x')\right]^{\text{diag}} = \left[\det{}_{TM}\left(\frac{\sinh[D(\omega)/2]}{D(\omega)/2}\right)\right]^{-1}\exp[\mathcal{R}(\omega)]\delta(p)\,, \tag{316}$$

where $D(\omega) = \omega^i D_i$ and $\mathcal{R}(\omega) = \omega^i\mathcal{R}_i$.

We implicitly assumed that there are no closed geodesics and that the equation of closed orbits of isometries has a unique solution. This is indeed the case on noncompact symmetric spaces (with negative curvature). On compact symmetric spaces this is not true, there are infinitely many closed geodesics and infinitely many closed orbits of isometries. The results for compact symmetric spaces (with positive curvature) can be obtained by an analytic continuation from the dual noncompact case [19]. That is why we proposed above to complexify our holonomy group. If the coordinates ω^i are complex, taking values in the subspace $\mathbb{R}^p_{\text{reg}}$ defined above, then the equation of closed orbits should have a unique trivial solution and the Jacobian is an analytic function. It is worth stressing once again that the regularized canonical coordinates cover the whole group except for a set of measure zero.

Now by using the above lemmas and the theorem we can compute the heat trace. We define the invariant (scalar curvature of the holonomy group)

$$R_H = -\frac{1}{4}\beta^{ij}F^k{}_{il}F^l{}_{jk}. \tag{317}$$

Theorem 7 *The heat trace of the operator L has the form*

$$\begin{aligned}\Theta(t) = \ & (4\pi t)^{-n/2}\int\limits_M d\text{vol } \operatorname{tr}_V\left[\det{}_{TM}\left(\frac{\sinh(t\mathcal{B})}{t\mathcal{B}}\right)\right]^{-1/2} \\ & \exp\left\{\left(\frac{1}{8}R + \frac{1}{6}R_H - \mathcal{R}^2 - Q\right)t\right\} \\ & \times \int\limits_{\mathbb{R}^n_{\text{reg}}} \frac{d\omega}{(4\pi t)^{p/2}}\,\beta^{1/2}\exp\left\{-\frac{1}{4t}\langle\omega,\beta\omega\rangle\right\}\cosh\left[\mathcal{R}(\omega)\right] \\ & \times\left[\det{}_{\mathcal{H}}\left(\frac{\sinh\left[F(\omega)/2\right]}{F(\omega)/2}\right)\right]^{1/2}\left[\det{}_{TM}\left(\frac{\sinh\left[D(\omega)/2\right]}{D(\omega)/2}\right)\right]^{-1/2},\end{aligned} \tag{318}$$

where $\beta = \det\beta_{ij}$, $\langle\omega,\beta\omega\rangle = \beta_{ij}\omega^i\omega^j$, *and* $F(\omega) = \omega^i F_i$.

One should also stress that these global solutions do not affect the asymptotics of the heat kernel. So, if all we want is the heat kernel asymptotics then one can use our results for any symmetric space even without regularization.

This equation can be used now to generate all heat kernel coefficients A_k for any locally symmetric space simply by expanding it in a power series in t. By using the standard Gaussian averages one can obtain now all heat kernel coefficients in terms of traces of various contractions of the matrices $D^a{}_{ib}$ and $F^j{}_{ik}$ with the matrix β^{ik}. All these quantities are curvature invariants and can be expressed directly in terms of the Riemann tensor.

4.8 Low-Energy Effective Action in Quantum General Relativity

We can apply now the obtained results for the heat trace to compute the low-energy one-loop effective action in quantum general relativity given by (70). In the Euclidean formulation we have

$$\Gamma_{(1)} = \frac{1}{2}\left(\log\operatorname{Det}\hat{L} - 2\log\operatorname{Det}F\right), \tag{319}$$

which, in the zeta-regularization, takes the form

$$\Gamma_{(1)} = -\frac{1}{2}\left(\zeta'_{\hat{L}}(0) - 2\zeta'_F(0)\right), \tag{320}$$

where $\zeta_{\hat{L}}(s)$ and $\zeta_F(s)$ are the zeta-functions of the graviton operator $\hat{L}$ and the ghost operator F. Now, let us define the total zeta-function by

$$\zeta_{GR}(s) = \zeta_{\hat{L}}(s) - 2\zeta_F(s). \tag{321}$$

Then the effective action is

$$\Gamma_{(1)} = -\frac{1}{2}\zeta'_{GR}(0). \tag{322}$$

Next, by using the definition of the zeta-function we obtain

$$\zeta_{GR}(s) = \frac{\mu^{2s}}{\Gamma(s)} \int_0^\infty dt\; t^{s-1} e^{t\lambda} \Theta_{GR}(t), \tag{323}$$

where

$$\Theta_{GR}(t) = \Theta_{\hat{L}}(t) - 2\Theta_F(t), \tag{324}$$

μ is a renormalization parameter introduced to preserve dimensions and $\Theta_{\hat{L}}(t)$ and $\Theta_F(t)$ are the heat traces of the operators $\hat{L}$ and F. Here λ is a sufficiently large negative infrared cutoff parameter introduced to regularize any infrared divergences, which are present if the operators $\hat{L}$ and F have negative modes. The parameter λ should be set to zero at the end of the calculations.

Now, notice that both operators $\hat{L}$ and F are of Laplace type, that is, $-\Delta + Q$, acting on pure tensor bundles; so, there is no Yang – Mills group here, $\tilde{\mathcal{R}}_{ab} = \mathcal{E}_{ab} = \mathcal{B}_{ab} = 0$. The operator $\hat{L}$ acts on the bundle $\mathcal{T}_{(2)} = T^*M \otimes T^*M$ of symmetric two-tensors and the operator F acts on sections of the tangent bundle $\mathcal{T}_{(1)} = TM$. The potentials, Q, for both operators are obviously read off from their definition

$$\left(Q_{(1)}\right)^a{}_b = -R^a{}_b\,, \tag{325}$$

$$\begin{aligned}\left(Q_{(2)}\right)_{cd}{}^{ab} = {} & -2R^{(a}{}_c{}^{b)}{}_d - 2\delta^{(a}{}_{(c}R^{b)}{}_{d)} + R_{cd}g^{ab} + \frac{2}{n-2}g_{cd}R^{ab} \\ & -\frac{1}{(n-2)}g_{cd}g^{ab}R + \delta^a{}_{(c}\delta^b{}_{d)}(R - 2\Lambda).\end{aligned} \tag{326}$$

The generators of the orthogonal group $SO(n)$ in the vector representation are

$$\left(\Sigma_{(1)ab}\right)^c{}_d = 2\delta^c{}_{[a}g_{b]d}. \tag{327}$$

The generators of the orthogonal group $SO(n)$ in the symmetric two-tensor representation are

$$\left(\Sigma_{(2)ab}\right)_{cd}{}^{ef} = -4\delta^{(e}{}_{[a}g_{b](d}\delta^{f)}{}_{c)}. \tag{328}$$

The generators of the holonomy group are

$$\mathcal{R}_{(1)i} = D_i, \tag{329}$$

and

$$\mathcal{R}_{(2)i} = -2D_i \vee I_{(1)}, \tag{330}$$

which, in component language, reads

$$\left(\mathcal{R}_{(1)i}\right)^a{}_b = D^a{}_{ib} \tag{331}$$

and

$$\left(\mathcal{R}_{(2)i}\right)_{cd}{}^{ab} = -2D^{(a}{}_{i(d}\delta^{b)}{}_{c)}. \tag{332}$$

The Casimir operators are

$$\left(\mathcal{R}^2_{(1)}\right)^a{}_b = -R^a{}_b\,, \tag{333}$$

and

$$\left(\mathcal{R}^2_{(2)}\right)_{cd}{}^{ab} = 2R^{(a}{}_d{}^{b)}{}_c - 2\delta^{(a}{}_{(c}R^{b)}{}_{d)}. \tag{334}$$

By using the results for the heat traces described above we obtain the total heat trace

$$\Theta_{GR}(t) = (4\pi t)^{-n/2} \int\limits_M d\text{vol}\;\; \exp\left\{\left(\frac{1}{8}R + \frac{1}{6}R_H\right)t\right\} \tag{335}$$

$$\times \int\limits_{\mathbb{R}^n_{\text{reg}}} \frac{d\omega}{(4\pi t)^{p/2}}\, \beta^{1/2} \exp\left\{-\frac{1}{4t}\langle\omega, \beta\omega\rangle\right\}\Psi(t;\omega)$$

$$\times \left[\det{}_{\mathcal{H}}\left(\frac{\sinh\left[F(\omega)/2\right]}{F(\omega)/2}\right)\right]^{1/2}\left[\det{}_{TM}\left(\frac{\sinh\left[D(\omega)/2\right]}{D(\omega)/2}\right)\right]^{-1/2}, \tag{336}$$

where

$$\begin{aligned}\Psi(t;\omega) = {} & \exp\left[-t(R-2\Lambda)\right]\operatorname{tr}_{\mathcal{T}_{(2)}}\exp\left(tV_{(2)}\right)\cosh\left[2D(\omega)\vee I_{(1)}\right] \\ & -2\operatorname{tr}_{TM}\exp\left(tV_{(1)}\right)\cosh\left[\,D(\omega)\right]\,,\end{aligned} \tag{337}$$

and the matrices $V_{(1)}$ and $V_{(2)}$ are defined by

$$\left(V_{(1)}\right)^{a}{}_{b} = 2R^{a}{}_{b}, \tag{338}$$

$$\left(V_{(2)}\right)_{cd}{}^{ab} = 4\delta^{(a}{}_{(c}R^{b)}{}_{d)} - R_{cd}g^{ab} - \frac{2}{n-2}g_{cd}R^{ab} + \frac{1}{(n-2)}g_{cd}g^{ab}R. \tag{339}$$

One can go further and compute the function $\Psi(t;\omega)$ by finding the eigenvalues of the endomorphisms $V_{(1)}$ and $V_{(2)}$. However, we will not do it here and leave the answer in the form (337). By using the obtained heat trace one can compute now the zeta-function and then the effective action. We would like to stress two points here. First of all, quantum general relativity is a non-renormalizable theory. Therefeore, even if one gets a final result via the zeta-regularization one should not take it too seriously. Second, our results for the heat kernel and, hence, for the effective action are essentially non-perturbative. They contain an infinite series of Feynmann diagrams and cannot be obtained in any perturbation theory. One could try now to use this result for the analysis of the ground state in quantum gravity. But this is a rather ambitious program for the future.

Acknowledgements I would like to thank the organizers of the summer school "New Paths Towards Quantum Gravity" and the workshop "Quantum Gravity: An Assessment," Bernhelm Booß-Bavnbek, Giampiero Esposito, and Matthias Lesch, for their kind invitation and for the financial support.

References

1. I.G. Avramidi, *The nonlocal structure of one-loop effective action via partial summation of asymptotic expansion*. Phys. Lett. B **236** (1990a), 443–449.
2. I.G. Avramidi, *The covariant technique for calculation of the heat kernel asymptotic expansion*. Phys. Lett. B **238** (1990b), 92–97.
3. I.G. Avramidi, *A covariant technique for the calculation of the one-loop effective action*. Nucl. Phys. B **355** (1991), 712–754; Erratum: Nucl. Phys. B **509** (1998), 557–558.
4. I.G. Avramidi, *A new algebraic approach for calculating the heat kernel in gauge theories*. Phys. Lett. B **305** (1993), 27–34.
5. I.G. Avramidi, *The heat kernel on symmetric spaces via integrating over the group of isometries*. Phys. Lett. B **336** (1994), 171–177.
6. I.G. Avramidi, *Covariant algebraic method for calculation of the low-energy heat kernel*. J. Math. Phys. **36** (1995), 5055–5070; Erratum: J. Math. Phys. **39** (1998), 1720.
7. I.G. Avramidi, *A new algebraic approach for calculating the heat kernel in quantum gravity*. J. Math. Phys. **37** (1996), 374–394.
8. I.G. Avramidi, *Green functions of higher-order differential operators*. J. Math. Phys. **39** (1998), 2889–2909.
9. I.G. Avramidi, *Covariant techniques for computation of the heat kernel*. Rev. Math. Phys. **11** (1999), 947–980.
10. I.G. Avramidi, *Heat kernel and quantum gravity*. Springer, Berlin (2000).
11. I.G. Avramidi, *Heat kernel in quantum field theory*. Nucl. Phys. Proc. Suppl. **104** (2002), 3–32.
12. I.G. Avramidi, *Heat kernel on homogeneous bundles*. Int. J. Geom. Methods Mod. Phys. **5** (2008), 1–23.

13. I.G. Avramidi, *Heat kernel on homogeneous bundles over symmetric spaces*. Comm. Math. Phys. **288** (2009), 963–1006.
14. A.O. Barvinsky, G.A. Vilkovisky, *The generalized Schwinger – De Witt technique in gauge theories and quantum gravity*. Phys. Rep. C **119** (1985), 1–74.
15. N. Berline, E. Getzler, M. Vergne, *Heat kernels and Dirac operators*. Springer, Berlin (1992).
16. N.D. Birrel, P.C.W. Davies, *Quantum fields in curved space*. Cambridge University Press, Cambridge (1982).
17. T. Branson, P.B. Gilkey, B. Ørsted, *Leading terms in the heat invariants*. Proc. Am. Math. Soc. **109** (1990), 437–450.
18. I.L. Buchbinder, S.D. Odintsov, I.L. Shapiro, *Effective action in quantum gravity*. IOP Publishing, Bristol (1992).
19. R. Camporesi, *Harmonic analysis and propagators on homogeneous spaces*. Phys. Rep. **196** (1990), 1–134.
20. B.S. De Witt, *Quantum theory of gravity II. The manifestly covariant theory*. Phys. Rev. **162** (1967), 1195–1238.
21. B.S. De Witt, *Quantum theory of gravity III. Applications of the covariant theory*. Phys. Rev. **162** (1967), 1239–1256.
22. B.S. De Witt, *Quantum field theory in curved spacetime*. Phys. Rep. C **19** (1975), 296–357.
23. B.S. De Witt, "Quantum gravity: New synthesis", in: S. Hawking, and W. Israel, (eds.), *General relativity*. Cambridge University Press, Cambridge (1979).
24. B.S. De Witt, *The global approach to quantum field theory*. Oxford University Press, Oxford (2003).
25. V.A. Fock, *The proper time in classical and quantum mechanics*. Izv. USSR Acad. Sci. Phys. **4–5** (1937), 551–568.
26. S.A. Fulling, *Aspects of quantum field theory in curved space-time*. Cambridge University Press, Cambridge (1989).
27. P.B. Gilkey, *Invariance theory, the heat equation, and the atiyah-singer index theorem*. CRC Press, Boca Raton (1995).
28. K. Kirsten, *Spectral functions in mathematics and physics*. CRC Press, Boca Raton (2001).
29. S. Minakshisundaram, A. Pleijel, *Some properties of the eigenfunctions of the Laplace operator on Riemannian manifolds*. Can. J. Math. **1** (1949), 242–256.
30. J.S. Schwinger, *On gauge invariance and vacuum polarization*. Phys. Rev. **82** (1951), 664–679.
31. J.S. Schwinger, *The theory of quantized fields V.* Phys. Rev. **93** (1954), 615–628.
32. R.T. Seeley, *Complex powers of an elliptic operator*. Proc. Symp. Pure Math. **10** (1967), 288–307.
33. A.E. Van de Ven, *Index-free heat kernel coefficients*. Class. Quant. Grav. **15** (1998), 2311–2344.
34. D.V. Vassilevich, *Heat kernel expansion: User's manual*. Phys. Rep., **388** (2003), 279–360.
35. S. Yajima, Y. Higasida, K. Kawano, S.-I. Kubota, Y. Kamo, S. Tokuo, *Higher coefficients in asymptotic expansion of the heat kernel*. Phys. Rep. Kumamoto Univ. **12** (2004), 39–62.

Chapter 5
Lectures on Cohomology, T-Duality, and Generalized Geometry

P. Bouwknegt

Abstract These are notes for lectures, originally entitled "Selected Mathematical Aspects of Modern Quantum Field Theory", presented at the Summer School "New Paths Towards Quantum Gravity", Holbæk, Denmark, 10–16 May 2008. My aim for these lectures was to introduce a mixture of physics and mathematics postgraduate students into a selection of exciting new developments on the interface of mathematics and quantum field theory. This write-up covers three topics: (1) cohomology and differential characters, (2) T-duality, and (3) generalized geometry. The three chapters can be read, more or less, independent of each other, but there is a common central theme, namely the occurrence of a (local) 2-form gauge field in certain quantum fields theories, the so-called B-field, which plays a role analogous to the electromagnetic gauge field.

The notes are suitable for beginning postgraduate students in mathematical physics with some background in differential geometry and algebraic topology, but some sections may need a slightly more sophisticated background. I hope these notes fill a gap between undergraduate coursework and current research at the cutting edge of the field. The notes certainly do not offer an exhaustive discussion of the topics mentioned above, but rather serve as an introduction after which the reader should feel comfortable to study research papers in these areas.

5.1 Cohomology and Differential Characters

In this section we will give a basic introduction to several cohomology theories and differential characters. We will be guided by the example of electromagnetism. A basic reference for the first part of this chapter is the textbook by Bott and Tu [4].

P. Bouwknegt (✉)
Department of Theoretical Physics, Research School of Physics and Engineering
Department of Mathematics, Mathematical Sciences Institute, Australian National University, Canberra, ACT 0200, Australia, `peter.bouwknegt@anu.edu.au`

Bouwknegt, P.: *Lectures on Cohomology, T-Duality, and Generalized Geometry*. Lect. Notes Phys. **807**, 261–311 (2010)
DOI 10.1007/978-3-642-11897-5_5

5.1.1 A Brief Review of de Rham and Čech Cohomology

5.1.1.1 Open Covers

The proper language for Čech cohomology is that of sheaves. For our purposes it suffices to work with good open covers (i.e. a cover such that each set and multiple intersections are contractible), as one can show that the definition of Čech cohomology does not depend on the choice of good open cover.

Let $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{I}}$ be a good open cover of X (we take X to be paracompact, so that it admits a partition of unity). Denote $U_{\alpha_0\ldots\alpha_q} = U_{\alpha_0} \cap \cdots \cap U_{\alpha_q}$.

Now, let $Y = Y_{\mathcal{U}} = \coprod_{\alpha\in\mathcal{I}} U_\alpha$ be the *nerve* of the open cover $\mathcal{U}$. Then we have a submersion $\pi : Y \to X$. Much of what follows remains true for general submersions $\pi : Y \to X$ (which is useful in the theory of bundle gerbes).

Multiple intersections can be written as fibred products, i.e.

$$\begin{aligned}
Y &= \{(x,\alpha) \in X \times \mathcal{I} \mid x \in U_\alpha\},\\
Y^{[2]} &= Y \times_X Y = \{(x,y,\alpha,\beta) \mid \pi(x) = \pi(y)\}\\
&= \{(x,\alpha,\beta) \mid x \in U_{\alpha\beta}\} = \coprod_{\alpha,\beta} U_{\alpha\beta},\\
Y^{[q+1]} &= \underbrace{Y \times_X Y \times_X \ldots \times_X Y}_{q+1} = \coprod_{\alpha_0,\ldots,\alpha_q} U_{\alpha_0\ldots\alpha_q}.
\end{aligned}$$

5.1.1.2 De Rham Cohomology

We denote by $\Omega^p(X) = \Gamma(\wedge^p T^*X)$ the set of (smooth) differential forms on X of degree p, by $\Omega^p_{\mathrm{cl}}(X)$ the set of closed forms, and by $\Omega^p(X)_{\mathbb{Z}}$ the set of forms with integral periods. Similarly, by $\Omega^{(p,q)} = \Omega^p(Y^{[q+1]})$, $p, q \geq 0$, we denote p-forms on $Y^{[q+1]}$, i.e. an $\omega \in \Omega^p(Y^{[q+1]})$ is a collection of p-forms $\omega_{\alpha_0\ldots\alpha_q} \in \Omega^p(U_{\alpha_0\ldots\alpha_q})$. For convenience we let $\Omega^{(p,-1)} = \Omega^p(Y^{[0]}) = \Omega^p(X)$.

The cohomology of the exterior derivative $d : \Omega^{(p,q)} \to \Omega^{(p+1,q)}$ is denoted by $H^p_{\mathrm{dR}}(Y^{[q+1]})$ and known as the *de Rham cohomology* of $Y^{[q+1]}$. In particular $H^p_{\mathrm{dR}}(M)$ denotes the cohomology of $d : \Omega^p(X) \to \Omega^{p+1}(X)$, the de Rham cohomology of X. Since each $U_{\alpha_0\ldots\alpha_q}$ is contractible we have by the Poincaré lemma

Theorem 1 *The following complex is exact for all $q \geq 0$,*

$$0 \longrightarrow \Omega^{(-1,q)} \longrightarrow \Omega^{(0,q)} \xrightarrow{d} \Omega^{(1,q)} \xrightarrow{d} \Omega^{(2,q)} \xrightarrow{d} \cdots$$

where

$$\Omega^{(-1,q)} \equiv \mathrm{Ker}(d : \Omega^{(0,q)} \to \Omega^{(1,q)}) = C(Y^{[q+1]}, \mathbb{R}).$$

5.1.1.3 Čech Cohomology

For G an abelian group ($\mathbb{R}$, $\underline{\mathbb{R}}$, Ω^q, etc.[1]) we define Čech cochains by $\check{C}^q(X,\mathsf{G}) = C(Y^{[q+1]},\mathsf{G})$, i.e. collections of smooth maps $\omega_{\alpha_0\ldots\alpha_q} : U_{\alpha_0\ldots\alpha_q} \to \mathsf{G}$, where we assume the maps $\omega_{\alpha_0\ldots\alpha_q}$ to be antisymmetric in their indices. The Čech differential $\delta : \check{C}^q(X,\mathsf{G}) \to \check{C}^{q+1}(X,\mathsf{G})$ is defined by

$$(\delta\omega)_{\alpha_0\ldots\alpha_{q+1}} = \sum_i (-1)^i \omega_{\alpha_0\ldots\widehat{\alpha}_i\ldots\alpha_{q+1}}\big|_{U_{\alpha_0\ldots\alpha_{q+1}}}, \tag{1}$$

where a hat over an index means omission of the index. It satisfies $\delta^2 = 0$. The cohomology of δ is denoted by $\check{H}^q(X,\mathsf{G})$ and known as the *Čech cohomology* of X with coefficients in G. Specializing to $\mathsf{G} = \Omega^p$, i.e. $\check{C}^q(X,\mathsf{G}) = \Omega^p(Y^{[q+1]})$, we have

Theorem 2 *The following complex is exact for all $p \geq 0$,*

$$0 \longrightarrow \Omega^{(p,-1)} \longrightarrow \Omega^{(p,0)} \xrightarrow{\delta} \Omega^{(p,1)} \xrightarrow{\delta} \Omega^{(p,2)} \xrightarrow{\delta} \cdots$$

where

$$\Omega^{(p,-1)} \equiv \mathrm{Ker}(\delta : \Omega^{(p,0)} \to \Omega^{(p,1)}) = \Omega^p(X).$$

There are two important exact sequences arising from a change in coefficients. The first one arises from

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\iota} \mathbb{R} \xrightarrow{e^{2\pi i\cdot}} \mathbb{R}/\mathbb{Z} \longrightarrow 0$$

and is given by

$$\cdots \longrightarrow \check{H}^p(X,\mathbb{Z}) \longrightarrow \check{H}^p(X,\mathbb{R}) \longrightarrow \check{H}^p(X,\mathbb{R}/\mathbb{Z}) \longrightarrow \check{H}^{p+1}(X,\mathbb{Z}) \longrightarrow \cdots. \tag{2}$$

The second one arises from

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\iota} \underline{\mathbb{R}} \xrightarrow{e^{2\pi i\cdot}} \underline{\mathbb{R}/\mathbb{Z}} \longrightarrow 0$$

and is given by

$$\cdots \longrightarrow \check{H}^p(X,\mathbb{Z}) \longrightarrow \check{H}^p(X,\underline{\mathbb{R}}) \longrightarrow \check{H}^p(X,\underline{\mathbb{R}/\mathbb{Z}}) \longrightarrow \check{H}^{p+1}(X,\mathbb{Z}) \longrightarrow \cdots \tag{3}$$

[1] Here $\underline{\mathbb{R}}$ denotes the sheaf of germs of $\mathbb{R}$-valued functions on a manifold X.

which, by using $\check{H}^p(M, \underline{\mathbb{R}}) = 0$, for $p > 0$, (see Theorem 1 for $p = 0$), leads to

$$H^p(X, \underline{\mathbb{R}/\mathbb{Z}}) \cong H^{p+1}(X, \mathbb{Z}), \qquad p > 0.$$

5.1.2 Electromagnetism

5.1.2.1 Curvature and Connection

We now identify some of the previously discussed de Rham and Čech cocycles in Maxwell's theory of electromagnetism in $D = 4$ dimensions. Electromagnetic fields are encoded in a field strength ('curvature') $F \in \Omega^2(X)$, satisfying the Maxwell and Bianchi equations

$$\begin{aligned} d(\star F) &= \star J \qquad \text{(Maxwell)}, \\ dF &= \star\widetilde{J} \qquad \text{(Bianchi)}, \end{aligned} \tag{4}$$

where J and $\widetilde{J}$ are the so-called electric and magnetic current 1-forms. The electric and magnetic charges are given by

$$\begin{aligned} e &= \int_{B^3} (\star J) = \int_{S^2} \star F, \\ g &= \int_{B^3} (\star\widetilde{J}) = \int_{S^2} F, \end{aligned} \tag{5}$$

where B^3 is a 3-ball containing the charge (could be of infinite extension) and $S^2 = \partial B^3$.[2]

Now, consider a good open cover $\mathcal{U} = \{U_\alpha\}$ of X. Locally, on U_α, we have

$$F = dA_\alpha\,, \qquad A_\alpha \in \Omega^1(U_\alpha). \tag{6}$$

On overlaps $U_{\alpha\beta}$ the gauge potentials ('connections') A_α are related by gauge transformations

$$(\delta A)_{\alpha\beta} \equiv A_\beta - A_\alpha = -d\Lambda_{\alpha\beta}\,, \qquad \Lambda_{\alpha\beta} \in \Omega^0(U_{\alpha\beta}). \tag{7}$$

There is a subtlety here. Both the curvature F and the connection A_α take values in the Lie algebra $\mathfrak{g} = \mathbb{R}$. However, $\mathbb{R}$ is both the Lie algebra corresponding to the noncompact group $\mathsf{G} = \mathbb{R}$ or the compact group $\mathsf{G} = \mathsf{U}(1) \cong \mathbb{R}/\mathbb{Z}$. In the first case we require

$$(\delta\Lambda)_{\alpha\beta\gamma} = \Lambda_{\beta\gamma} - \Lambda_{\alpha\gamma} + \Lambda_{\alpha\beta} = 0, \tag{8}$$

[2] It should be remarked that we have not yet observed configurations with magnetic charge $g \neq 0$ (magnetic monopoles) in nature.

while in the second case it would be more natural to write the gauge transformations in a multiplicative form, defining $g_{\alpha\beta} = \exp(2\pi i \Lambda_{\alpha\beta}) \in \mathsf{U}(1)$,

$$\begin{aligned}(\delta A)_{\alpha\beta} &= A_\beta - A_\alpha = -\frac{1}{2\pi i} g_{\alpha\beta}^{-1} d g_{\alpha\beta} = -\frac{1}{2\pi i} d \log g_{\alpha\beta}, \\ (\delta g)_{\alpha\beta\gamma} &= g_{\beta\gamma} g_{\alpha\gamma}^{-1} g_{\alpha\beta} = 1,\end{aligned} \tag{9}$$

or, in terms of $\Lambda_{\alpha\beta}$,

$$\begin{aligned}(\delta A)_{\alpha\beta} &= A_\beta - A_\alpha = -d\Lambda_{\alpha\beta}, \\ (\delta\Lambda)_{\alpha\beta\gamma} &= \Lambda_{\beta\gamma} - \Lambda_{\alpha\gamma} + \Lambda_{\alpha\beta} = n_{\alpha\beta\gamma}, \\ (\delta n)_{\alpha\beta\gamma\delta} &= n_{\beta\gamma\delta} - n_{\alpha\gamma\delta} + n_{\alpha\beta\delta} - n_{\alpha\beta\gamma} = 0,\end{aligned} \tag{10}$$

where $n_{\alpha\beta\gamma} \in \mathbb{Z}$.

Geometrically, the $\{g_{\alpha\beta}\}$ define the transitions functions of a principal $\mathsf{U}(1)$ bundle P over X, while $\{A_\alpha\}$ defines a connection on P.

5.1.2.2 Gauge Transformations

Under a gauge transformation $h_\alpha \in \Omega^0(U_\alpha)$, we have

$$\begin{aligned}A'_\alpha &= A_\alpha + \frac{1}{2\pi i} d \log h_\alpha = A_\alpha + \widetilde{d} h_\alpha \,, \\ g'_{\alpha\beta} &= h_\alpha g_{\alpha\beta} h_\beta^{-1} \,,\end{aligned} \tag{11}$$

or, in terms of $h_\alpha = \exp(2\pi i b_\alpha)$,

$$\begin{aligned}A'_\alpha &= A_\alpha + d b_\alpha \,, \\ \Lambda'_{\alpha\beta} &= \Lambda_{\alpha\beta} + b_\alpha - b_\beta + m_{\alpha\beta} \,, \\ n'_{\alpha\beta\gamma} &= n_{\alpha\beta\gamma} + m_{\beta\gamma} - m_{\alpha\gamma} + m_{\alpha\beta} \,,\end{aligned} \tag{12}$$

for some set of integers $m_{\alpha\beta} \in \mathbb{Z}$.

Thus, in particular, we see that $\{g_{\alpha\beta}\} \in \check{H}^1(X, \underline{\mathbb{R}/\mathbb{Z}})$ and $\{n_{\alpha\beta\gamma}\} \in \check{H}^2(X, \mathbb{Z})$ is the image of $\{g_{\alpha\beta}\}$ under the isomorphism $H^1(X, \underline{\mathbb{R}/\mathbb{Z}}) \to \check{H}^2(X, \mathbb{Z})$. Also $[F] \in H^2_{\mathrm{dR}}(X)$ is the image of $\{n_{\alpha\beta\gamma}\} \in \check{H}^2(X, \mathbb{Z})$ into $\check{H}^2(X, \underline{\mathbb{R}}) \cong H^2_{\mathrm{dR}}(X)$.

Equations (9), (10), (11), (12) can be pictorially represented in the following diagram in which the de Rham differential runs horizontally and the Čech differential vertically.

$$\begin{tikzcd}
n_{\alpha\beta\gamma} \arrow[r] & \star & & \\
m_{\alpha\beta} \arrow[u, dotted] \arrow[r, dotted] & \Lambda_{\alpha\beta} \arrow[u] \arrow[r] & \star & \\
 & b_\alpha \arrow[u, dotted] \arrow[r, dotted] & A_\alpha \arrow[u] \arrow[r] & \star \\
 & & & F \arrow[u]
\end{tikzcd} \tag{13}$$

5.1.2.3 Dirac Quantization

For $\mathsf{G} = \mathsf{U}(1)$ the curvature has integral periods, i.e. $F \in \Omega^2_{\mathrm{cl}}(X)_{\mathbb{Z}}$. This is known as Dirac quantization. The physical argument goes as follows.

We consider a sphere S^2 surrounding all magnetic charges. Let $S^2 = S^2_+ \cup S^2_-$, and let $A_\pm$ be the gauge fields on $S^2_\pm$. Then

$$\begin{aligned}
\int_{S^2} F = \int_{S^2_+} F + \int_{S^2_-} F = \int_{S^1} A_+ - \int_{S^1} A_- \\
= -\frac{1}{2\pi i} \int_{S^1} d \log g_{+-} \ \in\ \mathbb{Z}\,.
\end{aligned}$$

Thus, to summarize, electromagnetism is characterized by 'transition functions' $\{g_{\alpha\beta}\} \in \check{H}^1(X, \underline{\mathbb{R}/\mathbb{Z}})$ (or equivalently $\{n_{\alpha\beta\gamma}\} \in \check{H}^2(X, \mathbb{Z})$) and by a field strength $F \in \Omega^2_{\mathrm{cl}}(X)_{\mathbb{Z}}$, a closed 2-form with integral periods, whose cohomology class $[F] \in H^2_{\mathrm{dR}}(X)$ is related to $\{n_{\alpha\beta\gamma}\} \in \check{H}^2(X, \mathbb{Z})$ by the 'tic-tac-toe' equations depicted in (13).

That is, electromagnetism schematically contains the following information:

$$\begin{tikzcd}
 & F \arrow[d] \\
\{n_{\alpha\beta\gamma}\} \arrow[r] & {[F]}
\end{tikzcd}$$

or, in other words, in physics we can measure more than just the electromagnetic field strength F. There are other gauge invariant objects. A prime example is the holonomy of A

$$\mathrm{hol}_\gamma(A) = \exp\left(2\pi i \oint_\gamma A\right) ,$$

which, depending on the context, is known as the Bohm–Aharanov phase, Berry phase, a Wilson loop, etc.

The goal is to encapture all this gauge-invariant information in a cohomology theory, i.e. we are aiming to construct a cohomology theory that completes the square (the discussion above pertains to $p = 2$)

$$\begin{tikzcd}
? \arrow[r] \arrow[d] & \Omega^p_{\mathrm{cl}}(X)_{\mathbb{Z}} \arrow[d] \\
\check{H}^p(X,\mathbb{Z}) \arrow[r] & H^p(X,\mathbb{R})
\end{tikzcd}$$

This will lead us to the so-called differential cohomologies.

5.1.3 The Čech – de Rham Complex

Before we get to the subject of differential cohomologies let us consider the the tic-tac-toe procedure relating Čech and de Rham cohomologies, as discussed in the previous section in the example of electromagnetism, in greater generality.

Theorem 3 *We have an isomorphism*

$$\check{H}^p(X,\mathbb{R}) \cong H^p_{\mathrm{dR}}(X). \tag{14}$$

Proof The proof is given by 'tic-tac-toe'ing through the double complex, known as the Čech – de Rham complex, which is depicted below.

$$\begin{tikzcd}
 & {} & {} & {} & {} & \\
0 \arrow[r] & C^2(X,\mathbb{R}) \arrow[r] \arrow[u] & \Omega^0(Y^{[3]}) \arrow[r, "-d"] \arrow[u] & \Omega^1(Y^{[3]}) \arrow[r, "-d"] \arrow[u] & \Omega^2(Y^{[3]}) \arrow[r, "-d"] \arrow[u, "\delta"'] & \cdots \\
0 \arrow[r] & C^1(X,\mathbb{R}) \arrow[r] \arrow[u, "\delta"'] & \Omega^0(Y^{[2]}) \arrow[r, "d"] \arrow[u, "\delta"'] & \Omega^1(Y^{[2]}) \arrow[r, "d"] \arrow[u, "\delta"'] & \Omega^2(Y^{[2]}) \arrow[r, "d"] \arrow[u, "\delta"'] & \cdots \\
0 \arrow[r] & C^0(X,\mathbb{R}) \arrow[r] \arrow[u, "\delta"'] & \Omega^0(Y) \arrow[r, "-d"] \arrow[u, "\delta"'] & \Omega^1(Y) \arrow[r, "-d"] \arrow[u, "\delta"'] & \Omega^2(Y) \arrow[r, "-d"] \arrow[u, "\delta"'] & \cdots \\
 & & \Omega^0(X) \arrow[r, "d"] \arrow[u] & \Omega^1(X) \arrow[r, "d"] \arrow[u] & \Omega^2(X) \arrow[r, "d"] \arrow[u] & \cdots \\
 & & 0 \arrow[u] & 0 \arrow[u] & 0 \arrow[u] &
\end{tikzcd}$$

The rows and colums in this double complex starting from a zero are exact due to Theorems 1 and 2, while the bottom row computes $H_{\mathrm{dR}}(X)$, and the leftmost column computes $\check{H}(X,\mathbb{R})$. Now, let $\omega^p \equiv \omega^{(p,-1)} \in \Omega^p_{\mathrm{cl}}(X)$ be a representative of $[\omega] \in H^p_{\mathrm{dR}}(X)$. Define $\omega^{(p,0)} \in \Omega^p(Y)$ by

$$\omega^{(p,0)}_\alpha = (\delta\omega^p)_\alpha = \omega^p \big|_{U_\alpha}. \tag{15}$$

Next we observe

$$d\omega^{(p,0)}_\alpha = d\delta\omega^p = \delta d\omega^p = 0, \tag{16}$$

hence by exactness of the de Rham complex (Theorem 1 for $q = 0$), we have

$$\omega^{(p,0)}_\alpha = d\omega^{(p-1,0)}_\alpha, \tag{17}$$

for some $\omega^{(p-1,0)} \in \Omega^{p-1}(Y)$. Now define $\omega^{(p-1,1)} = \delta\omega^{(p-1,0)} \in \Omega^{p-1}(Y^{[2]})$. Continuing the process we construct $\omega^{(p-k,k)} \in \Omega^{p-k}(Y^{[k+1]})$ such that

$$\begin{aligned}
d\omega^{(p,-1)} &= 0\\
\delta\omega^{(p,-1)} - \quad d\omega^{(p-1,0)} &= 0\\
\vdots \qquad &= \vdots\\
\delta\omega^{(p-k,k)} + (-1)^k\, d\omega^{(p-k-1,k+1)} &= 0,\\
\vdots \qquad &= \vdots\\
\delta\omega^{(-1,p)} \qquad &= 0
\end{aligned}$$

and conclude that we have constructed a Čech cocycle $\check{\omega}^p = \omega^{(-1,p)}$, $\delta\check{\omega}^p = 0$. To conclude the latter equation, observe

$$d\delta\omega^{(-1,p)} = \delta d\omega^{(-1,p)} = (-1)^p \delta d(d\omega^{(0,p-1)}) = 0\,,$$

but since Ker d is zero at this step, we have $\delta\omega^{(-1,p)} = 0$. The various cocycles are depicted below.

$$
\begin{tikzcd}
0 \arrow[r] & 0 & & \\
\check{\omega}^p \arrow[r] \arrow[u] & \omega^{(0,p)} \arrow[r] \arrow[u] & 0 & \\
 & \omega^{(0,p-1)} \arrow[r] \arrow[u] & \omega^{(1,p-1)} \arrow[r] \arrow[u] & 0
\end{tikzcd}
$$

$$\cdots$$

$$
\begin{tikzcd}
{} \arrow[r] & \omega^{(p-1,1)} \arrow[r] & 0 & \\
 & \omega^{(p-1,0)} \arrow[r] \arrow[u] & \omega^{(p,0)} \arrow[r] \arrow[u] & 0 \\
 & & \omega^p \arrow[r] \arrow[u] & 0 \arrow[u]
\end{tikzcd}
$$

The procedure clearly generalizes the discussion in the case of electromagnetism. Examining the ambiguities in the various choices, one can easily verify that we have constructed a map $H^p_{\mathrm{dR}}(X) \to \check{H}^p(X, \mathbb{R})$. Arguing the other way around we get the inverse of this map, so this establishes the isomorphism. $\square$

By carefully studying the proof above, it is clear that an intermediate step in the proof is the construction of a $p+1$-tuple:

$$\omega = (\omega^{(-1,p)}, \omega^{(0,p-1)}, \ldots, \omega^{(p,-1)}), \tag{18}$$

such that $D\omega = 0$, where

$$D = \delta + (-1)^{q+1} d \tag{19}$$

on $\omega^{(p,q)}$. Note

$$D^2 = \delta^2 + (-1)^q (d\delta - \delta d) + d^2 = 0, \tag{20}$$

so D defines a cohomology $H^p_D(X)$, where $\deg(\omega^{(p,q)}) = p + q + 1$. It is clear that $H^p_D(X) \cong \check{H}^p(X, \mathbb{R}) \cong H^p_{\mathrm{dR}}(X)$.

One can ask what happens for $\check{H}^p(X, \mathbb{Z})$, the Čech cohomology with integer coefficients. In this case we still have a double complex

$$\begin{tikzcd}
 & \ & \ & \ & \ \\
C^2(X,\mathbb{Z}) \arrow[r] \arrow[u] & \Omega^0(Y^{[3]}) \arrow[r, "-d"] \arrow[u] & \Omega^1(Y^{[3]}) \arrow[r, "-d"] \arrow[u] & \Omega^2(Y^{[3]}) \arrow[r, "-d"] \arrow[u, "\delta"'] & \cdots \\
C^1(X,\mathbb{Z}) \arrow[r] \arrow[u, "\delta"'] & \Omega^0(Y^{[2]}) \arrow[r, "d"] \arrow[u, "\delta"'] & \Omega^1(Y^{[2]}) \arrow[r, "d"] \arrow[u, "\delta"'] & \Omega^2(Y^{[2]}) \arrow[r, "d"] \arrow[u, "\delta"'] & \cdots \\
C^0(X,\mathbb{Z}) \arrow[r] \arrow[u, "\delta"'] & \Omega^0(Y) \arrow[r, "-d"] \arrow[u, "\delta"'] & \Omega^1(Y) \arrow[r, "-d"] \arrow[u, "\delta"'] & \Omega^2(Y) \arrow[r, "-d"] \arrow[u, "\delta"'] & \cdots \\
 & \Omega^0(X) \arrow[r, "d"] \arrow[u] & \Omega^1(X) \arrow[r, "d"] \arrow[u] & \Omega^2(X) \arrow[r, "d"] \arrow[u] & \cdots \\
 & 0 \arrow[u] & 0 \arrow[u] & 0 \arrow[u] &
\end{tikzcd}$$

but the rows are no longer exact at the first term. Hence, in the proof above, we can in general no longer find a Čech cocycle $\check{\omega}^p = \pm\omega^{(-1,p)}$ such that $d\omega^{(-1,p)} = \delta\omega^{(0,p-1)}$. The tic-tac-toe procedure the other way around still works, so we still have a map $\check{H}^p(X,\mathbb{Z}) \to H^p_{\mathrm{dR}}(X) \cong \check{H}^p(X,\mathbb{R})$. This map is, in general, no longer surjective nor injective (cf. (2)).

5.1.4 Differential Cohomologies

As discussed in the case of electromagnetism, the purpose of differential cohomology is to find a 'differential model' that completes the square

$$\begin{tikzcd}
? \arrow[r] \arrow[d] & \Omega^p_{\mathrm{cl}}(X)_{\mathbb{Z}} \arrow[d] \\
\check{H}^p(X,\mathbb{Z}) \arrow[r] & H^p(X,\mathbb{R})
\end{tikzcd} \tag{21}$$

There are (at least) three such models, whose cohomology will generically be denoted by $\mathbb{H}^p(X)$. These are

$$\mathbb{H}^p(X) = \begin{cases} H^p(X,\mathbb{D}^p) & \text{Deligne cohomology} \\ \widehat{H}^p(X,\mathbb{R}/\mathbb{Z}) & \text{Cheeger–Simons differential characters.} \\ \check{H}(p)^p(X) & \text{Cheeger–Simons cohomology/sparks} \end{cases} \tag{22}$$

In fact, for all models we have short exact sequences

$$0 \longrightarrow H^{p-1}(X,\mathbb{R}/\mathbb{Z}) \longrightarrow \mathbb{H}^p(X) \longrightarrow \Omega^p_{\mathrm{cl}}(X)_{\mathbb{Z}} \longrightarrow 0 \tag{23}$$

and

$$0 \longrightarrow \Omega^{p-1}(X)/\Omega^{p-1}_{\mathrm{cl}}(X)_{\mathbb{Z}} \longrightarrow \mathbb{H}^p(X) \longrightarrow H^p(X,\mathbb{Z}) \longrightarrow 0 \tag{24}$$

which, in fact, fit into the extended commutative diagram [36, 37]

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & H^{p-1}(X,\mathbb{R})/H^{p-1}_{\mathrm{free}}(X,\mathbb{Z}) \arrow[r] \arrow[d] & \Omega^{p-1}(X)/\Omega^{p-1}_{\mathrm{cl}}(X)_{\mathbb{Z}} \arrow[r] \arrow[d] & d\Omega^{p-1}(X) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \check{H}^{p-1}(X,\mathbb{R}/\mathbb{Z}) \arrow[r] \arrow[d] & \mathbb{H}^p(X) \arrow[r] \arrow[d] & \Omega^p_{\mathrm{cl}}(X)_{\mathbb{Z}} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \check{H}^p_{\mathrm{tors}}(X,\mathbb{Z}) \arrow[r] \arrow[d] & \check{H}^p(X,\mathbb{Z}) \arrow[r] \arrow[d] & \check{H}^p_{\mathrm{free}}(X,\mathbb{Z}) \arrow[r] \arrow[d] & 0 \\
 & 0 & 0 & 0 &
\end{tikzcd} \tag{25}$$

where we have denoted

$$\check{H}^p_{\mathrm{free}}(X,\mathbb{Z}) = \mathrm{Coker}(\check{H}^p(X,\mathbb{Z}) \to \check{H}^p(X,\mathbb{R})) \tag{26}$$

and

$$\check{H}^p_{\mathrm{tors}}(X,\mathbb{Z}) = \mathrm{Ker}(\check{H}^p(X,\mathbb{Z}) \to \check{H}^p(X,\mathbb{R})). \tag{27}$$

5.1.4.1 Deligne Cohomology

The idea of Deligne cohomology is transparent from our electromagnetism example. In order to recover all of the physics in electromagnetism we need to modify the Čech – de Rham complex in two ways. First of all we need to work with Čech cochains which take values in the integers. Second, we need to cut off the Čech – de Rham complex on the right so as to retain the curvature $F \in \Omega^2_{\mathrm{cl}}(X)$ itself, rather than just its cohomology class; i.e. instead of the tic-tac-toe

$$\begin{tikzcd}
n_{\alpha\beta\gamma} \arrow[r] & \star & & \\
m_{\alpha\beta} \arrow[u, dotted] \arrow[r, dotted] & \Lambda_{\alpha\beta} \arrow[u] \arrow[r] & \star & \\
 & b_\alpha \arrow[u, dotted] \arrow[r, dotted] & A_\alpha \arrow[u] \arrow[r] & \star \\
 & & & F \arrow[u]
\end{tikzcd}$$

we only want to consider the part

$$
\begin{tikzcd}
n_{\alpha\beta\gamma} \arrow[r] & \star & \\
m_{\alpha\beta} \arrow[u, dotted] \arrow[r, dotted] & \Lambda_{\alpha\beta} \arrow[u] \arrow[r] & \star \\
& b_\alpha \arrow[u, dotted] \arrow[r, dotted] & A_\alpha \arrow[u]
\end{tikzcd}
$$

In general, we cut off the double complex by restricting to forms of degree less than p (we call this subcomplex $\mathbb{D}^p$). So, a Deligne cochain ω of degree q in $\mathbb{D}^p$ is a tuple

$$
\omega = \begin{cases} (\omega^{(-1,q)}, \omega^{(0,q-1)}, \ldots, \omega^{(p-1,q-p)}) & \text{for } q \geq p \\ (\omega^{(-1,q)}, \omega^{(0,q-1)}, \ldots, \omega^{(q-1,0)}) & \text{for } q < p, \end{cases} \tag{28}
$$

where $\omega^{(-1,q)} \in \check{C}^q(X, \mathbb{Z})$ and $\omega^{(k,l)} \in \Omega^{(k,l)} = \Omega^k(Y^{[l+1]})$, for $k, l \geq 0$. The differential $D = \delta + (-1)^{q+1} d$ is the same as in the Čech-de Rham complex,[3] but then restricted to the subcomplex $\mathbb{D}^p$. The corresponding cohomology, known as Deligne or Deligne–Beilinson cohomology, is denoted by $H^q(X, \mathbb{D}^p)$.

The following diagram shows the tic-tac-toes for the cases $q < p$, $q > p$.

$$
\begin{tikzcd}
C^p(X,\mathbb{Z}) & & & \star & & & \\
& & & & \star & & \\
& \star & & & & \star & (q > p) \\
& & \star & & & & \star \\
& & & \star & & & \\
& & & (q < p) & \star & & \Omega^{p-1}(Y)
\end{tikzcd}
$$

[3] Where, as before, d is replaced by $(2\pi i)^{-1} d \log$ on 0-forms.

It is clear that in these cases the cohomology is easily expressed in terms of known cohomologies. The interesting case is $q = p$. The following theorem summarizes the results.

Theorem 4 *We have*

$$H^q(X, \mathbb{D}^p) = \begin{cases} \check{H}^{q-1}(X, \mathbb{R}/\mathbb{Z}) & \text{for } q < p \\ \check{H}^q(X, \mathbb{Z}) & \text{for } q > p, \end{cases} \tag{29}$$

while for $q = p$, $\mathbb{H}^p(X, \mathbb{D}^p)$ fits in the exact sequences discussed in the previous section. In particular, the maps completing the square (21) *are given by*

$$\begin{tikzcd}
{[(\omega^{(-1,p)},} \arrow[d] & \cdots & {,\omega^{(p-1,0)})]} \arrow[d, "\omega^p = d\omega^{(p-1,0)}"] \\
\check{H}^p(X, \mathbb{Z}) \arrow[dr] & & \Omega^p_{\mathrm{cl}}(X)_{\mathbb{Z}} \arrow[dl] \\
& H^p(X, \mathbb{R}) &
\end{tikzcd}$$

5.1.4.2 Cheeger–Simons Differential Characters

In this section we discuss the basics of Cheeger–Simons differential characters. This section requires some knowledge of singular (co-)homology theory (see, e.g., [46]). In particular, we denote singular p-chains by $C_p(X)$, the differential on $C_p(X)$ by ∂, and the closed p-chains by $Z_p(X)$. The dual objects in singular cohomology are denoted by $C^p(X)$, ∂^*, and $Z^p(X)$, respectively.

The idea behind Cheeger–Simons differential characters is, as we have discussed in the case of electromagnetism, to consider in addition to F other gauge-invariant quantities such as the holonomies along paths $\gamma \in Z_1(X)$:

$$\mathrm{hol}_\gamma(A) = \exp\left(2\pi i \oint_\gamma A\right). \tag{30}$$

Let

$$c(\gamma) = \frac{1}{2\pi i} \log \mathrm{hol}_\gamma(A) = \oint_\gamma A \ \in \ \mathbb{R}/\mathbb{Z}, \tag{31}$$

then, for $\gamma \in C_2(X)$, we have

$$c(\partial\gamma) = \oint_{\partial\gamma} A = \int_\gamma F, \tag{32}$$

establishing the relation between the holonomies and the curvature F. Generalizing this to higher degrees leads to

Definition 1 [24, 16] A (Cheeger–Simons) differential character of degree p is a pair (χ, ω), where χ is a homomorphism $\chi : Z_{p-1}(X) \to \mathbb{R}/\mathbb{Z}$ and $\omega \in \Omega^p(X)$ such that

$$\chi(\partial\gamma) = \int_\gamma \omega \bmod \mathbb{Z}, \qquad \forall \gamma \in C_p(X). \tag{33}$$

The group of all differential characters of degree p is denoted by $\widehat{H}^p(X, \mathbb{R}/\mathbb{Z})$.

Lemma 1 *If* $(\chi, \omega) \in \widehat{H}^p(X, \mathbb{R}/\mathbb{Z})$, *then* $\omega \in \Omega^p_{\mathrm{cl}}(X)_{\mathbb{Z}}$.

Proof If $\partial\gamma = 0$, then

$$0 = \chi(\partial\gamma) = \int_\gamma \omega \bmod \mathbb{Z}, \tag{34}$$

hence ω has integral periods. Furthermore, for all $\gamma \in C_{p+1}(X)$ we have

$$0 = \chi(\partial^2\gamma) = \int_{\partial\gamma} \omega \bmod \mathbb{Z} = \int_\gamma d\omega \bmod \mathbb{Z}, \tag{35}$$

hence $d\omega = 0$. □

We also have a map $\widehat{H}^p(X, \mathbb{R}/\mathbb{Z}) \to \check{H}^p(X, \mathbb{Z})$ constructed as follows. Lift χ to $c' : Z_{p-1}(X) \to \mathbb{R}$ and interpret $\omega \in C^p(X)$. Then

$$\sigma = \omega - \partial^* c' \tag{36}$$

is a map $C_p(X) \to \mathbb{Z}$, satisfying $\partial^*\sigma = 0$, i.e. $\sigma \in \check{C}^p(X, \mathbb{Z})$ is closed. One can verify explicitly that the maps constructed above complete the square (21), and in fact that $\widehat{H}^p(X, \mathbb{R}/\mathbb{Z})$ fits into the commutative diagram (25) as claimed.

Given a Deligne class $\omega = (\omega^{(0,p-1)}, \ldots, \omega^{(p,-1)})$ and a p-chain $\gamma \in Z_p(X)$, there is a formula for the holonomy, generalizing (30)

$$\mathrm{hol}_\gamma(\omega) = \prod_{k=0}^{p} \exp\left(2\pi i \sum_{\{\sigma^k\}} \int_{\sigma^k} \omega^{(k,p-1-k)}\right), \tag{37}$$

where the set $\{\sigma^k\}$ denotes the k-cycles in a simplicial decomposition of γ. We refer to [21] for details.

5.1.4.3 Cheeger–Simons Differential Cohomology

The observation that underlies Cheeger–Simons differential cohomology [29, 36, 37, 39] is that in the tuple determining the Deligne cohomology only the first and the last term are important, since the intermediate terms can be obtained by tic-tac-toe'ing. This can be rephrased as saying that we define the complex $\check{C}(p)^{\bullet}(X)$ by the homotopy Cartesian square

$$\begin{array}{ccc} \check{C}(p)^{\bullet}(X) & \longrightarrow & \Omega^{\bullet \geq p}(X) \\ \big\downarrow & & \big\downarrow \\ \check{C}^{\bullet}(X,\mathbb{Z}) & \longrightarrow & \check{C}^{\bullet}(X,\mathbb{R}) \end{array} \tag{38}$$

that is,

$$\check{C}(p)^q(X) = \begin{cases} \check{C}^q(X,\mathbb{Z}) \times \check{C}^{q-1}(X,\mathbb{R}) \times \Omega^q(X)\,, & q \geq p, \\ \check{C}^q(X,\mathbb{Z}) \times \check{C}^{q-1}(X,\mathbb{R})\,, & q < p. \end{cases} \tag{39}$$

An element in $\check{C}(p)^q(X)$ will be denoted as a triple (c, h, ω), with the understanding that $\omega = 0$ for $q < p$.

The differential is given by

$$d(c, h, \omega) = (\delta c, \omega - c - \delta h, d\omega) \tag{40}$$

and the cohomology, called Cheeger–Simons cohomology in [39], by $\check{H}(p)^q(M)$.

Theorem 5 *[39] We have $\check{H}(p)^q(X) \cong H^q(X, \mathbb{D}^p)$ and $\check{H}(p)^p(X) \cong \widehat{H}^p(X, \mathbb{R}/\mathbb{Z})$.*

We will not prove this theorem here, but rather illustrate the various maps in one particular case. Namely, an element in $\check{H}(p)^p(X)$ is a triple $(c, h, \omega) \in \check{C}^p(X,\mathbb{Z}) \times \check{C}^{p-1}(X,\mathbb{R}) \times \Omega^p(X)$ such that

$$\delta c = 0, \qquad d\omega = 0, \qquad \delta h = \omega - c, \tag{41}$$

upto elements of the form $(\delta\tilde{c}, -\tilde{c} - \delta\tilde{h}, 0)$. We can now construct a map $\check{H}(p)^p(X) \to \widehat{H}^p(X, \mathbb{R}/\mathbb{Z})$, as follows: we let $[(c, h, \omega)] \mapsto (\chi, \omega)$ where $\chi : Z_{p-1}(X) \to \mathbb{R}/\mathbb{Z}$ is defined as

$$\chi(\gamma) = h(\gamma) \bmod \mathbb{Z} \tag{42}$$

for $\gamma \in Z_{p-1}(X)$. Note that χ is only uniquely defined on $Z_{p-1}(X)$, but not on $C_{p-1}(X)$ as $\delta\tilde{h}(\gamma) = \tilde{h}(\partial\gamma) = 0$ for $\gamma \in Z_{p-1}(X)$.

This concludes our discussion of differential cohomologies. We have seen three models for $\mathbb{H}^p(X)$ and have argued that $\mathbb{H}^2(X)$ is relevant to electromagnetism or, more precisely, classifies equivalence classes of principal $\mathsf{U}(1)$-bundles with connection. Similarly it turns out that $\mathbb{H}^3(X)$ classifies equivalence classes of ($\mathsf{U}(1)$-) gerbes with connection. A discussion of (bundle) gerbes is beyond the scope of these lectures, but see [52] (and references therein) for an excellent introduction.

5.2 T-Duality

5.2.1 Introduction to T-Duality

Target space duality, also known as T-duality is a particular symmetry of string theory or, more generally, a duality between different string theories (see [30] for a comprehensive review). It turns out that T-duality is related to various constructions in mathematics, such as Takai duality, mirror symmetry, Fourier–Mukai transform.

To introduce T-duality, let us consider the basic occurrence of T-duality in string theory. Consider thereto a closed string moving in a D-dimensional spacetime (target space) of the form $M \times S^1$, i.e. $D = \dim(M \times S^1)$.[4] The degrees of freedom of the string are encoded in a map $X : \Sigma \to M \times S^1$, describing the embedding of the two-dimensional string worldsheet Σ into the target spacetime. In terms of local coordinates (σ, τ), $\tau \in \mathbb{R}, 0 \leq \sigma \leq \pi$, on the string worldsheet and local coordinates X^N, $N = 1, \ldots, D$ on the target manifold, the dynamics of the string is described by some two-dimensional field theory consisting of D scalar fields $X^N(\sigma, \tau)$.

Suppose the circle S^1 has radius R and lies in the Dth direction. Upon quantization of the string, and for that matter any quantum field theory of particles, the momentum p^D in the direction of the circle becomes quantized in units of $1/R$ (Bohr quantization rule), i.e. $p^D = n/R$, [5] with $n \in \mathbb{Z}$, ensuring, for example, the single-valuedness of the exponential $\exp(ip \cdot X)$. The string, however, distinguishes itself from a theory of particles in the sense that strings can wind around compact directions. While in flat space the condition that a string is closed translates into the boundary condition $X^N(0, \tau) = X^N(\pi, \tau)$, for all τ, in the presence of the circle the string can wind around the circle m times, i.e. satisfies boundary conditions $X^D(0, \tau) = X^D(\pi, \tau) + mR$, for all τ. The energy spectrum of a string with *momentum number n* and m is schematically given by *winding number*

$$E = \left(\frac{n}{R}\right)^2 + (mR)^2 + \cdots, \tag{43}$$

where the dots stand for terms independent of R and depend on the details of the manifold M. We conclude that the spectrum of the string is invariant under the simultaneous interchange of R and $1/R$ (the circle with its 'dual' circle) and momentum and winding numbers. In fact, this duality not only holds for the energy spectrum but all physical observables (such as scattering amplitudes) that can be computed in this theory. The string theory on $M \times S^1_R$ is completely equivalent to

[4] Throughout this chapter we use various notation for the circle interchangeably, namely $\mathbb{R}/\mathbb{Z}$, S^1, $\mathsf{U}(1)$, or $\mathbb{T}$. In general we use the first two if we only consider the circle as a manifold, while the latter two notations are used when viewing the circle as a Lie group (with group multiplication written multiplicatively).

[5] Here and in the discussion that follows, we omit constants such as 2π which are irrelevant for the argument.

the same string theory on $M \times S^1_{1/R}$. In other words, this string theory cannot be used to determine whether we are living in a universe in which spacetime has a compact direction S^1_R as opposed to $S^1_{1/R}$. This phenomenon is known as T-duality.

If one includes supersymmetry, i.e. considers superstrings, the situation gets slightly more complicated and T-duality becomes, in general, a duality between different superstring theories (such as between type IIA and type IIB). We will not focus on superstrings in this review, but instead consider a different generalization, namely to strings on manifolds which have locally defined circles (e.g. circle bundles, circle fibrations) and move in the presence of a background flux. In particular we will focus on the topological properties of the spacetime and flux and their duals. The discussion can be generalized to manifolds with a (higher rank) torus action.

5.2.2 The Buscher Rules

To generalize the discussion in the previous section, consider a closed string moving in a target space(time) M, described by a map $X : \Sigma \to M$ (or $X^M(\sigma, \tau)$ in local coordinates). The low-energy effective field theory for the string is given by coupling the string modes to the massless fields in its spectrum, which always include the metric $g_{MN}(X)$ of the target manifold M, a so-called B-field $B_{MN}(X)$ (a locally defined 2-form gauge field) and a dilaton $\Phi(X)$. The action is given by a so-called nonlinear sigma model action

$$S[X] = \int d\sigma d\tau \Big(\sqrt{h} h^{\alpha\beta} g_{MN}(X) \partial_\alpha X^M \partial_\beta X^N + \varepsilon^{\alpha\beta} B_{MN}(X) \partial_\alpha X^M \partial_\beta X^N + \alpha' \sqrt{h} R(h) \Phi(X) \Big), \tag{44}$$

where $h_{\alpha\beta}$ denotes some (auxiliary) metric on the string worldsheet, with curvature $R(h)$. Conformal invariance of the above sigma model allows us to (locally) choose a flat metric $h_{\alpha\beta} = \eta_{\alpha\beta}$. For simplicity, we will also ignore the dilaton in the discussion that follows. Hence, upon introducing complex coordinates $(z, \bar{z})$, instead of (σ, τ), the action (44) takes the form

$$S[X] = \int d^2 z \; Q_{MN}(X) \partial X^M \bar{\partial} X^N, \tag{45}$$

where $Q_{MN} = g_{MN} + B_{MN}$.

Now suppose the action (45) has a $U(1)^N$ isometry, then we may choose local coordinates $X^M = (X^\mu, X^m)$, $(m = 1, \ldots, N,\ \mu = N+1, \ldots, D)$, such that the isometry acts by $X^m \to X^m + \varepsilon^m$, i.e. the Killing vectors are given by $\kappa_m = \partial/\partial X^m$ and $\mathcal{L}_{\kappa_m} Q_{MN} = 0$. That is, $Q_{MN}(X^\mu, X^m) = Q_{MN}(X^\mu)$ is independent of the coordinates X^m along the isometry.

Upon decomposing

$$Q_{MN} = \begin{pmatrix} Q_{\mu\nu} & Q_{\mu n} \\ Q_{m\nu} & Q_{mn} \end{pmatrix}, \tag{46}$$

the action can be written as

$$\begin{aligned} S[X] = \int d^2z \Big(& Q_{\mu\nu}(X)\partial X^\mu \bar\partial X^\nu + Q_{\mu n}(X)\partial X^\mu \bar\partial X^n + Q_{m\nu}(X)\partial X^m \bar\partial X^\nu \\ & + Q_{mn}(X)\partial X^m \bar\partial X^n \Big). \end{aligned} \tag{47}$$

Now, as mentioned before, this action has a (translation) symmetry $X^m(z,\bar z) \to X^m(z,\bar z) + \varepsilon^m$ for constants ε^m. In order to explore the consequences of this symmetry we explore a common trick in quantum field theory known as 'gauging the symmetry'. In order to write down an action which is invariant under the local symmetry $X^m(z,\bar z) \to X^m(z,\bar z) + \varepsilon^m(z,\bar z)$ we introduce $\mathsf{U}(1)^N$ gauge fields $(A^m(z,\bar z), A^m(z,\bar z))$ and couple them minimally to the fields $X^m(z,\bar z)$, i.e. we replace the derivates $\partial, \bar\partial$ by covariant derivatives

$$\begin{aligned} \partial X^m &\to DX^m = \partial X^m + A^m , \\ \bar\partial X^m &\to \bar D X^m = \bar\partial X^m + \bar A^m . \end{aligned} \tag{48}$$

In addition we introduce auxiliary fields $\widehat{X}_m$ to the action, whose equations of motion imply that the $\mathsf{U}(1)^N$ curvature $F^m = \partial \bar A^m - \bar\partial A^m$ is vanishing; i.e. we consider

$$\begin{aligned} S[X] = \int d^2z \Big(& Q_{\mu\nu}(X)\partial X^\mu \bar\partial X^\nu + Q_{\mu n}(X)\partial X^\mu \bar D X^n + Q_{m\nu}(X) DX^m \bar\partial X^\nu \\ & + Q_{mn}(X) DX^m \bar D X^n + \widehat{X}_m \big(\partial \bar A^m - \bar\partial A^m\big)\Big) . \end{aligned} \tag{49}$$

It can now easily be checked that the action (49) has a local $\mathsf{U}(1)^N$ gauge symmetry given by

$$\begin{aligned} X^m(z,\bar z) &\to X^m(z,\bar z) + \varepsilon^m(z,\bar z) , \\ A^m(z,\bar z) &\to A^m(z,\bar z) - \partial\varepsilon^m(z,\bar z) , \\ \bar A^m(z,\bar z) &\to \bar A^m(z,\bar z) - \bar\partial\varepsilon^m(z,\bar z) . \end{aligned} \tag{50}$$

We can think of the sigma model (49) as having a target space, locally described by coordinates $X^\mu, X_m, \widehat{X}_m$). This turns out to be the *correspondence space* to be discussed in later sections.

It turns out that, despite having introduced both additional degrees of freedom and additional symmetries, we have done so precisely such that they cancel out. That is, we claim that the action (49) is equivalent, at least at a semi-classical level, to the action (47); namely, starting with (49) and integrating out the auxiliary coordinate $\widehat{X}_m$ forces $F^m = \partial \bar{A}^m - \bar{\partial} A^m = 0$. [6] This implies $A^m = \partial \widetilde{X}^m$, $\bar{A}^m = \bar{\partial} \widetilde{X}^m$. Choosing the gauge in which $\widetilde{X}^m = 0$ then leads back to (47). On the other hand, one easily verifies that by first integrating out the gauge fields $(A^m,\ \bar{A}^m)$ and then gauge fixing, one reproduces the action (47) in terms of dual coordinates $(X^\mu, \widehat{X}_m)$ coupled to

$$\widehat{Q}_{MN} = \begin{pmatrix} \widehat{Q}_{\mu\nu} & \widehat{Q}_{\mu n} \\ \widehat{Q}_{m\nu} & \widehat{Q}_{mn} \end{pmatrix} = \begin{pmatrix} Q_{\mu\nu} - Q_{\mu n}(Q^{-1})_{mn} Q_{n\nu} & Q_{\mu m}(Q^{-1})_{mn} \\ -(Q^{-1})_{mn} Q_{n\nu} & (Q^{-1})_{mn} \end{pmatrix}. \tag{51}$$

The local transformation rules between (g_{MN}, B_{MN}) and $(\widehat{g}_{MN}, \widehat{B}_{MN})$, as encoded by (51), are known as the *Buscher rules* [19, 20]. Sigma models (47), related by the Buscher rules (51), are completely equivalent (at least at the semi-classical level) and correspondingly describe equivalent (or dual) string theories.

In the case of a $\mathsf{U}(1)$ isometry (i.e. $N = 1$) we can easily unravel the original symmetric and antisymmetric components $(\widehat{g}_{MN}, \widehat{B}_{MN})$ of $\widehat{Q}_{MN}$ in (51), and we obtain

$$\begin{aligned}
\widehat{g}_{\bullet\bullet} &= \frac{1}{g_{\bullet\bullet}}, \\
\widehat{g}_{\mu\bullet} &= \frac{B_{\mu\bullet}}{g_{\bullet\bullet}}, \\
\widehat{g}_{\mu\nu} &= g_{\mu\nu} - \frac{1}{g_{\bullet\bullet}}\left(g_{\mu\bullet} g_{\nu\bullet} - B_{\mu\bullet} B_{\nu\bullet}\right), \\
\widehat{B}_{\mu\bullet} &= \frac{g_{\mu\bullet}}{g_{\bullet\bullet}}, \\
\widehat{B}_{\mu\nu} &= B_{\mu\nu} - \frac{1}{g_{\bullet\bullet}}\left(g_{\mu\bullet} B_{\nu\bullet} - g_{\nu\bullet} B_{\mu\bullet}\right),
\end{aligned} \tag{52}$$

where, for notational simplicity, we have indicated the circle direction by a $\bullet$. We easily see that (52) is a generalization of the example in Sect. 5.2.1.

Our discussion of the Buscher rules has been purely local. Nevertheless it is possible to extract the global, or topological, information contained in the Buscher rules (52). To this end, suppose now our spacetime is of the form $N \times Y$, where Y is a principal circle bundle $\pi : Y \to X$ over X.[7] Let A be a connection on the circle bundle Y. Locally

[6] Integrating out an auxiliary coordinate means to solve for the equations of motion of this coordinate and substitute the solution back into the action.

[7] We will ignore N in the remainder as it plays no role in our discussion. It should be remembered, however, that in order for our spacetime theory to define a consistent string theory the total spacetime $N \times Y$ needs to satisfy certain constraints.

$$A = A_M dx^M = dx^\bullet + \overline{A}_\mu dx^\mu . \tag{53}$$

We can decompose both the canonical metric on Y and the B-field, with respect to the base, in terms of the connection

$$\begin{aligned} g &= \overline{g}_{\mu\nu} dx^\mu dx^\nu + (dx^\bullet + \overline{A}_\mu dx^\mu)^2 \\ B &= \tfrac{1}{2}\overline{B}_{\mu\nu} dx^\mu \wedge dx^\nu + \overline{B}_\mu dx^\mu \wedge (dx^\bullet + \overline{A}_\nu dx^\nu), \end{aligned} \tag{54}$$

where $\frac{1}{2}\overline{B}_{\mu\nu} dx^\mu \wedge dx^\nu$ and $\overline{B}_\mu dx^\mu$ are a 2-form and a 1-form on the base X, respectively. Then, by inserting (54) into the Buscher rules (52), we see that the Buscher rules essentially correspond to the interchange $\overline{A}_\mu \leftrightarrow \overline{B}_\mu$. To be precise, starting with

$$g_{MN} = \begin{pmatrix} \overline{g}_{\mu\nu} + \overline{A}_\mu \overline{A}_\nu & \overline{A}_\mu \\ \overline{A}_\nu & 1 \end{pmatrix}, \qquad B_{MN} = \begin{pmatrix} \overline{B}_{\mu\nu} + (\overline{B}_\mu \overline{A}_\nu - \overline{A}_\mu \overline{B}_\nu) & \overline{B}_\mu \\ -\overline{B}_\nu & 0 \end{pmatrix}, \tag{55}$$

the Buscher rules give

$$\widehat{g}_{MN} = \begin{pmatrix} \overline{g}_{\mu\nu} + \overline{B}_\mu \overline{B}_\nu & \overline{B}_\mu \\ \overline{B}_\nu & 1 \end{pmatrix}, \qquad \widehat{B}_{MN} = \begin{pmatrix} \overline{B}_{\mu\nu} & \overline{A}_\mu \\ -\overline{A}_\nu & 0 \end{pmatrix}. \tag{56}$$

Denoting the coordinate of the dual circle by $\hat{x}^\bullet$, we can interpret $\widehat{A} = d\hat{x}^\bullet + \overline{B}_\mu dx^\mu$, locally, as a connection on a dual circle bundle $\widehat{\pi} : \widehat{Y} \to X$. We deduce from (56) that on the correspondence space $Y \times_X \widehat{Y} = \{(y, \hat{y}) \in Y \times \widehat{Y} \mid \pi(y) = \widehat{\pi}(\hat{y})\}$, with local coordinates $(x^\mu, x^\bullet, \hat{x}^\bullet)$,

$$\widehat{B} = B + A \wedge \widehat{A} - dx^\bullet \wedge d\widehat{x}^\bullet, \tag{57}$$

so that

$$\widehat{H} - H = d(A \wedge \widehat{A}) = F \wedge \widehat{A} - A \wedge \widehat{F} x, \tag{58}$$

where $F = dA$ and $\widehat{F} = d\widehat{A}$ are the curvatures of A and $\widehat{A}$, respectively, and are (globally) defined forms on M. Equation (58) actually makes sense globally on $Y \times_X \widehat{Y}$. Rewriting this equation as

$$H - \widehat{F} \wedge A = \widehat{H} - F \wedge \widehat{A}, \tag{59}$$

we see that the left-hand side is a form on Y, while the right-hand side is a form on $\widehat{Y}$. Thus, in order to have equality, we conclude that both have to equal a form H_3 defined on X, i.e.

$$\begin{aligned} H &= H_3 + \widehat{F} \wedge A x, \\ \widehat{H} &= H_3 + F \wedge \widehat{A} x. \end{aligned} \tag{60}$$

We note that these equations imply that

$$F = \widehat{\pi}_* \widehat{H}, \qquad \widehat{F} = \pi_* H, \tag{61}$$

where π_* and $\widehat{\pi}_*$ are the integrations over the S^1 fibres of Y and $\widehat{Y}$, respectively (i.e. the push-forward maps in cohomology). In other words, the H-flux and first Chern class of the circle bundle are exchanged under T-duality. It is believed this duality extends to Čech classes as well.

We can now summarize the topological content of T-duality for principal circle bundles as follows:

Theorem 6 *[6, 7] Given a pair $(Y, [H])$ of an (isomorphism class of) circle bundle $\pi : Y \to X$ corresponding to a class $[F] \in H^2(X, \mathbb{Z})$, with H-flux $[H] \in H^3(Y, \mathbb{Z})$, the T-dual is a pair $(\widehat{Y}, [\widehat{H}])$, with $\widehat{\pi} : \widehat{Y} \to X$ a dual circle bundle corresponding to a class $[\widehat{F}] \in H^2(X, \mathbb{Z})$, and $[\widehat{H}] \in H^3(\widehat{Y}, \mathbb{Z})$*

$$\begin{tikzcd} \mathbb{T} \arrow[r] & Y \arrow[d, "\pi"'] \\ & X \end{tikzcd} \qquad \begin{tikzcd} \mathbb{T} \arrow[r] & \widehat{Y} \arrow[d, "\widehat{\pi}"'] \\ & X \end{tikzcd} \tag{62}$$

They are related by

$$[F] = \widehat{\pi}_*[\widehat{H}], \qquad [\widehat{F}] = \pi_*[H], \tag{63}$$

such that on the correspondence space

$$\begin{tikzcd} & Y \times_X \widehat{Y} \arrow[dl, "p"] \arrow[dr, "\widehat{p}"'] & \\ Y \arrow[dr, "\pi"'] & & \widehat{Y} \arrow[dl, "\widehat{\pi}"] \\ & X & \end{tikzcd} \tag{64}$$

we have

$$p^*[H] = \widehat{p}^*[\widehat{H}]. \tag{65}$$

5.2.3 Gysin Sequences and Dimensional Reduction

Let us first review how the global content of the Buscher rules, i.e. Theorem 6, is encoded in the Gysin sequence for the principal circle bundle $\pi : Y \to X$ (cf. [6, 7, 10]). Principal circle bundles are classified, up to isomorphism, by the Euler class $\chi(Y) \in H^2(X, \mathbb{Z})$ or equivalently by the first Chern class $c_1(L_Y) \in H^2(X, \mathbb{Z})$ of the associated line bundle $L_Y = Y \times_{\mathbb{T}} \mathbb{C}$. Given a principal circle bundle $\pi : Y \to X$, we have the pullback map $\pi^* : H^k(X, \mathbb{Z}) \to H^k(Y, \mathbb{Z})$ and the push-forward map ('integration over the S^1-fibre' in the case of forms) $\pi_* : H^k(Y, \mathbb{Z}) \to H^{k-1}(X, \mathbb{Z})$. These maps nicely fit into a long exact sequence in cohomology, the so-called Gysin sequence

$$\longrightarrow H^k(X, \mathbb{Z}) \xrightarrow{\pi^*} H^k(Y, \mathbb{Z}) \xrightarrow{\pi_*} H^{k-1}(X, \mathbb{Z}) \xrightarrow{\delta} H^{k+1}(X, \mathbb{Z}) \longrightarrow \tag{66}$$

where the map $\delta : H^{k-1}(X, \mathbb{Z}) \to H^{k+1}(X, \mathbb{Z})$ is given, on a class $\omega \in H^{k-1}(X, \mathbb{Z})$, by $\delta\omega = [F] \cup \omega$. Here, $[F] \in H^2(X, \mathbb{Z})$ is the Euler class of Y (i.e. the curvature of a connection on Y).

Considering the $k = 3$ segment of the Gysin sequence (66), we see that any class $[H] \in H^3(Y, \mathbb{Z})$, i.e. any H-flux on Y, gives rise to a class $\pi_*[H] \in H^2(X, \mathbb{Z})$, which can be interpreted as $[\widehat{F}]$, the Euler class of a T-dual circle bundle $\hat{\pi} : \widehat{Y} \to X$. Furthermore, we have $[F] \cup [\widehat{F}] = 0$ in $H^4(X, \mathbb{Z})$. Conversely, by considering the Gysin sequence corresponding to the T-dual circle bundle $\hat{\pi} : \widehat{Y} \to X$, we conclude from $[F] \cup [\widehat{F}] = [\widehat{F}] \cup [F] = 0$ that $[F] = \widehat{\pi}_*[\widehat{H}]$ for some class $[\widehat{H}] \in H^3(\widehat{Y}, \mathbb{Z})$. This is precisely the content of (63).

From the Gysin sequence we can of course only determine the element $[\widehat{H}] \in H^3(\widehat{Y}, \mathbb{Z})$ up to an element in $\pi^*(H^3(X, \mathbb{Z}))$. To fix this ambiguity, we need some extra input. The extra input, of course, is that T-duality should not affect that part of the H-flux that 'lives' on the base manifold X. This is equivalent to demanding that $p^*[H] - \hat{p}^*[\widehat{H}] = 0$ in $H^3(Y \times_X \widehat{Y}, \mathbb{Z})$ as in (65)

The above considerations are summarized in the following diagram:

$$\begin{tikzcd}
 & & {[H]} \arrow[r, mapsto] \arrow[d, phantom, "\in"] & {[\widehat{F}]} \arrow[r, mapsto] \arrow[d, phantom, "\in"] & 0 & \\
\arrow[r] & H^3(X, \mathbb{Z}) \arrow[r, "\pi^*"] & H^3(Y, \mathbb{Z}) \arrow[r, "\pi_*"] & H^2(X, \mathbb{Z}) \arrow[r, "{[F]\cup}"] \arrow[d, dotted] & H^4(X, \mathbb{Z}) \arrow[r] & {} \\
\arrow[r] & H^3(X, \mathbb{Z}) \arrow[r, "\widehat{\pi}^*"] & H^3(\widehat{Y}, \mathbb{Z}) \arrow[r, "\widehat{\pi}_*"] & H^2(X, \mathbb{Z}) \arrow[r, "{[\widehat{F}]\cup}"] & H^4(X, \mathbb{Z}) \arrow[r] & {} \\
 & & {[\widehat{H}]} \arrow[r, mapsto] \arrow[u, phantom, "\in"] & {[F]} \arrow[r, mapsto] \arrow[u, phantom, "\in"] & 0 &
\end{tikzcd}$$

To illuminate the T-duality rules even further we can recast the Gysin sequence in a dimensionally reduced form following [10]. We will do this only at the level of forms, i.e. de Rham cohomology.

Let κ denote the (globally defined) Killing vector field corresponding to the $\mathsf{U}(1)$-isometry, and let $\Omega^k(Y)_{S^1}$ denote the space of k-forms invariant under the isometry, i.e. $\mathcal{L}_\kappa \Omega = 0$. Let us also choose a connection A on Y. Then we have a map

$$f_A : \Omega^k(X) \oplus \Omega^{k-1}(X) \to \Omega^k(Y)_{S^1} , \quad (\Omega_{(k)}, \Omega_{(k-1)}) \mapsto \pi^*\Omega_{(k)} + A \wedge \pi^*\Omega_{(k-1)}, \tag{67}$$

with inverse

$$f_A^{-1} : \Omega^k(Y)_{S^1} \to \Omega^k(X) \oplus \Omega^{k-1}(X) , \qquad \Omega \mapsto (\Omega - A \wedge \pi_*\Omega, \pi_*\Omega). \tag{68}$$

A simple computation shows

$$(d \circ f_A)(\Omega_{(k)}, \Omega_{(k-1)}) = (d\Omega_{(k)} + F \wedge \Omega_{(k-1)}) - A \wedge d\Omega_{(k-1)}. \tag{69}$$

Thus, upon defining a modified differential $D : \Omega^k(X) \oplus \Omega^{k-1}(X) \to \Omega^{k+1}(X) \oplus \Omega^k(X)$ by

$$D(\Omega_{(k)}, \Omega_{(k-1)}) = (d\Omega_k + F \wedge \Omega_{(k-1)}, -d\Omega_{(k-1)}), \tag{70}$$

we have $d \circ f_A = f_A \circ D$. It is straightforward to check that $D^2 = 0$, and hence that D defines a cohomology $H^k_D(X) \equiv H^k(\Omega^\bullet(X) \oplus \Omega^{\bullet-1}(X), D)$. Furthermore, because of the commutativity of the diagram

$$\begin{tikzcd} \Omega^k(X) \oplus \Omega^{k-1}(X) \arrow[r, "\cong", "f_A"'] \arrow[d, "D"'] & \Omega^k(Y)_{S^1} \arrow[d, "d"] \\ \Omega^{k+1}(X) \oplus \Omega^k(X) \arrow[r, "\cong", "f_A"'] & \Omega^{k+1}(Y)_{S^1} \end{tikzcd} \tag{71}$$

we have the result

$$H^k(Y) \cong H^k_D(X). \tag{72}$$

While the explicit isomorphism (67) depends on the choice of connection A, it is easily verified that the isomorphism (72) is independent of the choice of A.

Now that we have a globally defined dimensional reduction of forms (68) and an identification of cohomology (72), it is straightforward to dimensionally reduce the Gysin sequence (66) (at the level of de Rham cohomology). The result is the following exact sequence

$$\cdots \longrightarrow H^k(X) \xrightarrow{\pi^*} H^k_D(X) \xrightarrow{\pi_*} H^{k-1}(X) \xrightarrow{\delta} H^{k+1}(X) \longrightarrow \cdots \tag{73}$$

where the various maps, on representatives of the cohomology, are given by

$$\begin{aligned} \pi^* : H^k(X) \to H^k_D(X)\,, &\qquad \pi^*(\Omega_{(k)}) = (\Omega_{(k)}, 0)\,, \\ \pi_* : H^k_D(X) \to H^{k-1}(X)\,, &\qquad \pi_*(\Omega_{(k)}, \Omega_{(k-1)}) = \Omega_{(k-1)}\,, \\ \delta : H^{k-1}(X) \to H^{k+1}(X)\,, &\qquad \delta(\Omega_{(k-1)}) = F \wedge \Omega_{(k-1)}\,. \end{aligned} \tag{74}$$

In fact, using these maps it is easy to prove that the Gysin sequence is exact (see [10] for details).

Let us use this dimensionally reduced formalism to show that T-duality leads to an isomorphism of *twisted cohomologies*. First, let us denote the space of even and odd forms on Y by $\Omega^{\bar{0}}(Y)$ and $\Omega^{\bar{1}}(Y)$, respectively, i.e.

$$\Omega^{\bar{\imath}}(Y) = \bigoplus_{i=\bar{\imath} \bmod 2} \Omega^i(Y)\,. \tag{75}$$

Then, given a representative H for a class $[H] \in H^3(Y)$, we can construct a 'twisted differential' $d_H : \Omega^{\bar{\imath}} \to \Omega^{\overline{\imath+1}}$ by

$$d_H\Omega = d\Omega + H \wedge \Omega\,. \tag{76}$$

Clearly, $(d_H)^2 = 0$ (since $dH = 0$). The cohomology of the $\mathbb{Z}_2$-graded complex $(\Omega^\bullet(Y), d_H)$ is known as the twisted cohomology $H^{\bar{\imath}}(Y, [H])$ of Y, with respect to the 3-form H. It is easy to see that while explicit representatives for twisted cohomology classes depend on the choice of H, the twisted cohomology itself only depends on the class $[H]$. Let us now examine what a twisted cohomology class looks like under the dimensional reduction.

Decomposing $H = H_{(3)} + A \wedge H_{(2)}$ and $\Omega = \Omega' + A \wedge \Omega''$ as in (67), we have

$$d_H\Omega = (d\Omega' + H_{(3)} \wedge \Omega' + F \wedge \Omega'') + A \wedge (-d\Omega'' - H_{(3)} \wedge \Omega'' + H_{(2)} \wedge \Omega')\,.$$

Thus, the condition for Ω to be a twisted cohomology class, i.e. $d_H\Omega = 0$, decomposes into two equations

$$\begin{aligned} d\Omega' + H_{(3)} \wedge \Omega' + F \wedge \Omega'' = 0\,, \\ d\Omega'' + H_{(3)} \wedge \Omega'' - H_{(2)} \wedge \Omega' = 0\,. \end{aligned} \tag{77}$$

Note that both equations do not depend on the choice of A and are described completely in terms of forms on X.

Now, consider the pair $((H_{(3)}, H_{(2)}), F) \in H^3_D(X) \oplus H^2(X)$. It follows from (74) and the discussion before, that the T-duality transformation in this dimensionally reduced formalism is given by

$$((\widehat{H}_{(3)}, \widehat{H}_{(2)}), \widehat{F}) = ((H_{(3)}, F), H_{(2)})\,. \tag{78}$$

Therefore, T-duality provides an isomorphism on twisted cohomology $T_* : H^\bullet(Y, [H]) \to H^{\bullet+1}(\widehat{Y}, [\widehat{H}])$, which is explicitly given by

$$(\widehat{\Omega}', \widehat{\Omega}'') = (\Omega'', -\Omega')\,. \tag{79}$$

That is, $d_H\Omega = 0$ iff $d_{\widehat{H}}\widehat{\Omega} = 0$. Of course, (79) agrees with the 'Hori formula' [40, 6]

$$T_*\Omega = \widehat{\Omega} = \int_{S^1} e^{A\wedge\widehat{A}}\,\Omega\,. \tag{80}$$

The discussion above can be lifted to K-theory and, in this more general setting, T-duality gives an isomorphism of the twisted K-theories of Y and $\widehat{Y}$, descending to an isomorphism between the twisted cohomologies of Y and $\widehat{Y}$, as expressed in the following commutative diagram [6]

$$\begin{array}{ccc} K^\bullet(Y,[H]) & \xrightarrow{T_!} & K^{\bullet+1}(\widehat{Y},[\widehat{H}]) \\ {\scriptstyle ch_H}\big\downarrow & & \big\downarrow{\scriptstyle ch_{\widehat{H}}} \\ H^\bullet(Y,[H]) & \xrightarrow{T_*} & H^{\bullet+1}(\widehat{Y},[\widehat{H}]) \end{array}\quad, \tag{81}$$

where ch_H denotes the twisted Chern character (see, e.g. [5]). A thorough discussion of twisted K-theory is beyond the scope of these lectures (see, however, Sect. 5.2.4). Let us just conclude by remarking that twisted K-theory is believed to be classifying D-brane charges in the background of H-flux [13, 50, 51, 60], and that the isomorphism (81) is consistent with the statement that the string theories on $(Y, [H])$ and $(\widehat{Y}, [\widehat{H}])$ are T-dual.

The dimensionally reduced formalism discussed in this section can be applied in the case of higher rank principal torus bundles as well and leads to a concrete description of the transformation of the characteristic classes when T-duality is applied to principal torus bundles with background H-flux. We refer to [11] for details.

5.2.4 T-Duality = Takai Duality

Using the fact that $\mathsf{PU}(\mathcal{H})$, the projective unitary group on a separable Hilbert space $\mathcal{H}$ has a central extension

$$1 \longrightarrow \mathsf{U}(1) \longrightarrow \mathsf{U}(\mathcal{H}) \longrightarrow \mathsf{PU}(\mathcal{H}) \longrightarrow 1 \tag{82}$$

together with the fact that $\mathsf{U}(\mathcal{H})$ is contractible (Kuiper's theorem), we have from the long exact sequence in homotopy that $\pi_k(\mathsf{PU}(\mathcal{H})) \cong \pi_{k-1}(\mathsf{U}(1))$. Hence $\pi_k(\mathsf{PU}(\mathcal{H})) \cong \mathbb{Z}$ for $k = 2$ and vanishing for $k \neq 2$. We conclude that $\mathsf{PU}(\mathcal{H})$ is a model for the classifying space $K(\mathbb{Z}, 2)$. Hence $B\mathsf{PU}(\mathcal{H}) = K(\mathbb{Z}, 3)$ and thus isomorphism classes of principal $\mathsf{PU}(\mathcal{H})$-bundles over Y are classified by $H^3(Y, \mathbb{Z})$.

Thus we can geometrize our data $(Y, [H])$ in terms of a principal $\mathsf{PU}(\mathcal{H})$-bundle P over the principal circle bundle Y. Equivalently, noting that $\mathsf{PU}(\mathcal{H}) = \mathrm{Aut}(\mathcal{K})$ where $\mathcal{K}$ denotes the algebra of compact operators on $\mathcal{H}$, we can replace P by the associated algebra bundle $\mathcal{E} = P \times_{\mathsf{PU}(\mathcal{H})} \mathcal{K}$. The space of continuous sections $\mathcal{A} = C(Y, \mathcal{E})$ is a stable, continuous-trace, C*-algebra with spectrum Y, and the $\mathbb{T}$-action on Y lifts uniquely to an $\mathbb{R}$-action on $\mathcal{A} = C(Y, \mathcal{E})$. Continuous-trace algebras like $\mathcal{A}$ are determined, upto Morita equivalence, by their so-called Dixmier – Douady class $\mathrm{DD}(\mathcal{A}) \in H^3(Y, \mathbb{Z})$ which, in the above construction, is given by $[H]$. Moreover, the twisted K-theory $K^\bullet(Y, [H])$ can be defined as the K-theory $K_\bullet(\mathcal{A})$ of $\mathcal{A}$.

For any C*-algebra $\mathcal{A}$, with an action of a group G, i.e. a homomorphism

$$\alpha : \mathsf{G} \to \mathrm{Aut}(\mathcal{A}) , \tag{83}$$

we can construct the crossed product C*-algebra $\mathcal{A} \times_\alpha \mathsf{G}$ as follows. First, we consider the set of compactly supported continuous functions $f : \mathsf{G} \to \mathcal{A}$ denoted by $C_c(\mathsf{G}, \mathcal{A})$. On $C_c(\mathsf{G}, \mathcal{A})$ we define a product and $*$-operator by

$$\begin{aligned} (f * g)(x) &= \int_{\mathsf{G}} f(y)\alpha_y(g(y^{-1}x))\, dy , \\ f^*(x) &= \alpha_x(f(x^{-1}))^* . \end{aligned} \tag{84}$$

Then, by embedding $C_c(\mathsf{G}, \mathcal{A}) \hookrightarrow \mathcal{B}(L^2(\mathsf{G}, \mathcal{A}))$ through $f \mapsto T_f$, where $T_f g = f * g$, we can define $\mathcal{A} \times_\alpha \mathsf{G}$ as the completion of $C_c(\mathsf{G}, \mathcal{A})$ with respect to the operator norm on $\mathcal{B}(L^2(\mathsf{G}, \mathcal{A}))$.

The statements of T-duality, cf. Theorem 6, can now be formulated in this context as follows:

Theorem 7 *[54] Let $\mathcal{A}$ be a continuous C*-algebra with spectrum Y and Dixmier – Douady invariant $[H]$, then $\mathcal{A} \rtimes_\alpha \mathbb{R}$ is again a continuous-trace algebra with spectrum $\widehat{Y}$ and Dixmier – Douady invariant $[\widehat{H}]$ as given by Theorem 6. Moreover*

- *The C*-algebra $\mathcal{A} \rtimes_\alpha \mathbb{R}$ admits an action $\widehat{\alpha}$ of the dual group $\widehat{\mathbb{R}}$ and we have*

$$(\mathcal{A} \rtimes_\alpha \mathbb{R}) \rtimes_{\widehat{\alpha}} \widehat{\mathbb{R}} \cong \mathcal{A} \otimes \mathcal{K}(L^2(\mathbb{R})) , \tag{85}$$

  hence the C-algebra obtained by taking the crossed product twice is Morita equivalent to the original algebra (this is known as Takai duality).*
- *We have an isomorphism (known as the Connes – Thom isomorphism)*

$$K_\bullet(\mathcal{A} \rtimes \mathbb{R}) \cong K_{\bullet+1}(\mathcal{A}) . \tag{86}$$

The methods of this section can be generalized to higher rank principal torus bundles [9, 11, 12, 47, 48], but several complications arise. First of all the $\mathsf{T} = \mathbb{T}^n$-action on the principal torus bundle Y need not always lift to an $\mathbb{R}^n$-action on $\mathcal{A} = C(Y, \mathcal{E})$. Even if it does, this lift need not be unique. Second, the crossed product $\mathcal{A} \rtimes \mathbb{R}^n$ need not be continuous trace, but rather, might correspond to a field of noncommutative tori. A discussion of these results is beyond the scope of these lectures.

5.2.5 T-Duality as a Duality of Loop Group Bundles

In this section we will discuss a geometric reformulation of T-duality as a duality of loop group bundles and make contact with the classifying space approach of [17]. The discussion closely follows [14], which is based on earlier ideas in [6, 7, 28].

Our starting point is to geometrize the H-flux, as in the previous section, in terms of an (isomorphism class of) principal $\mathsf{PU}(\mathcal{H})$-bundle $\tilde{\pi} : P \to Y$ over the circle bundle $\pi : Y \to X$. We will shortly see that these geometrical data are equivalent to a principal $L\mathsf{PU}(\mathcal{H}) \rtimes \mathbb{T}$-bundle over X. Let us first recall the definition of the semi-direct product $L\mathsf{G} \rtimes \mathbb{T}$. First of all, $\mathbb{T}$ acts on $L\mathsf{G}$ by $\mathbb{T} \times L\mathsf{G} \to L\mathsf{G}\,, \quad (t, \gamma) \mapsto t \cdot \gamma$, where $(t \cdot \gamma)(s) = \gamma(ts)$. The semi-direct product $L\mathsf{G} \rtimes \mathbb{T}$ is then defined by the multiplication law $(\gamma_1, t_1) \circ (\gamma_2, t_2) = ((t_2 \cdot \gamma_1)\gamma_2, t_1 t_2)$. Equivalently, we can think of the semi-direct product $L\mathsf{G} \rtimes \mathbb{T}$ as the split short exact sequence

$$\begin{tikzcd} 1 \arrow[r] & L\mathsf{G} \arrow[r, "\iota"] & L\mathsf{G} \rtimes \mathbb{T} \arrow[r, "\rho"] & \mathbb{T} \arrow[r] \arrow[l, bend left, "s"] & 1 \end{tikzcd} \tag{87}$$

where for $(\gamma, t) \in L\mathsf{G} \rtimes \mathbb{T}$, we have $\rho(\gamma, t) = t$ and $\iota(\gamma) = (\gamma, 1)$. We will refer to ρ as the 'momentum homomorphism.'

We have

Theorem 8 *[3, 14, 53] Let* G *be a simply connected Lie group. We have a 1–1 correspondence between isomorphism classes of principal* G*-bundles* $\tilde{\pi} : P \to Y$ *over principal* $\mathbb{T}$*-bundles* $\pi : Y \to X$ *and isomorphism classes of principal* $L\mathsf{G} \rtimes \mathbb{T}$*-bundles* $\Pi : Q \to X$*, i.e.*

$$\left(\begin{tikzcd} \mathsf{G} \arrow[r] & P \arrow[d, "\tilde{\pi}"] \\ \mathbb{T} \arrow[r] & Y \arrow[d, "\pi"] \\ & X \end{tikzcd} \right) \cong \left(\begin{tikzcd} L\mathsf{G} \rtimes \mathbb{T} \arrow[r] & Q \arrow[d, "\Pi"] \\ & X \end{tikzcd} \right) \tag{88}$$

Proof We will first discuss the correspondence (88) explicitly, using transition functions. Let $\mathcal{U} = \{U_\alpha\}$ be a good cover of X for which we have trivializations $\phi_\alpha : \pi^{-1}(U_\alpha) \xrightarrow{\simeq} U_\alpha \times \mathbb{T}$. We write $\phi_\alpha(y) = (\pi(y), s_\alpha(y))$, where the 'section' $s_\alpha : \pi^{-1}(U_\alpha) \to \mathbb{T}$ satisfies $s_\alpha(yt) = s_\alpha(y)t$, for $t \in \mathbb{T}$ (group action written multiplicatively). The transition functions $g_{\alpha\beta} : U_{\alpha\beta} \to \mathbb{T}$ are defined by

$$\left(\phi_\alpha \circ \phi_\beta^{-1}\right)(x,t) = (x, g_{\alpha\beta}(x)t)\,, \quad \text{i.e.} \quad g_{\alpha\beta}(x) = s_\alpha(y)s_\beta(y)^{-1}\,, \tag{89}$$

where $y \in \pi^{-1}(x) \subset Y$. This definition does not depend on the choice of $y \in \pi^{-1}(x)$. The transition functions satisfy the cocycle identity, $g_{\alpha\beta}(x)g_{\beta\gamma}(x) = g_{\alpha\gamma}(x)$, $x \in U_{\alpha\beta\gamma}$. We also recall that the $\mathbb{T}$-bundle can be reconstructed from the transition functions by setting $E = \coprod_\alpha (U_\alpha \times \mathbb{T})/\sim$, where we identify $(x,t) \sim (x,t')$ on $U_{\alpha\beta} \times \mathbb{T}$ iff $t = g_{\alpha\beta}(x)t'$.

Let $V_\alpha = \pi^{-1}(U_\alpha)$. Since V_α is homotopic to $\mathbb{T}$ and since $\pi_1(\mathsf{G}) = 0$ by assumption, the G-bundle $\tilde{\pi} : P \to Y$ trivializes over V_α. Denote the local trivialization by $\tilde{\phi}_\alpha : \tilde{\pi}^{-1}(V_\alpha) \xrightarrow{\simeq} V_\alpha \times \mathsf{G}$, the corresponding section by $\tilde{s}_\alpha : \tilde{\pi}^{-1}(V_\alpha) \to \mathsf{G}$, and the transition functions by $\tilde{g}_{\alpha\beta} : V_{\alpha\beta} \to \mathsf{G}$.

We will now show how to use these data to define a principal $L\mathsf{G} \rtimes \mathbb{T}$-bundle Q over M. We define it by declaring that the transition functions $G_{\alpha\beta} : U_{\alpha\beta} \to L\mathsf{G} \times \mathbb{T}$ are given by

$$G_{\alpha\beta}(x) = \left(\tilde{g}_{\alpha\beta}\left(\phi_\beta^{-1}(x,\cdot)\right), g_{\alpha\beta}(x)\right)\,, \qquad x \in U_{\alpha\beta}\,. \tag{90}$$

Then, for $x \in U_{\alpha\beta\gamma}$, one has

$$\begin{aligned} G_{\alpha\beta}(x)G_{\beta\gamma}(x) &= \left(\tilde{g}_{\alpha\beta}\left(\phi_\beta^{-1}(x,\cdot)\right), g_{\alpha\beta}(x)\right) \circ \left(\tilde{g}_{\beta\gamma}\left(\phi_\gamma^{-1}(x,\cdot)\right), g_{\beta\gamma}(x)\right) \\ &= \left(g_{\beta\gamma}(x)\cdot\tilde{g}_{\alpha\beta}\left(\phi_\beta^{-1}(x,\cdot)\right)\tilde{g}_{\beta\gamma}\left(\phi_\gamma^{-1}(x,\cdot)\right), g_{\alpha\beta}(x)g_{\beta\gamma}(x)\right) \\ &= \left(\tilde{g}_{\alpha\beta}\left(\phi_\gamma^{-1}(x,\cdot)\right)\tilde{g}_{\beta\gamma}\left(\phi_\gamma^{-1}(x,\cdot)\right), g_{\alpha\beta}(x)g_{\beta\gamma}(x)\right) \\ &= \left(\tilde{g}_{\alpha\gamma}\left(\phi_\gamma^{-1}(x,\cdot)\right), g_{\alpha\gamma}(x)\right) = G_{\alpha\gamma}(x)\,, \end{aligned} \tag{91}$$

where we have used the cocycle properties of the transition functions $g_{\alpha\beta}$ and $\tilde{g}_{\alpha\beta}$. Conversely, given a principal $L\mathsf{G} \rtimes \mathbb{T}$-bundle Q over X with transition functions $G_{\alpha\beta} : U_{\alpha\beta} \to L\mathsf{G} \rtimes \mathbb{T}$, we can reconstruct the transition functions of a $\mathbb{T}$-bundle Y over X and the G-bundle P over Y as follows. First, we let $g_{\alpha\beta} = \rho(G_{\alpha\beta})$ (cf. (87)) be the transition function of the principal $\mathbb{T}$-bundle $\pi : Y \to X$ and $\tilde{g}_{\alpha\beta}(y) = \jmath(G_{\alpha\beta}(\pi(y)))(s_\beta(y))$ the transition function of the G-bundle, where $\jmath(\gamma,t) = \gamma$ is a left splitting of (87). This construction clearly is the inverse of the construction described above. $\square$

Now consider $\mathsf{G} = \mathsf{PU}(\mathcal{H})$. In that case one easily checks that since $\pi_2(\mathsf{PU}(\mathcal{H})) = \mathbb{Z}$ (as opposed to simply connected, compact Lie groups for which $\pi_2(\mathsf{G}) = 0$), we have $\pi_1(L\mathsf{G} \rtimes \mathbb{T}) = \mathbb{Z} \oplus \mathbb{Z}$. Hence, loosely speaking, we have two circles sitting inside $L\mathsf{G} \rtimes \mathbb{T}$. The first circle is recovered from the momentum homomorphism in (87), while the second circle can be recovered as follows. The group $\mathsf{G} = \mathsf{PU}(\mathcal{H})$ has a canonical central extension $\mathsf{U}(\mathcal{H})$

$$1 \longrightarrow \mathbb{T} \longrightarrow \mathsf{U}(\mathcal{H}) \longrightarrow \mathsf{PU}(\mathcal{H}) \longrightarrow 1 \tag{92}$$

The central extension (92) has a connection compatible with the group structure. The holonomy of this connection is a homomorphism hol : $L\mathsf{G} \to \mathbb{T}$. We let $\mathsf{N} = \mathrm{Ker}(\mathrm{hol} : L\mathsf{G} \to \mathbb{T})$, a normal subgroup of $L\mathsf{G}$. We thus have an exact sequence

$$1 \longrightarrow \mathsf{N} \rtimes \mathbb{T} \xrightarrow{\iota} L\mathsf{G} \rtimes \mathbb{T} \xrightarrow{\omega} \mathbb{T} \longrightarrow 1 \tag{93}$$

where ω, the winding homomorphism, is defined by $\omega(\gamma, t) = \mathrm{hol}(\gamma)$. If this were again a split exact sequence, like (93), we could apply the reconstruction theorem and obtain a dual $\mathsf{PU}(\mathcal{H})$-bundle over a dual $\mathbb{T}$-bundle. However, the sequence (93) is not split and hence $\mathsf{N} \rtimes \mathbb{T}$ is not isomorphic to $L\mathsf{G}$. From the exact sequence in homotopy it follows easily, however, that $\mathsf{N} \rtimes \mathbb{T}$ and $L\mathsf{G}$ are homotopy equivalent. After performing this homotopy, we can perform the reconstruction. The result being our T-dual circle bundle with T-dual H-flux described in the introduction. A proof of this statement follows from the work of Bunke and Schick [17]. In order to relate this discussion to [17], we need to reformulate the equivalence of Theorem 8 in terms of classifying spaces. It turns out that the classifying space of principal $L\mathsf{G} \rtimes \mathbb{T}$-bundles $B(L\mathsf{G} \rtimes \mathbb{T})$ is equal to $R = E\mathbb{T} \times_{\mathbb{T}} BL\mathsf{G}$, which arose in [17] as the classifying space of (equivalence classes of pairs $(Y, [H])$. This equivalence arises by chasing the following diagram (see [14] for details):

$$\begin{tikzcd}
 & & \mathbb{T} \arrow[d] & \mathbb{T} \arrow[d] \\
B\mathsf{G} & BL\mathsf{G} \arrow[l, "\widehat{\pi}"'] \arrow[d] & Y \arrow[l, "\psi"'] \arrow[ll, bend right] \arrow[r, "\widehat{\phi}"'] \arrow[d] & E\mathbb{T} \arrow[d] \\
 & B(L\mathsf{G} \rtimes \mathbb{T}) \arrow[rr, bend right, "\widehat{\pi}"'] & X \arrow[l, "\phi"'] \arrow[r, "\tilde{\phi}"] & B\mathbb{T}
\end{tikzcd}$$

While R has the natural structure of a principal $BL\mathsf{G} \cong (K(\mathbb{Z}, 3) \rtimes K(\mathbb{Z}, 2))$-bundle over $B\mathbb{T} = K(\mathbb{Z}, 2)$, there is another way of interpreting R, namely as a $K(\mathbb{Z}, 3)$ homotopy fibration over $K(\mathbb{Z}, 2) \times K(\mathbb{Z}, 2)$ [17, 47, 48]. Moreover, there is a map $T : R \to R$ such that $T^* : H^2(R, \mathbb{Z}) \to H^2(R, \mathbb{Z})$ exchanges the two generators, and $T \circ T$ is homotopic to the identity on R. It turns out that $T : R \to R$ implements

T-duality for principal circle bundles with background flux (see [17]): A principal $L\mathsf{G} \rtimes \mathbb{T}$-bundle Q over X has two natural characteristic classes of degree 2 on X. One of these is the first Chern class of the associated circle bundle over X, $c_1(Y)$, and the other is given by integration over the fibre of Y of the Dixmier – Douady invariant of P (i.e. the H-flux $[H] \in H^3(Y, \mathbb{Z})$). We denote these by $c(Q)$ and $d(Q)$, respectively, and they are the pullback under the classifying map $\phi : X \to R$ of the generators of $H^2(R, \mathbb{Z})$.

Hence, in terms of classifying maps, the T-dual principal $L\mathsf{G} \rtimes \mathbb{T}$-bundle $\widehat{Q}$ over X is defined by considering the continuous map $T \circ f : X \to B(L\mathsf{G} \rtimes \mathbb{T})$ and by associating to it $\widehat{Q} = (T \circ f)^*(E(L\mathsf{G} \rtimes \mathbb{T}))$. It follows that T-duality exchanges the entries of the pair $(c(Q), d(Q))$, and that T-duality applied twice gives a bundle that is isomorphic to Q, since $T \circ T \sim \mathrm{I}_R$. We summarize this as follows:

Theorem 9 *Let* $\mathsf{G} = \mathsf{PU}(\mathcal{H})$. *Given a principal* $L\mathsf{G} \rtimes \mathbb{T}$*-bundle,*

$$\begin{array}{ccc} L\mathsf{G} \rtimes \mathbb{T} & \longrightarrow & Q \\ & & \big\downarrow \Pi \\ & & X \end{array}$$

with classifying map $\phi : X \to B(L\mathsf{G} \rtimes \mathbb{T})$*, then there exists a T-dual principal* $L\mathsf{G} \rtimes \mathbb{T}$*-bundle,*

$$\begin{array}{ccc} L\mathsf{G} \rtimes \mathbb{T} & \longrightarrow & \widehat{Q} \\ & & \big\downarrow \widehat{\Pi} \\ & & X \end{array}$$

with classifying map $T \circ \phi : X \to B(L\mathsf{G} \rtimes \mathbb{T})$*, which has the following properties:*

1. $\widehat{\widehat{Q}}$ *is isomorphic to* Q*;*
2. $c(\widehat{Q}) = d(Q)$ *and* $d(\widehat{Q}) = c(Q)$.

To summarize, geometrically speaking, T-duality can be viewed as the exchange of the momentum and winding homomorphisms of (87) and (93).

5.3 Generalized Geometry

Aspects of generalized geometry occur in many papers in both the physics and mathematics literature since the 1980s, but the subject itself was only recently formalized by Hitchin [38] and further workedout by his students [23, 34, 35]. In this section, I will highlight several aspects of generalized geometry, mostly from an algebraic point of view, hopefully complementing the current literature on the subject.

5.3.1 Cartan Relations

Let M be a smooth (i.e. C^∞), d-dimensional, manifold, [8] TM its tangent bundle, T^*M its cotangent bundle, and let $\Gamma(TM)$ and $\Gamma(T^*M)$ denote the vectorspace of smooth sections of TM and T^*M, respectively (vectorfields and 1-forms). Alternatively, we can think of $\Gamma(TM)$ as the set of derivations of $C^\infty(M)$ as for $X \in \Gamma(TM)$, $f, g \in C^\infty(M)$, we have the Leibnitz rule

$$X(fg) = X(f)g + fX(g)\,. \tag{94}$$

We also have a Lie bracket $[\ ,\ ] : \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$, satisfying

(L1) $[X, Y] = -[Y, X]$,
(L2) $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$,
(L3) $[X, fY] = f[X, Y] + X(f)Y$,

for $X, Y, Z \in \Gamma(TM)$, $f \in C^\infty(M)$.

We have the following operations on forms $\Omega^k(M) = \Gamma(\wedge^k T^*M)$.

(a) A *differential* $d : \Omega^k(M) \to \Omega^{k+1}(M)$ (a.k.a. the *exterior differential*), defined by

$$\begin{aligned}(d\omega)(X_0, \dots, X_k) &= \sum_{i=0}^{k} (-1)^i X_i \omega(X_0, \dots, \widehat{X_i}, \dots, X_k) \\ &+ \sum_{0 \le i < j \le k} (-1)^{i+j} \omega([X_i, X_j], X_0, \dots, \widehat{X_i}, \dots, \widehat{X_j}, \dots, X_k).\end{aligned} \tag{95}$$

(b) A *contraction* $\imath_X : \Omega^k(M) \to \Omega^{k-1}(M)$ for $X \in \Gamma(TM)$ defined by

$$(\imath_X \omega)(X_1, \dots, X_{k-1}) = \omega(X, X_1, \dots, X_{k-1})\,. \tag{96}$$

(c) A *Lie derivative* $\mathcal{L}_X : \Omega^k(M) \to \Omega^k(M)$ defined by

$$\mathcal{L}_X = d\imath_X + \imath_X d \equiv \{d, \imath_X\}\,. \tag{97}$$

The various operations satisfy the following system of equations (Cartan formulas):

$$\begin{aligned}\{\imath_X, \imath_Y\} &= 0\,, \\ \{d, \imath_X\} &= \mathcal{L}_X\,, \\ [\mathcal{L}_X, \imath_Y] &= \imath_{[X,Y]}\,, \\ [\mathcal{L}_X, \mathcal{L}_Y] &= \mathcal{L}_{[X,Y]}\,, \\ [d, \mathcal{L}_X] &= 0\,.\end{aligned} \tag{98}$$

[8] We change the notation for a manifold from X to M, as compared to the previous sections, to avoid confusion with the notation for a vectorfield.

In fact, these equations express the fact that the Lie bracket $[\ ,\ ]$ is the *derived bracket* of the exterior differential [43]. All properties of the Lie bracket follow from the graded Jacobi identities for the operators $d, \imath_X$, and $\mathcal{L}_X$ of degrees $1, -1$, and 0, respectively. Recall that a graded Lie algebra is defined by the identities

$$\begin{aligned}
[A, B] &= -(-1)^{|A||B|}[B, A]\,, \\
[A, [B, C]] &= [[A, B], C] + (-1)^{|A||B|}[B, [A, C]]\,.
\end{aligned} \tag{99}$$

Thus, for instance, we have

$$\begin{aligned}
0 = [d, \{\imath_X, \imath_Y\}] &= [\{d, \imath_X\}, \imath_Y] - [\imath_X, \{d, \imath_Y\}] \\
&= [\mathcal{L}_X, \imath_Y] - [\imath_X, \mathcal{L}_Y] = \imath_{[X,Y]} + \imath_{[Y,X]}\,,
\end{aligned} \tag{100}$$

from which it follows, using the fact that $\imath_X\omega = 0$ for all $\omega \in \Omega(M)$ iff $X = 0$, that $[X, Y] = -[Y, X]$.

5.3.2 Lie Algebroids

In this section we briefly introduce the concept of a Lie algebroid. We refer to [49] for a complete treatment and historical account of the subject.

Definition 2 A *Lie algebroid* $(A, [\ ,\ ], \rho)$ over M is a vector bundle $A \to M$, together with a Lie bracket $[\ ,\ ]$ on ΓA and a bundle map ('anchor') $\rho : A \to TM$, such that

(A1) $\rho[X, Y] = [\rho(X), \rho(Y)]$, for all $X, Y \in \Gamma A$,
(A2) $[X, fY] = f[X, Y] + (\rho(X)f)\, Y$, for all $X, Y \in \Gamma A$, $f \in C^\infty(M)$.

Remark 1 Note, in fact, that requirement (A1) is superfluous as it follows from (A2) by evaluating $[X, [Y, fZ]]$ in two different ways, using the Jacobi identity and (A2), i.e. on the one hand

$$\begin{aligned}
[X, [Y, fZ]] &= [X, f[Y, Z]] + [X, (\rho(Y)f)Z] \\
&= f[X, [Y, Z]] + (\rho(X)f)[Y, Z] + (\rho(Y)f)[X, Z] + \rho(X)(\rho(Y)f)Z,
\end{aligned}$$

while on the other hand

$$\begin{aligned}
[X, [Y, fZ]] &= [[X, Y], fZ] + [Y, [X, fZ]] \\
&= f[[X, Y], Z] + (\rho([X, Y])f)Z + f[Y, [X, Z]] + (\rho(Y)f)[X, Z] \\
&\quad + (\rho(X)f)[Y, Z] \\
&\quad + \rho(Y)(\rho(X)f)Z \\
&= f[X, [Y, Z]] + (\rho([X, Y])f)Z + (\rho(Y)f)[X, Z] + (\rho(X)f)[Y, Z] \\
&\quad + (\rho(Y)(\rho(X)f))Z\,.
\end{aligned}$$

Equating the two expressions gives (A1).

Given a Lie algebroid $\mathcal{L} = (A, [\ ,\ ], \rho)$ over M, there exists a standard differential $d_A : \Omega^k(A) \to \Omega^{k+1}(A)$ on $\Omega^\bullet(A) = \Gamma(\wedge^\bullet A^*)$ given by an analogue of the Cartan formula (cf. (95))

$$
\begin{aligned}
(d_A\omega)(X_0, \ldots, X_k) &= \sum_{i=0}^{k} (-1)^i \rho(X_i)\omega(X_0, \ldots, \widehat{X_i}, \ldots, X_k) \\
&\quad + \sum_{0\leq i<j\leq k} (-1)^{i+j} \omega([X_i, X_j], X_0, \ldots, \widehat{X_i}, \ldots, \widehat{X_j}, \ldots, X_k), \qquad (101)
\end{aligned}
$$

as well as a contraction $\iota_X : \Omega^{k+1}(A) \to \Omega^k(A)$, for $X \in \Gamma A$, given by

$$
(\iota_X\omega)(X_0, \ldots, X_{k-1}) = \omega(X, X_0, \ldots, X_{k-1}). \qquad (102)
$$

Together they satisfy a system of Cartan formulas (98), where $\mathcal{L}_X = \{d_A, \iota_X\}$.

5.3.3 Generalized Geometry

5.3.3.1 Courant Bracket

We will now generalize familiar constructions on TM to the so-called generalized tangent bundle $\mathcal{T}M = TM \oplus T^*M$. This is known as *generalized geometry*.

The first observation is that there is a natural field of nondegenerate symmetric bilinear forms on sections of $\mathcal{T}M$, namely for $X + \xi, Y + \eta \in \Gamma(TM \oplus T^*M)$ we put

$$
\langle X + \xi, Y + \eta \rangle = \tfrac{1}{2}(\iota_X\eta + \iota_Y\xi). \qquad (103)
$$

The symmetry group of this form is the orthogonal group

$$
\mathsf{O}(TM \oplus T^*M) = \{\mathbb{A} \in \mathsf{GL}(TM \oplus T^*M) \mid \langle \mathbb{A}\cdot, \mathbb{A}\cdot \rangle = \langle \cdot, \cdot \rangle\}.
$$

Since the bilinear form has signature (d, d), we have $\mathsf{O}(TM \oplus T^*M) \cong \mathsf{O}(d, d)$. The Lie algebra

$$
\mathfrak{o}(TM \oplus T^*M) = \{\mathbb{Q} \in \mathsf{M}(TM \oplus T^*M) \mid \langle \mathbb{Q}\cdot, \cdot \rangle + \langle \cdot, \mathbb{Q}\cdot \rangle = 0\}
$$

consists of matrices of the form

$$
\mathbb{Q} = \begin{pmatrix} A & \beta \\ b & -A^T \end{pmatrix}, \qquad (104)
$$

where

$$\begin{aligned} A &: \Gamma(TM) \to \Gamma(TM)\,, & A^T &: \Gamma(T^*M) \to \Gamma(T^*M)\,, \\ \beta &: \Gamma(T^*M) \to \Gamma(TM)\,, & b &: \Gamma(TM) \to \Gamma(T^*M)\,, \end{aligned} \tag{105}$$

satisfy $\beta^T = -\beta$ and $b^T = -b$. Hence we can think of b as a 2-form $b \in \Gamma(\wedge^2 T^*M) = \Omega^2(M)$ by $\imath_X b = b(X)$, and similarly β as a bivector $\beta \in \Gamma(\wedge^2 TM)$. We thus see that generalized geometry, in particular, naturally incorporates 2-forms, i.e. B-fields. The finite transformations corresponding to b and β are given by

$$\begin{aligned} e^b &\equiv \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}, & e^b(X+\xi) &= X + \xi + \imath_X b\,, \\ e^\beta &\equiv \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}, & e^\beta(X+\xi) &= X + \imath_\xi \beta + \xi\,. \end{aligned} \tag{106}$$

We refer to e^b as a B-field transform.

The Courant bracket on $\Gamma(TM \oplus T^*M)$, which plays a similar role in generalized geometry as the Lie bracket on $\Gamma(TM)$, is defined as [25]

$$[\![X+\xi, Y+\eta]\!] = [X,Y] + \mathcal{L}_X\eta - \mathcal{L}_Y\xi - \tfrac{1}{2}d(\imath_X\eta - \imath_Y\xi). \tag{107}$$

The Courant bracket does not, in general, behave nicely under $\mathsf{O}(TM \oplus T^*M)$ transformations. However,

$$[\![e^b(X+\xi), e^b(Y+\eta)]\!] = e^b[\![X+\xi, Y+\eta]\!] - \imath_X\imath_Y db\,,$$

hence B-field transforms give rise to an automorphism of the Courant bracket iff $db = 0$, i.e. $b \in \Omega^2_{\mathrm{cl}}(M)$. The above computation suggests, however, to define a *twisted Courant bracket*, for $H \in \Omega^3_{\mathrm{cl}}(M)$, by

$$[\![X+\xi, Y+\eta]\!]_H = [\![X+\xi, Y+\eta]\!] + \imath_X\imath_Y H\,. \tag{108}$$

The (twisted) Courant bracket is obviously skew symmetric, but does not satisfy the other properties of a Lie bracket. Specifically

Theorem 10 *Let $A, B, C \in \Gamma(TM \oplus T^*M)$ and $f \in C^\infty(M)$. The Courant bracket satisfies the following properties:*

(i) $[\![A,B]\!]_H = -[\![B,A]\!]_H$
(ii) $\mathrm{Jac}(A,B,C) = [\![[\![A,B]\!]_H, C]\!]_H + [\![[\![B,C]\!]_H, A]\!]_H + [\![[\![C,A]\!]_H, B]\!]_H = d\,\mathrm{Nij}(A,B,C)$ *where*
$\mathrm{Nij}(A,B,C) = \frac{1}{3}\left(\langle[\![A,B]\!]_H, C\rangle + \langle[\![B,C]\!]_H, A\rangle + \langle[\![C,A]\!]_H, B\rangle\right)$ *is the Nijenhuis operator*

(iii) $[\![A, fB]\!]_H = f[\![A, B]\!]_H + (\rho(A)f)B - \langle A, B\rangle df$, *where* $\rho : TM \oplus T^*M \to TM$*, a morphism of vector bundles, is the projection onto the first factor*
(iv) $e^b\,[\![A, B]\!]_H = [\![e^b\,A, e^b\,B]\!]_{H+db}$

We recall that a subbundle $E \subset TM \oplus T^*M$ is called *isotropic* if $\langle A, B\rangle = 0$, for all $A, B \in \Gamma E$, and *involutive* if $[\![A, B]\!]_H \in \Gamma E$, for all $A, B \in \Gamma E$. E is called a *Dirac structure* if E is a maximal isotropic, involutive subbundle.

Properties (i)–(iii) now clearly imply the following:

Theorem 11 *If E is an isotropic, involutive subbundle of $TM \oplus T^*M$, then E is a Lie algebroid.*

5.3.3.2 Clifford Algebra on $TM \oplus T^*M$

In generalized geometry, the role of the contraction in the Cartan relations is taken by a certain representation of the Clifford algebra $\mathrm{Cliff}(TM \oplus T^*M)$. We recall

Definition 3 The Clifford algebra $\mathrm{Cliff}(TM \oplus T^*M)$ is the algebra with generators γ_A, $A \in \Gamma(TM \oplus T^*M)$ and relations

$$\{\gamma_A, \gamma_B\} = 2\langle A, B\rangle\,. \tag{109}$$

The following statement is verified by straightforward calculation

Lemma 2 *We have a representation of the Clifford algebra* $\mathrm{Cliff}(TM \oplus T^*M)$ *on* $\Omega^\bullet(M) = \Gamma(\wedge^\bullet T^*M)$ *given by*

$$\gamma_{X+\xi}\cdot\varphi = \iota_X\varphi + \xi\wedge\varphi\,, \qquad X + \xi \in \Gamma(TM \oplus T^*M),\ \varphi \in \Omega^\bullet(M)\,. \tag{110}$$

Thus we can identify spinors for $TM \oplus T^*M$ with forms $\Omega^\bullet(M)$.

Now, for a given $\varphi \in \Omega^\bullet(M)$, such that $\varphi \neq 0$ pointwise, we denote by

$$E_\varphi = \{X + \xi \in \Gamma(TM \oplus T^*M) \mid \gamma_{X+\xi}\cdot\varphi = 0\} \tag{111}$$

the annihilator bundle of $\varphi \in \Omega^\bullet(M)$. It is clear that E_φ is an isotropic subbundle of $TM \oplus T^*M$. Moreover,

Definition 4 The element $\varphi \in \Omega^\bullet(M)$ is called a *pure spinor* if E_φ is a maximally isotropic subbundle of $TM \oplus T^*M$.

We end by remarking that E_φ is involutive if $d_H\varphi = 0$, where as before $d_H = d + H\wedge$. This follows easily from the generalized Cartan relations to be discussed in the next section.

5.3.3.3 Courant Bracket as a Derived Bracket

In order to write the (twisted) Courant bracket as a derived bracket it is useful to introduce a closely related bracket on $\Gamma(TM \oplus T^*M)$, the so-called (twisted)

Dorfmann bracket or *Loday bracket*. For $X+\xi, Y+\eta \in \Gamma(TM \oplus T^*M)$, and $H \in \Omega^3_{\mathrm{cl}}(M)$, it is defined by

$$(X+\xi) \circ_H (Y+\eta) = [X,Y] + \mathcal{L}_X\eta - \iota_Y d\xi + \iota_X\iota_Y H \,. \tag{112}$$

It is related to the (twisted) Courant bracket by

$$A \circ_H B = [\![A,B]\!]_H + d\langle A,B\rangle \,, \qquad A, B \in \Gamma(TM \oplus T^*M) \,, \tag{113}$$

or, conversely,

$$[\![A,B]\!]_H = \tfrac{1}{2}(A \circ_H B - B \circ_H A) \,, \tag{114}$$

i.e. the Courant bracket is the skew symmetrization of the Dorfmann bracket. The Dorfmann bracket is not skew symmetric, but behaves better than the Courant bracket in the sense that it satisfies the properties (cf. Theorem 10)

(i) $A \circ_H (B \circ_H C)) = (A \circ_H B) \circ_H C + B \circ_H (A \circ_H C)$
(ii) $A \circ_H (fB) = f(A \circ_H B) + (\rho(A)f)B$

for $A, B, C \in \Gamma(TM \oplus T^*M)$ and $f \in C^\infty(M)$.

We can now formulate the relations analogous to (98), which establish the Courant algebra as a derived algebra [43, 44, 1, 40].

Theorem 12 *Let $A, B \in \Gamma(TM \oplus T^*M)$. Then, on $\Omega^\bullet(M)$ we have the following relations*

$$\begin{aligned} \{\gamma_A, \gamma_B\} &= 2\langle A,B\rangle \,, \\ \{d_H, \gamma_A\} &= \mathcal{L}_A \,, \\ [\mathcal{L}_A, \gamma_B] &= \gamma_{A\circ_H B} \,, \\ [\mathcal{L}_A, \mathcal{L}_B] &= \mathcal{L}_{A\circ_H B} = \mathcal{L}_{[\![A,B]\!]_H} \,, \\ [d_H, \mathcal{L}_A] &= 0 \,, \end{aligned} \tag{115}$$

where $\mathcal{L}_{X+\xi}\varphi = \mathcal{L}_X\varphi + (d\xi + \iota_X H) \wedge \varphi$.

Properties of the Dorfmann and Courant bracket (as summarized in, e.g. Theorem 10) follow straightforwardly from the Jacobi identities of the graded Lie algebra (115), as in Sect. 5.3.1.

Furthermore, anti-symmetrizing the third relation in (115) gives

$$\begin{aligned} \gamma_A\gamma_B \cdot d_H\varphi = d_H(\gamma_B\gamma_A \cdot \varphi) + \gamma_B \cdot d_H(\gamma_A \cdot \varphi) - \gamma_A \cdot d_H(\gamma_B \cdot \varphi) \\ + \gamma_{[\![A,B]\!]_H} \cdot \varphi \,, \end{aligned} \tag{116}$$

which in particular implies that the annihilator bundle E_φ of (111) is involutive if $d_H\varphi = 0$. More precisely, if we introduce a filtration of $\Omega^\bullet(M)$

$$F^0 \subset F^1 \subset \ldots \subset F^d \equiv \Omega^\bullet(M) \tag{117}$$

by

$$F^k = \{\psi \in \Omega^\bullet(M) \mid \gamma_{A_1} \ldots \gamma_{A_{k+1}} \cdot \psi = 0\,,\ \forall A_1, \ldots, A_{k+1} \in \Gamma(TM \oplus T^*M)\}\,, \tag{118}$$

then E_φ is involutive iff $d_H\varphi \in F^1$. [Note that, in general, $d_H(F^i) \subset F^{i+3}$ and $d_H(F^i) \subset F^{i+1}$ if E_φ is involutive.]

5.3.4 Courant Algebroids

The notion of a Courant algebroid was introduced in [45]. The following definition is taken from [44].

Definition 5 A *Courant algebroid* $(E, \circ, \langle\ ,\ \rangle, \rho)$ over M is a vector bundle $E \to M$, with a Loday bracket $\circ$ on ΓE, a morphism of vector bundles ('anchor') $\rho : E \to TM$, and a field of nondegenerate symmetric bilinear forms $\langle\ ,\ \rangle$ on the fibres of E satisfying[9]

(C1) $X \circ (Y \circ Z) = (X \circ Y) \circ Z + Y \circ (X \circ Z)$,
(C2) $\rho(X)\langle Y, Z\rangle = \langle X, Y \circ Z + Z \circ Y\rangle$,
(C3) $\rho(X)\langle Y, Z\rangle = \langle X \circ Y, Z\rangle + \langle Y, X \circ Z\rangle$,

for all $X, Y, Z \in \Gamma E$.

The Loday bracket in a Courant algebroid $\mathcal{C}$ is, in general, not skew symmetric. The skew symmetrization of the Loday bracket of $\mathcal{C}$ is known as the Courant bracket

$$[\![X, Y]\!] = \tfrac{1}{2}\,(X \circ Y - Y \circ X)\ . \tag{119}$$

We have the following [44, 59]:

Theorem 13 *In any Courant algebroid* $(E, \circ, \langle\ ,\ \rangle, \rho)$ *over* M *the following relations hold for the Loday bracket:*

(C4) $\rho(X \circ Y) = [\rho(X), \rho(Y)]$,
(C5) $X \circ fY = f(X \circ Y) + (\rho(X)f)Y$,
(C6) $X \circ Y + Y \circ X = 2\,D\langle X, Y\rangle$.

Here we have defined $D = \frac{1}{2}\rho^* d\ :\ C^\infty(M) \to \Gamma E$, *where* $\rho^*\ :\ \Gamma(T^*M) \to \Gamma E$ *is the adjoint of* ρ, *and we have identified* $E \cong E^*$ *under the isomorphism given by* $\langle\ ,\ \rangle$. *Equivalently,*

$$\langle Df, X\rangle = \frac{1}{2}\rho(X) f\ . \tag{120}$$

[9] We use the same notation for the map $\rho : \Gamma E \to \Gamma(TM)$ induced by $\rho : E \to TM$.

These translate into the following relations for the Courant bracket (119)

(C1′) $\rho([\![X,Y]\!]) = [\rho(X),\rho(Y)]$,
(C2′) $[\![X, fY]\!] = f[\![X,Y]\!] + (\rho(X)f)Y - \langle X,Y\rangle Df$,
(C3′) $\rho\, D = 0$*, or equivalently* $\langle Df, Dg\rangle = 0$,
(C4′) $[\![X,[\![Y,Z]\!]]\!] + [\![Y,[\![Z,X]\!]]\!] + [\![Z,[\![X,Y]\!]]\!] = D\,\mathrm{Nij}(X,Y,Z)$*, where* $\mathrm{Nij}(X,Y,Z) = \frac{1}{3}\left(\langle [\![X,Y]\!],Z\rangle + \langle [\![Y,Z]\!],X\rangle + \langle [\![Z,X]\!],Y\rangle\right)$ *is the Nijenhuis operator.*[10]

Proof

(C5) We prove (C5) by evaluating $\rho(X)\langle fY, Z\rangle$ in two different ways. First,

$$\begin{aligned}\rho(X)\langle fY,Z\rangle = \rho(X)(f\langle Y,Z\rangle) &= (\rho(X)f)\langle Y,Z\rangle + f(\rho(X)\langle Y,Z\rangle)\\ &= (\rho(X)f)\langle Y,Z\rangle + f\langle X\circ Y,Z\rangle + f\langle Y, X\circ Z\rangle\,,\end{aligned}$$

by using the derivation property of $\rho(X)$ on functions, and then (C3). On the other hand, using (C3), we have

$$\rho(X)\langle fY,Z\rangle = \langle X\circ fY, Z\rangle + f\langle Y, X\circ Z\rangle\,. \tag{121}$$

Equation (C5) then follows by virtue of the nondegenerateness of the bilinear form.

(C4) Equation (C4) follows by evaluating $(X\circ(Y\circ fZ))$ in two different ways (cf. Sect. 5.3.2), namely

$$\begin{aligned}(X\circ(Y\circ fZ)) &= X\circ f(Y\circ Z) + X\circ(\rho(Y)fZ)\\ &= f(X\circ(Y\circ Z)) + (\rho(X)f)(Y\circ Z) + (\rho(Y)f)(X\circ Z)\\ &\quad + \rho(X)(\rho(Y)f)Z,\end{aligned}$$

while on the other hand, using the Leibnitz property (C1)

$$\begin{aligned}(X\circ(Y\circ fZ)) &= ((X\circ Y)\circ fZ) + Y\circ(X\circ fZ)\\ &= f((X\circ Y)\circ Z) + \rho(X\circ Y)fZ + f(Y\circ(X\circ Z))\\ &\quad + (\rho(Y)f)(X\circ Z) + (\rho(X)f)(Y\circ Z) + \rho(Y)(\rho(X)f)Z.\end{aligned}$$

Equating the two results immediately gives (C4).

(C6) We have

$$\langle D\langle Y,Y\rangle, X\rangle = \frac{1}{2}\rho(X)\langle Y,Y\rangle = \langle X, Y\circ Y\rangle\,,$$

and thus $D\langle Y,Y\rangle = Y\circ Y$, which by polarization proves (C6).

[10] Note that Nij is not a tensor, e.g. $\mathrm{Nij}(fX,Y,Z) \neq \mathrm{Nij}(X,Y,Z)$.

It follows that

$$[\![X, Y]\!] = X \circ Y - D\langle X, Y\rangle \,. \tag{122}$$

The properties of the Courant bracket follow trivially from this relation, except for (C3$'$) which is obtained by applying ρ to (C6) and using (C4). Note that (C3$'$) in turn implies $\rho\,\rho^* = 0$, i.e. we have a complex $T^*M \to E \to TM$. □

The various maps fit into the following commutative diagram, in which the horizontal and vertical lines are complexes

$$\begin{tikzcd}
 & & & \Gamma(T^*M) \arrow[d, "\rho^*"] & & \\
0 \arrow[r] & \mathrm{Ker}\, D \arrow[r] & C^\infty(M) \arrow[r, "D"] \arrow[ur, "d"] & \Gamma E \arrow[r] \arrow[d, "\rho"] & \mathrm{Coker}\, D \arrow[r] & 0 \\
 & & & \Gamma(TM) & &
\end{tikzcd}$$

The horizontal complex has the structure of an L_∞-algebra (or, more precisely, L_2-algebra) [55, 56], where the maps $l_k : \wedge^k V \to V$, $k \geq 2$, are explicitly given by (here $X_i \in \Gamma E$, $f \in C^\infty(M)$).

$$\begin{aligned}
l_2(X_1 \wedge X_2) &= [\![X_1, X_2]\!], && \text{in degree } 0\,,\\
l_2(X_1 \wedge f) &= \langle X, Df\rangle, && \text{in degree } 1\,,\\
l_3(X_1 \wedge X_2 \wedge X_3) &= -\mathrm{Nij}(X_1, X_2, X_3), && \text{in degree } 0\,,
\end{aligned}$$

with all other l_k vanishing.

The proof is by straightforward calculation. The only nontrivial identity needed is

Lemma 3 *For $X \in \Gamma E$ and $f \in C^\infty(M)$, we have*

$$D\langle X, Df\rangle = [\![X, Df]\!] \,.$$

Proof We have

$$\begin{aligned}
\langle X \circ Df, Y\rangle &= \rho(X)\langle Df, Y\rangle - \langle Df, X \circ Y\rangle \\
&= \tfrac{1}{2}\rho(X)\rho(Y)f - \tfrac{1}{2}\rho(X \circ Y)f \\
&= \tfrac{1}{2}\rho(Y)\rho(X)f = 2\langle D\langle Df, X\rangle, Y\rangle \,,
\end{aligned}$$

thus

$$X \circ Df = 2D\langle X, Df\rangle \,.$$

Now, using (C6)

$$Df \circ X + X \circ Df = 2D\langle X, Df\rangle,$$

hence $Df \circ X = 0$. Putting things together, we have

$$[\![X, Df]\!] = \tfrac{1}{2}(X \circ Df - Df \circ X) = D\langle X, Df\rangle\,,$$

as claimed. □

5.3.4.1 Exact Courant Algebroids

Definition 6 An *exact Courant algebroid* is a Courant algebroid $(E, \circ, \langle\ ,\ \rangle, \rho)$ over M, which fits into a short exact sequence

$$0 \longrightarrow T^*M \xrightarrow{\rho^*} E \xrightarrow{\rho} TM \longrightarrow 0$$

Every exact Courant algebroid admits an isotropic splitting s, i.e. a bundle map $s : TM \to E$ such that $\rho s = 1$ and $\langle sX, sY\rangle = 0$ for all $X, Y \in \Gamma(TM)$. We say that two exact Courant algebroids are equivalent if they differ by a choice of isotropic splitting. Also note that the dual map $s^* : E \to T^*M$ provides a splitting on the left, i.e. $s^*\rho^* = 1$. An exact Courant algebroid is sometimes also called a *generalized tangent bundle* and denoted by $\mathcal{T}M$.

Lemma 4 *Let $(E, \circ, \langle\ ,\ \rangle, \rho)$ be a Courant algebroid over M. Then*

1. *$\rho^*(T^*M)$ is an isotropic subspace of E.*
2. *$[\![\ ,\ ]\!]\big|_{\rho^*(T^*M)} = 0$.*

Proof

1. We have $\langle \rho^*\xi, \rho^*\eta\rangle = \xi(\rho\rho^*\eta) = 0$ since $\rho\rho^* = 0$.
2. For $\xi, \eta \in \Gamma(T^*M)$, and $Z \in \Gamma E$, we have

$$\begin{aligned}\langle \rho^*\xi \circ \rho^*\eta, Z\rangle &= \rho(\rho^*\xi)\langle \rho^*\eta, Z\rangle - \langle \rho^*\eta, \rho^*\xi \circ Z\rangle \\ &= -\eta(\rho(\rho^*\xi \circ Z)) = -\eta([\rho\rho^*\xi, \rho(Z)]) = 0\,,\end{aligned}$$

and hence $\rho^*\xi \circ \rho^*\eta = 0$, and thus $[\![\rho^*\xi, \rho^*\eta]\!] = 0$, for all $\xi, \eta \in \Gamma(T^*M)$, which proves the lemma.

□

Theorem 14 [15, 18, 57, 58] *We have*

1. *Equivalence classes of exact Courant algebroids are classified by $H^3(M, \mathbb{R})$.*
2. *Under the identification $E \cong s(TM) \oplus \rho^*(T^*M)$, the bilinear form and Courant bracket on E reduce to the standard bilinear form and Courant bracket on $TM \oplus T^*M$, twisted by $H \in \Omega^3_{\mathrm{cl}}(M)$.*

Proof Let E be an exact Courant algebroid with isotropic splitting s, i.e.

$$\begin{tikzcd} 0 \arrow[r] & T^*M \arrow[r, "\rho^*"] & E \arrow[r, "\rho"] \arrow[l, bend left, "s^*"] & TM \arrow[r] \arrow[l, bend left, "s"] & 0 \end{tikzcd}$$

We have

$$\begin{aligned}\langle sX + \rho^*\xi, sY + \rho^*\eta\rangle &= \langle sX, sY\rangle + \langle \rho^*\xi, sY\rangle + \langle sX, \rho^*\eta\rangle + \langle \rho^*\xi, \rho^*\eta\rangle \\ &= \xi(\rho sY) + \eta(\rho sX) = \xi(Y) + \eta(X) = \iota_X\eta + \iota_Y\xi\,,\end{aligned}$$

where we have used Lemma 4. Next we have

$$(sX + \rho^*\xi)\circ(sY + \rho^*\eta) = sX\circ sY + sX\circ\rho^*\eta + \rho^*\xi\circ sY + \rho^*\xi\circ\rho^*\eta\,.$$

We evaluate the terms one by one. First of all

$$\rho^*\xi\circ\rho^*\eta = 0\,,$$

because of Lemma 4. The second term satisfies

$$\rho(sX\circ\rho^*\eta) = \rho sX\circ\rho\rho^*\eta = 0\,,$$

hence $sX\circ\rho^*\eta \in \rho^*(\Gamma(T^*M))$. Now, for $Z \in \Gamma(TM)$ we have

$$\begin{aligned} s^*(sX\circ\rho^*\eta)(Z) &= 2\langle sX\circ\rho^*\eta, sZ\rangle \\ &= 2(\rho sX)\langle\rho^*\eta, sZ\rangle - 2\langle\rho^*\eta, sX\circ sZ\rangle \\ &= X(\eta(Z)) - \eta([X,Z]) = \mathcal{L}_X\iota_Z\eta - \iota_{[X,Z]}\eta \\ &= \iota_Z\mathcal{L}_X\eta = (\mathcal{L}_X\eta)(Z)\,,\end{aligned}$$

hence $sX\circ\rho^*\eta = \rho^*(\mathcal{L}_X\eta)$. Similarly, for the third term, $\rho(\rho^*\xi\circ sY) = 0$ and

$$\begin{aligned} s^*(\rho^*\xi\circ sY)(Z) &= 2\langle\rho^*\xi\circ sY, sZ\rangle \\ &= -2\langle sY\circ\rho^*\xi, sZ\rangle + 4\langle D\langle sY, \rho^*\xi\rangle, sZ\rangle \\ &= -\mathcal{L}_Y\xi(Z) + \iota_Z d\iota_Y\xi = -(\iota_Y d\xi)(Z)\,,\end{aligned}$$

hence $\rho^*\xi\circ sY = -\rho^*(\iota_Y d\xi)$. It remains to compute the first term. We have

$$\rho(sX\circ sY) = [\rho sX, \rho sY] = [X, Y]\,,$$

but in general $sX\circ sY \neq s([X,Y])$ or, equivalently, $s^*(sX\circ sY) \neq 0$. Define $H(X,Y) \in \Gamma(T^*M)$ by

$$H(X,Y) = s^*(sX\circ sY)\,.$$

We claim that H is $C^\infty(M)$-linear and that $H(X,Y)(Z)$ is totally antisymmetric. First of all, to prove linearity,

$$H(X, fY) = fs^*(sX \circ sY) + s^*(X(f)sY) = fs^*(sX \circ sY) = fH(X,Y)\,,$$

where we have used $s^*s = 0$. To prove antisymmetry

$$\begin{aligned}
&H(X,Y) + H(Y,X) = s^*(sX \circ sY + sY \circ sX) = s^*D\langle sX, sY\rangle = 0\\
&H(X,Y)(Z) = 2\langle sX \circ sY, sZ\rangle = 2\rho(sX)\langle sY, sZ\rangle\\
&\quad -2\langle sY, sX \circ sZ\rangle = -H(X,Z)(Y)\,.
\end{aligned}$$

We conclude that

$$sX \circ sY = s([X,Y]) - \rho^*(\iota_X \iota_Y H)$$

for some $H \in \Omega^3$. One can check that for $A, B, C \in s(\Gamma(TM)) \oplus \rho^*(\Gamma(T^*M))$,

$$(A \circ (B \circ C)) = ((A \circ B) \circ C) + (B \circ (A \circ C)) + \iota_{\rho(A)}\iota_{\rho(B)}\iota_{\rho(C)} dH,$$

it follows that $dH = 0$. Now suppose we choose a different splitting. Say we have $s_i : TM \to E$, $i = 1, 2$, such that $\rho(s_1 - s_2) = 0$. Then the exactness implies that there exists a unique $B(X) \in \Gamma(T^*M)$ such that $s_1(X) - s_2(X) = \rho^*(B(X))$. We conclude that $B(X) = -s_1^* s_2(X) = s_2^* s_1(X)$, which implies that $B(X)(Y)$ is antisymmetric. Clearly, by its definition, B is $C^\infty(M)$-linear, so that $B \in \Omega^2(M)$.

Now, we show that

$$dB = H_1 - H_2\,,$$

where H_i is the closed 3-form corresponding to the splitting s_i, $i = 1, 2$. Indeed, from $\rho^* H_i(X,Y) = s_i X \circ s_i Y - s_i[X,Y]$, it follows

$$\begin{aligned}
\rho^*(H_1(X,Y) - H_2(X,Y)) &= s_1 X \circ s_1 Y - s_2 X \circ s_2 Y - \rho^* B([X,Y])\\
&= s_2 X \circ \rho^* B(Y) + \rho^* B(X) \circ s_2 Y - \rho^* B([X,Y])\,,
\end{aligned}$$

where we have used $s_1(X) - s_2(X) = \rho^*(B(X))$. Therefore

$$\begin{aligned}
H_1(X,Y,Z) \quad - \quad & H_2(X,Y,Z) = s_2^* \rho^*(H_1(X,Y) - H_2(X,Y))(Z)\\
= \quad & s_2^*(s_2 X \circ \rho^* B(Y))(Z) + s_2^*(\rho^* B(X) \circ s_2 Y)(Z)\\
- \ & s_2^*(\rho^* B([X,Y]))(Z)\\
= \quad & 2\langle s_2 X \circ \rho^* B(Y), s_2 Z\rangle + 2\langle \rho^* B(X) \circ s_2 Y, s_2 Z\rangle - B([X,Y],Z)\,.
\end{aligned}$$

We evaluate the first two terms on the right-hand side separately

$$\begin{aligned}
\langle s_2X \circ \rho^*B(Y), s_2Z\rangle &= X(B(Y,Z)) - 2\langle \rho^*B(Y), s_2X \circ s_2Z\rangle \\
&= X(B(Y,Z)) - B(Y,[X,Z]), \\
2\langle \rho^*B(X) \circ s_2Y, s_2Z\rangle &= -2\langle s_2Y \circ \rho^*B(X), s_2Z\rangle + 4\langle D\langle \rho^*B(X), s_2Y\rangle, s_2Z\rangle \\
&= -Y(B(X,Z)) + B(X,[Y,Z]) + 2\langle d(B(X,Y)), Z\rangle \\
&= -Y(B(X,Z)) + B(X,[Y,Z]) + Z(B(X,Y))\,.
\end{aligned}$$

Combining all terms proves the claim, i.e.

$$\begin{aligned}
H_1(X,Y,Z) - H_2(X,Y,Z) &= X(B(Y,Z)) - Y(B(X,Z)) + Z(B(X,Y)) \\
&\quad - B([X,Y],Z) + B([X,Z],Y) - B([Y,Z],X) \\
&= dB(X,Y,Z)\,.
\end{aligned}$$

Finally, let us now, given a representative of a class $[H] \in H^3(M,\mathbb{R})$, explicitly construct the Courant algebroid. We choose an open cover and a representative of $[H]$ in $H^3_D(M,\mathbb{R}) \cong H^3_{\mathrm{dR}}(M,\mathbb{R})$, i.e. a four-tuple $(\Lambda_{\alpha\beta\gamma}, A_{\alpha\beta}, B_\alpha, H)$ in the Čech – de Rham complex (over $\mathbb{R}$). We then construct E by means of the *clutching construction*

$$E = \coprod_\alpha (TM \oplus T^*M)_{|U_\alpha} / \sim,$$

where we identify $X+\xi \in \Gamma(TM \oplus T^*M)_{|U_\alpha}$ with $Y+\eta \in \Gamma(TM \oplus T^*M)_{|U_\beta}$ on overlaps $U_{\alpha\beta}$ iff $Y = X$ and $\eta = \xi + \iota_X dA_{\alpha\beta}$. Consistency on triple overlaps $U_{\alpha\beta\gamma}$ follows from $dA_{\beta\gamma} - dA_{\alpha\gamma} + dA_{\alpha\beta} = (d\delta A)_{\alpha\beta\gamma} = (d^2\Lambda)_{\alpha\beta\gamma} = 0$. On $TM_{|U_\alpha}$ the splitting is given by

$$s_{|U_\alpha} \;:\; X \mapsto X + \iota_X B_\alpha \tag{123}$$

and consistency on overlaps follows from $B_\beta - B_\alpha = dA_{\alpha\beta}$. Moreover, since (123) is just a B-transform (cf. (106)) we see that the Courant bracket on E is simply the H-twisted Courant bracket (108). □

5.3.5 Generalized Complex Geometry

We briefly introduce some of the concepts of *generalized complex geometry* with the aim of showing that T-duality acts naturally on all these structures.

5.3.5.1 Generalized Complex Structures

We recall that an *almost complex structure* on a manifold M is an endomorphism $J : TM \to TM$, such that $J^2 = -1$. Given an almost complex structure we can decompose the complexified manifold $TM_{\mathbb{C}} = TM \otimes \mathbb{C}$ as

$$TM_{\mathbb{C}} = T^{(1,0)}M \oplus T^{(0,1)}M, \tag{124}$$

where $T^{(1,0)}M = \mathrm{Ker}(J - i)$ and $T^{(0,1)}M = \mathrm{Ker}(J + i)$ are the $\pm i$ eigenspaces of J. We refer to elements of $\Gamma(T^{(1,0)}M)$ and $\Gamma(T^{(0,1)}M)$ as holomorphic and anti-holomorphic vectorfields, respectively. Similarly, we can decompose

$$\Omega^k(M) = \bigoplus_{p+q=k} \Omega^{p,q}(M), \qquad \Omega^{p,q}(M) = \Gamma(\wedge^p T^{(1,0)*}M \oplus \wedge^q T^{(0,1)*}M). \tag{125}$$

Definition 7 A *complex structure* is an almost complex structure J which is integrable with respect to the Lie bracket on $\Gamma(TM)$, i.e. which is such that $[X, Y] \in \Gamma(T^{(1,0)}M)$, for all $X, Y \in \Gamma(T^{(1,0)}M)$, and $[X, Y] \in \Gamma(T^{(0,1)}M)$, for all $X, Y \in \Gamma(T^{(0,1)}M)$.

If we denote by $\Pi_\pm = \frac{1}{2}(1 \mp iJ)$ the projection operators on the $J = \pm i$ eigenspaces, integrability of the almost complex structure is equivalent to the statement that, for all $X, Y \in \Gamma(TM)$, we have

$$\Pi_\mp[\Pi_\pm X, \Pi_\pm Y] = 0, \tag{126}$$

or, equivalently,

$$\mathrm{N}(X, Y) = [X, Y] - [JX, JY] + J[X, JY] + J[JX, Y] = 0. \tag{127}$$

The tensor $\mathrm{N} \in \Gamma(\wedge^2 T^*M \otimes TM)$ is called the *Nijenhuis tensor* in complex geometry.

The generalization of these considerations to generalized geometry is straightforward. First

Definition 8 A *generalized almost complex structure* is a $\mathbb{J} \in \mathrm{O}(TM \oplus T^*M)$ satisfying $\mathbb{J}^2 = -1$.

Given a generalized almost complex structure we can decompose

$$(TM \oplus T^*M) \otimes \mathbb{C} = E_+ \oplus E_-, \tag{128}$$

where $E_\pm = \mathrm{Ker}(1 \pm i\mathbb{J})$ are the $\pm i$ eigenspaces of $\mathbb{J}$. Both E_+ and E_- are maximally isotropic eigenspaces of $(TM \oplus T^*M) \otimes \mathbb{C}$ since $\mathbb{J} \in \mathrm{O}(TM \oplus T^*M)$.

There are two main examples of generalized almost complex structures

(i) If J is an almost complex structure, then

$$\mathbb{J}_J = \begin{pmatrix} -J & 0 \\ 0 & J^T \end{pmatrix} \tag{129}$$

is a generalized almost complex structure and $E_+ = T^{(0,1)}M \oplus T^{(1,0)*}M$.

(ii) If ω is a nondegenerate 2-form, then

$$\mathbb{J}_\omega = \begin{pmatrix} 0 & -\omega^{-1} \\ \omega & 0 \end{pmatrix} \tag{130}$$

is a generalized almost complex structure.

Definition 9 A *(twisted) generalized complex structure* is a generalized almost complex structure $\mathbb{J}$ which is integrable with respect to the (twisted) Courant bracket on $\Gamma(TM \oplus T^*M)$.

We have an equation identical to (126) for $\mathbb{J}$ to be a generalized complex structure, where now $\Pi_\pm$ are the projection operators on $E_\pm$.

Examples of generalized complex structures are

(i) $\mathbb{J}_J$ in (129) is a generalized complex structure if J is a complex structure.
(ii) $\mathbb{J}_\omega$ in (130) is a generalized complex structure if $d\omega = 0$, i.e. if M has a *symplectic structure* with symplectic form ω.

5.3.5.2 Generalized Kähler Structures

The last structure we want to briefly discuss is that of a generalized Kähler structure. We recall

Definition 10 A *Hermitian structure* on a manifold M is

(a) an almost complex structure J,
(b) a Hermitian metric $g : TM \times TM \to \mathbb{R}$, i.e. a metric such that $g(JX, JY) = g(X, Y)$, for all $X, Y \in \Gamma(TM)$.

Given a Hermitian structure on M we can define a nondegenerate 2-form ω on M by

$$\omega(X, Y) = g(JX, Y)\,.$$

The compatibility of the various structures is expressed in terms of commutativity of the following diagram

$$\begin{tikzcd} TM \arrow[rr, "g"] & & T^*M \\ & TM \arrow[ul, "J"] \arrow[ur, "\omega"'] & \end{tikzcd} \tag{131}$$

Finally,

Definition 11 A *Kähler structure* on M is a triple (g, J, ω) of a Hermitian metric g, a complex structure J, and a symplectic form ω, compatible in the sense of (131).

In order to define a generalized Kähler structure we first need the concept of a generalized metric.

Definition 12 A *generalized metric* on $\mathcal{T}M = TM \oplus T^*M$ is a self-adjoint $\mathbb{G} \in \mathsf{O}(TM \oplus T^*M)$, such that $\langle \mathbb{G}A, A\rangle > 0$ for all $A \in \Gamma(TM \oplus T^*M)$, $A \neq 0$.

A generalized metric satisfies $\mathbb{G}^2 = \mathbb{G}\mathbb{G}^T = 1$. If we let $C_\pm = \mathrm{Ker}(\mathbb{G} \mp 1)$, then we have bundle isomorphisms $\pi_\pm : C_\pm \to TM$ (since TM is maximally isotropic). We can then define $g(X, Y) = \langle \pi_+^{-1}(X), \pi_+^{-1}(Y)\rangle$. In fact, one can find a $b \in \Omega^2(M)$ such that

$$C_+ = \mathrm{graph}_{g+b} = \{X + (g + b)(X) \mid X \in \Gamma(TM)\}. \tag{132}$$

This corresponds to a generalized metric

$$\mathbb{G}^b = \begin{pmatrix} -g^{-1}b & g^{-1} \\ g - bg^{-1}b & bg^{-1} \end{pmatrix}, \tag{133}$$

which is B-transform of

$$\mathbb{G} = \begin{pmatrix} 0 & g^{-1} \\ g & 0 \end{pmatrix}, \tag{134}$$

i.e. $\mathbb{G}^b = e^b \mathbb{G} e^{-b}$.

Finally, we have

Definition 13 A *generalized Kähler structure* is a triple $(\mathbb{G}, \mathbb{J}_1, \mathbb{J}_2)$ of orthogonal maps $\mathcal{T}M \to \mathcal{T}M$ such that $\mathbb{J}_1$ and $\mathbb{J}_2$ define a pair of commuting generalized complex structures and $\mathbb{G} = -\mathbb{J}_1\mathbb{J}_2$ is a generalized metric on $\mathcal{T}M$.

We summarize this definition by the following diagram:

$$\begin{tikzcd} \mathcal{T}M \arrow[rr, "\mathbb{G}"] & & \mathcal{T}M \\ & \mathcal{T}M \arrow[ul, "\mathbb{J}_1"] \arrow[ur, "\mathbb{J}_2"'] & \end{tikzcd}$$

An example of a generalized Kähler structure is, of course, an ordinary Kähler structure, where we take $\mathbb{J}_1 = \mathbb{J}_J$ and $\mathbb{J}_2 = \mathbb{J}_\omega$. In that case

$$\mathbb{G} = -\mathbb{J}_J\mathbb{J}_\omega = \begin{pmatrix} -J & 0 \\ 0 & J^T \end{pmatrix}\begin{pmatrix} 0 & -\omega^{-1} \\ \omega & 0 \end{pmatrix} = \begin{pmatrix} 0 & g^{-1} \\ g & 0 \end{pmatrix},$$

which is manifestly positive definite with respect to $\langle\ ,\ \rangle$, as $\langle \mathbb{G}(X + \xi), X + \xi\rangle = g(X, X) + g^{-1}(\xi, \xi)$.

5.3.6 T-Duality in Generalized Geometry

In the final section we show that generalized geometry provides a natural framework to discuss T-duality (see, for example, [2, 26, 27, 31–33] for related considerations). We start with the dimensionally reduced framework of Sect. 5.2.3. If we choose a connection A on the principal circle bundle $\pi : E \to M$ and 'dimensionally reduce' $\Omega \in \Omega^k(E)_{S^1}$ and $(X, \Xi) \in \Gamma(TE \oplus T^*E)_{S^1}$ as in

$$\Omega = \Omega_{(k)} + A \wedge \Omega_{(k-1)}\,, \qquad X = x + f\partial_A\,, \qquad \Xi = \xi + gA, \tag{135}$$

where $\Omega_{(k)}, \Omega_{(k-1)} \in \Omega^\bullet(M)$, $x \in \Gamma(TM)$, $\xi \in \Omega^1(M)$, $f, g \in C^\infty(M)$, and similarly for $\widehat{E}$, then we have isomorphisms $\tau : \Omega^\bullet(E)_{S^1} \to \Omega^\bullet(\widehat{E})_{S^1}$ and $\phi : \Gamma(TE \oplus T^*E)_{S^1} \to \Gamma(T\widehat{E} \oplus T^*\widehat{E})_{S^1}$ given by

$$\begin{aligned} \tau(\Omega_{(k)} + A \wedge \Omega_{(k-1)}) &= -\Omega_{(k-1)} + \widehat{A} \wedge \Omega_{(k)}\,, \\ \phi(x, f; \xi, g) &= (x, g; \xi, f)\,. \end{aligned} \tag{136}$$

Theorem 15 *[22, 23, 34, 35] We have*

(a)The map τ induces a chain map on the differential complexes $(\Omega^\bullet(E)_{S^1}, d_H) \to (\Omega^\bullet(\widehat{E})_{S^1}, d_{\widehat{H}})$, i.e. $\tau \circ d_H = -d_{\widehat{H}} \circ \tau$ and, hence, an isomorphism on twisted cohomology.
(b)The map ϕ is orthogonal with respect to the pairing on $TE \oplus T^*E$, hence induces an isomorphism of Clifford algebras.
(c)For $B \in \Gamma(TE \oplus T^*E)_{S^1}$ we have $\tau(\gamma_B \cdot \Omega) = \gamma_{\phi(B)} \cdot \tau(\Omega)$, hence τ induces an isomorphism of Clifford modules $\tau : \Omega^\bullet(E)_{S^1} \to \Omega^\bullet(\widehat{E})_{S^1}$.
(d)ϕ preserves the twisted Courant bracket.

Proof The proof of (a)–(c) is by straightforward verification (see Sect. 5.2.3 for the proof of (a)). Part (d) follows from the defining relations of the Courant bracket as a derived bracket (cf. Theorem 12), but can also easily be observed by writing out the Courant bracket in the dimensionally reduced formalism, i.e.

$$\begin{aligned} &[\![(x_1, f_1; \xi_1, g_1), (x_2, f_2; \xi_2, g_2)]\!]_{F,H} = ([x_1, x_2], x_1(f_2) - x_2(f_1) + \iota_{x_1}\iota_{x_2}F; \\ &(\mathcal{L}_{x_1}\xi_2 - \mathcal{L}_{x_2}\xi_1) + (g_2\iota_{x_1}F - g_1\iota_{x_2}F) - \tfrac{1}{2}d(\iota_{x_1}\xi_2 - \iota_{x_2}\xi_1) \\ &\quad + \tfrac{1}{2}(df_1 g_2 + f_2 dg_1 - f_1 dg_2 - df_2 g_1) + \iota_{x_1}\iota_{x_2}H_{(3)} + (f_2\iota_{x_1}H_2 - f_1\iota_{x_2}H_{2)}), \\ &x_1(g_2) - x_2(g_1) + \iota_{x_1}\iota_{x_2}H_{(2)}), \end{aligned}$$

where $F = dA$ and $H = H_{(3)} + A \wedge H_{(2)}$. □

Since T-duality preserves both the bilinear form and the Courant bracket, it preserves properties such as subbundles being isotropic and involutive. Thus the theorem implies that T-duality preserves all the important structures present in generalized geometry (such as generalized complex structures, generalized Kähler structures) and that therefore generalized geometry is a convenient framework for T-duality.

Finally, we show that the Buscher rules have a natural interpretation in the context of generalized geometry as well. Consider thereto the generalized metric $\mathbb{G}^b$ in (133) or equivalently the graph of $Q = g + b$ in (133). Decomposing Q as in (46), the graph is given in local coordinates by

$$\begin{pmatrix} X^\mu \\ X^m \\ Q_{\mu N} X^N \\ Q_{mN} X^N \end{pmatrix} = \begin{pmatrix} X^\mu \\ X^m \\ Q_{\mu\nu} X^\nu + Q_{\mu n} X^n \\ Q_{m\nu} X^\nu + Q_{mn} X^n \end{pmatrix} .$$

Upon performing a T-duality transformation analogous to ϕ in (136), the graph transforms to

$$\begin{pmatrix} X^\mu \\ Q_{m\nu} X^\nu + Q_{mn} X^n \\ Q_{\mu\nu} X^\nu + Q_{\mu n} X^n \\ X^m \end{pmatrix} = \begin{pmatrix} N \cdot X \\ M \cdot X \end{pmatrix}, \tag{137}$$

where

$$M = \begin{pmatrix} Q_{\mu\nu} & Q_{\mu n} \\ 0 & 1 \end{pmatrix}, \qquad N = \begin{pmatrix} 1 & 0 \\ Q_{m\nu} & Q_{mn} \end{pmatrix} .$$

Now observe that the right-hand side of (137) is the graph of

$$MN^{-1} = \begin{pmatrix} Q_{\mu\nu} - Q_{\mu n}(Q^{-1})_{mn} Q_{n\nu} & Q_{\mu m}(Q^{-1})_{mn} \\ -(Q^{-1})_{mn} Q_{n\nu} & (Q^{-1})_{mn} \end{pmatrix} = \widehat{Q}, \tag{138}$$

which are precisely the Buscher rules of (51). This shows that under T-duality the generalized metric transforms according to the Buscher rules!

A generalization of the results in this section to higher rank principal torus bundles will appear in [8] (see also [42]).

Acknowledgements I would like to thank the organizers of the summer school, Bernhelm Booß-Bavnbek, Giampiero Esposito, and Matthias Lesch, for the invitation to lecture and their persistent encouragement and patience while writing up my lecture notes. I would also like to thank all participants for the pleasant atmosphere and all the feedback. This research was supported under the Australian Research Council's Discovery Projects scheme.

References

1. A. Alekseev, P. Xu, *Courant algebroids and derived brackets*, unfinished manuscript, available from `http://www.math.psu.edu/ping/papers.html`.
2. O. Ben-Bassat, *Mirror symmetry and generalized complex manifolds*, J. Geom. Phys. **56** (2006) 533–558 [`arXiv:math.AG/0405303`].
3. A. Bergman, U. Varadarajan, *Loop groups, Kaluza-Klein reduction and M-theory*, J. High Energy Phys. **06** (2005) 043 [`arXiv:hep-th/0406218`].
4. R. Bott, L.W. Tu, *Differential Forms in Algebraic Topology*, Springer Verlag, New York (1982).
5. P. Bouwknegt, A. Carey, V. Mathai, M. Murray, D. Stevenson, *Twisted K-theory and K-theory of bundle gerbes*, Comm. Math. Phys. **228** (2002) 17–45 [`arXiv:hep-th/0106194`].
6. P. Bouwknegt, J. Evslin, V. Mathai, *T-duality: Topology change from H-flux*, Commun. Math. Phys. **249** (2004) 383–415 [`arXiv:hep-th/0306062`].
7. P. Bouwknegt, J. Evslin, V. Mathai, *On the topology and H-flux of T-dual manifolds*, Phys. Rev. Lett. **92** (2004) 181601 [`arXiv:hep-th/0312052`].
8. P. Bouwknegt, J. Garretson, P. Kao, *Generalized fluxes and T-duality in generalized geometry*, in preparation.
9. P. Bouwknegt, K. Hannabuss, V. Mathai, *T-duality for principal torus bundles*, J. High Energy Phys. **03** (2004) 018 [`arXiv:hep-th/0312284`].
10. P. Bouwknegt, K. Hannabuss, V. Mathai, *T-duality for principal torus bundles and dimensionally reduced Gysin sequences*, Adv. Theor. Math. Phys. **9** (2005) 749–773, [`arXiv:hep-th/0412268`].
11. P. Bouwknegt, K. Hannabuss, V. Mathai, *Nonassociative tori and applications to T-duality*, Commun. Math. Phys. **264** (2006) 41–69 [`arXiv:hep-th/0412092`].
12. P. Bouwknegt, K. Hannabuss, V. Mathai, *C^*-algebras in tensor categories*, To appear in the proceedings of the conference “Motives, Quantum Field Theory, and Pseudodifferential Operator”, Boston, 2–13 June 2008, Clay Mathematics Proceedings, vol. **12**, A. Carey, D. Ellwood, S. Paycha, S. Rosenberg (eds.) [`arXiv:math.QA/0702802`].
13. P. Bouwknegt, V. Mathai, *D-branes, B-fields and twisted K-theory*, J. High Energy Phys. **03** (2000) 007 [`arXiv:hep-th/0002023`].
14. P. Bouwknegt, V. Mathai, *T-duality as a duality of loop group bundles*, J. Phys. A: Math. Theor. **42** (2009) 162001 [`arXiv:0902.4341 [hep-th]`].
15. P. Bressler, A. Chervov, *Courant algebroids*, J. Math. Sciences **128** (2005) 3030–3053 [`arXiv:hep-th/0212195`].
16. J.-L. Brylinski, *Loop spaces, characteristic classes and geometric quantization*, Prog. Math. **107** Birhäuser, Boston (1993).
17. U. Bunke, T. Schick, *On the topology of T-duality*, Rev. Math. Phys. **17** (2005) 77–112, [`arXiv:math/0405132`].
18. H. Bursztyn, M. Crainic, *Dirac geometry, quasi-Poisson actions and D/G-valued moment maps*, J. Diff. Geom. **82** (2009) 501–566 [`arXiv:0710.0639 [math.DG]`].
19. T. Buscher, *A symmetry of the string background field equations*, Phys. Lett. **B194** (1987) 59-62.
20. T. Buscher, *Path integral derivation of quantum duality in nonlinear sigma models*, Phys. Lett. **B201** (1988) 466–472.
21. A. Carey, S. Johnson, M. Murray, *Holonomy on D-branes*, J. Geom. Phys. **52** (2004) 186–216 [`arXiv:hep-th/0204199`].
22. G. Cavalcanti, M. Gualtieri, *T-duality with NS-flux and generalized complex structures*, unpublished manuscript, 2007.
23. G. Cavalcanti, *New aspects of the dd^c-lemma*, PhD thesis, University of Oxford, 2004 [`arXiv:math.DG/0501406`].
24. J. Cheeger and J. Simons, *Differential characters and geometric invariants*, Lecture Notes in Math. **1167** (1985) 50–80.
25. T.J. Courant, *Dirac manifolds*, Trans. A.M.S. **319** (1990) 631–661.

26. A. Dabholkar, C. Hull, *Generalised T-duality and non-geometric backgrounds*, J. High Energy Phys. **05** (2006) 009 [arXiv:hep-th/0512005].
27. I. Ellwood, *NS-NS fluxes in Hitchin's generalized geometry*, J. High Energy Phys. **12** (2007) 084 [arXiv:hep-th/0612100].
28. J. Evslin, *Twisted K-theory from monodromies*, J. High Energy Phys. **11** (2003) 037 [arXiv:hep-th/0302081].
29. D. Freed, *Classical Chern-Simons theory, Part II*, Houston J. Math. **28** (2002) 293–310.
30. A. Giveon, M. Porrati, E. Rabinovici, *Target space duality in String Theory*, Phys. Rept. **244** (1994) 77–202 [arXiv:hep-th/9401139].
31. M. Graña, J. Louis, D. Waldram, *$SU(3) \times SU(3)$ compactification and mirror duals of magnetic fluxes*, J. High Energy Phys. **04** (2007) 101 [arXiv:hep-th/0612237].
32. M. Graña, R. Minasian, M. Petrini, A. Tomasiello, *Supersymmetric backgrounds from generalized Calabi-Yau manifolds*, J. High Energy Phys. **08** (2004) 046 [arXiv:hep-th/0406137].
33. M. Graña, R. Minasian, M. Petrini, D. Waldram, *T-duality, generalized geometry and non-geometric backgrounds*, J. High Energy Phys. **04** (2009) 075 [arXiv:0807.4527 [hep-th]].
34. M. Gualtieri, *Generalized Complex Geometry*, PhD thesis, University of Oxford, 2003 [arXiv:math.DG/0401221].
35. M. Gualtieri, Generalized Complex Geometry, [arXiv:math.DG/0703298].
36. R. Harvey, B. Lawson, *From sparks to grundles – differential characters*, Comm. Anal. Geom. **14** (2006) 25–28 [arXiv:math.DG/0306193].
37. R. Harvey, B. Lawson, J. Zweck, *The de Rham-Federer theory of differential characters and character duality*, Amer. J. Math. **125** (2003) 791–847 [arXiv:math.DG/0512251].
38. N. Hitchin, *Generalized Calabi-Yau manifolds*, Quart. J. Math. Oxford Ser. **5**4 (2003) 281–308 [arXiv:math.DG/0209099].
39. M.J. Hopkins, I.M. Singer, *Quadratic functions in geometry, topology, and M-theory*, J. Diff. Geom. **70** (2005) 329–452 [arXiv:math.AT/0211216].
40. K. Hori, *D-branes, T-duality, and index theory*, Adv. Theor. Math. Phys. **3** (1999) 281–342 [arXiv:hep-th/9902102].
41. S. Hu, B. Uribe, *Extended manifolds and extended equivariant cohomology*, J. Geom. Phys. **59** (2009) 104–131 [arXiv:math.DG/0608319].
42. P. Kao, *T-duality and Poisson-Lie T-duality in generalized geometry*, PhD thesis, Australian National University, June 2008.
43. Y. Kosmann-Schwarzbach, *Derived brackets*, Lett. Math. Phys. **69** (2001) 61–87 [arXiv:math.DG/0312524].
44. Y. Kosmann-Schwarzbach, "*Quasi, twisted, and all that ... in Poisson geometry and Lie algebroid theory*", in *The Breadth of Symplectic and Poisson Geometry*, Progr. Math. **232** (2005) 363–389, Birkhäuser Boston, Boston (2005) [arXiv:math.SG/0310359].
45. Z.-J. Liu, A. Weinstein, P. Xu, *Manin triples for Lie bialgebroids*, J. Diff. Geom. **45** (1997) 547–574 [arXiv:dg-ga/9508013].
46. W.S. Massey, "Singular homology theory", *Graduate Texts in Mathematics* **70**, Springer-Verlag, New York (1980).
47. V. Mathai, J. Rosenberg, *T-duality for torus bundles via noncommutative topology*, Comm. Math. Phys. **253** (2005) 705–721 [arXiv:hep-th/0401168].
48. V. Mathai, J. Rosenberg *T-duality for torus bundles with H-fluxes via noncommutative topology, II: the high-dimensional case and the T-duality group*, Adv. Theor. Math. Phys. **10** (2006) 123–158 [arXiv:hep-th/0508084].
49. K.C. Mackenzie, "General theory of Lie groupoids and Lie algebroids", *London Mathematical Society Lecture Note Series* vol. 123, Cambridge University Press, Cambridge (2005).
50. R. Minasian, G. Moore, *K-theory and Ramond-Ramond charge*, J. High Energy Phys. **11** (1997) 002 [arXiv:hep-th/9710230].
51. G. Moore, E. Witten, *Self duality, Ramond-Ramond fields, and K-theory*, J. High Energy Phys. **05** (2000) 032 [arXiv:hep-th/9912279].
52. M. Murray, *An introduction to bundle gerbes*, to appear in the proceedings of the conference in honor of Nigel Hitchin's 60th Birthday [arXiv:0712.1651 [math.DG]].

53. M. Murray, D. Stevenson, *Higgs fields, bundle gerbes and string structures*, Comm. Math. Phys. **243** (2003) 541–555 [`arXiv:math/0106179`].
54. I. Raeburn, J. Rosenberg, *Crossed products of continuous-trace C^*-algebras by smooth actions*, Trans. Amer. Math. Soc. **305** (1988) 1–45.
55. D. Roytenberg, *Courant Algebroids, Derived Brackets and Even Symplectic Supermanifolds*, PhD thesis, University of California Berkeley, 1999 [`arXiv:math.DG/9910078`].
56. D. Roytenberg, *Quasi-Lie bialgebroids and twisted Poisson manifolds*, Lett. Math. Phys. **61** (2002) 123–137 [`arXiv:math.QA/0112152`].
57. P. Ševera, Letters to A. Weinstein, available at `http://sophia.dtp.fmph.uniba.sk/severa/letters/`.
58. P. Ševera, A.Weinstein, *Poisson geometry with a 3-form background*, Prog. Theor. Phys. Suppl. **144** (2001) 145–154 [`arXiv:math.SG/0107133`].
59. K. Uchino, *Remarks on the definition of a Courant algebroid*, Lett. Math. Phys. **60** (2002) 171–175 [`arXiv:math.DG/0204010`].
60. E. Witten, *D-Branes and K-Theory*, J. High Energy Phys. **12** (1998) 019 [`arXiv:hep-th/9810188`].

Chapter 6
Stochastic Geometry and Quantum Gravity: Some Rigorous Results

H. Zessin

In Erinnerung an Herbert Ziezold

Abstract The aim of these lectures is a short introduction into some recent developments in stochastic geometry which have one of its origins in simplicial gravity theory (see Regge Nuovo Cimento **19**: 558–571, 1961). The aim is to define and construct rigorously point processes on spaces of Euclidean simplices in such a way that the configurations of these simplices are simplicial complexes. The main interest then is concentrated on their curvature properties. We illustrate certain basic ideas from a mathematical point of view. An excellent representation of this area can be found in Schneider and Weil (*Stochastic and Integral Geometry*, Springer, Berlin, 2008. German edition: Stochastische Geometrie, Teubner, 2000). In Ambjørn et al. (*Quantum Geometry* Cambridge University Press, Cambridge, 1997) you find a beautiful account from the physical point of view. More recent developments in this direction can be found in Ambjørn et al. ("Quantum gravity as sum over spacetimes", Lect. Notes Phys. **807**. Springer, Heidelberg, 2010). After an informal axiomatic introduction into the conceptual foundations of Regge's approach the *first lecture* recalls the concepts and notations used. It presents the fundamental zero-infinity law of stochastic geometry and the construction of cluster processes based on it. The *second lecture* presents the main mathematical object, i.e. Poisson–Delaunay surfaces possessing an intrinsic random metric structure. The *third and fourth lectures* discuss their ergodic behaviour and present the two-dimensional Regge model of pure simplicial quantum gravity. We terminate with the formulation of basic open problems. Proofs are given in detail only in a few cases. In general the main ideas are developed. Sufficiently complete references are given.

H. Zessin (✉)
Fakultät für Mathematik, Universität Bielefeld, Universitätsstraße 25, D-33615, Germany
`zessin@math.uni-bielefeld.de`

Based on lectures held at the Instituto de cibernetica, matematica y fisica (ICIMAF), La Habana (Cuba), in March 2006.

Zessin, H.: *Stochastic Geometry and Quantum Gravity: Some Rigorous Results*. Lect. Notes Phys. **807**, 313–335 (2010)
DOI 10.1007/978-3-642-11897-5_6

6.1 An Axiomatic Introduction into Regge's Model

For a better understanding of the following mathematical reasoning I first present an *informal axiomatic* approach to Regge's [13] model of pure quantum gravity. (We shall use already some notions which will be made precise later.)

Consider a space X of geometric objects (cells, simplices) x in $E = \mathbb{R}^d,\ d = 4$, where x denotes a finite, affinely independent subset of E. On the next level we consider *configurations* μ of such cells in E, which are *locally finite* in the sense that

$$\zeta_B(\mu) = cd\{x \in \mu | b(x) \in B\} < +\infty,\ B \in \mathcal{B}_0(E). \tag{1}$$

This means that for any bounded Borel subset B of E, the number of cells in the configuration μ with barycentre hitting B is finite. Note that ζ_B depends only on the configuration of barycentres of μ.

We now restrict the class of cell configurations to *spacetimes*: One is only interested in those μ which are simplicial complexes; we then write $\mu \in \Gamma$. Thus with a cell x in the configuration μ, also all subsets $y \subseteq x$ are elements of μ; moreover, if the convex hulls of two cells meet then they meet in a common face.

The *first three axioms* can now be formulated as follows:

A$_1$ The spacetime (or cosmos) μ is a four-dimensional simplicial complex. This means that any maximal simplex $x \in \mu$ has cardinality 5.

A$_2$ The spacetime μ is a *random simplicial complex in E*, i.e. μ is realized in the collection Γ according to some law (probability) P on Γ, its σ-field $\mathcal{B}_\Gamma$ being generated by the random variables ζ_B, where B runs through all bounded Borel sets of E. Furthermore *P is of first order*, i.e. each counting variable ζ_B is integrable with respect to P. Moreover, *P is not the empty spacetime*, i.e. the event $\{\zeta_E > 0\}$ is certain, i.e. has probability 1.

A$_3$ The random spacetime is *stationary*, i.e. the underlying law P is invariant with respect to all translations, induced by the translations of the Euclidean space E.

To summarize: A model for a stationary random simplicial complex in E is thus given by a probability space $(\Gamma, \mathcal{B}_\Gamma, P)$, on which the mean number of cells hitting any bounded region with their barycentres is finite and the vacuum is not realized. Intuitively, P represents a random mechanism, which realizes four-dimensional simplicial complexes of cells.

In the sequel we consider only the two-dimensional example of a *Poisson–Delaunay surface in E*. This random surface is obtained by first realizing a stationary Poisson point process in $\mathbb{R}^2$, each point being marked independently by a (random) strictly positive number. We interpret its realizations as a curved surface consisting of triangles as follows: In a first step a Poissonian realization of positions in $\mathbb{R}^2$ gives rise to a *Delaunay tesselation* of the plane into triangles; and in a second step the marks of their vertices are used to redefine the lengths of the triangle sides.

(The Euclidian distances are discarded!) In this way two-dimensional curved spacetimes appear. Finally, all triangles are augmented by their sides and vertices, each of them appearing only once. In this way one obtains two-dimensional stationary random simplicial complexes. The assumption of stationarity for P allows the introduction of the notion of a *Palm measure* P^0 *of* P *with respect to the barycentres*. This is a finite measure on Γ, because P is of first order. Its normalization P_0 is the conditional law given that 0 is the barycentre of the realized simplicial complex. This will be a fundamental notion in the sequel.

Another fundamental concept is *curvature*. This is needed to formulate the fourth axiom. Regge proposed in his article the following discrete version of the *Einstein–Hilbert action*. In the present two-dimensional situation it is concentrated in the vertices a of μ and 0 otherwise. Thus if a is a vertex on μ for some $\mu \in \Gamma$ the *Regge action* is defined by

$$g_{\mathcal{R}}(a, \mu) = -(g_c - g_v)(a, \mu), \tag{2}$$

where $g_{\mathrm{v}}(a, \mu)$ is the sum of areas of all triangles t in μ having a as vertex; and $g_{\mathrm{c}}(a, \mu)$ denotes the *curvature of* μ *in its vertex* a, given by the deficit angle of μ in a.

It is important here to note that the notions of area or angle of a triangle t depend on the (random) metric chosen on the sides of t. Recall from above that this metric is induced by the marks in the vertices.

We are now in the position to state the last axiom:

A₄ P is curved with respect to the Regge action, i.e. $g_{\mathcal{R}}$ is integrable with respect to the Palm measure P^0 of P.

Observe here, since $g_{\mathcal{R}}$ is given in advance, the last axiom is an assumption on the random spacetime P (point of view of classical statistical mechanics).

To summarize: The Regge model of pure quantum gravity is given by a quadruple $(\Gamma, \mathcal{B}_\Gamma, P, g_{\mathcal{R}})$ which satisfies the axioms $\mathbf{A_1}, \ldots, \mathbf{A_4}$.

This model only makes assumptions on the *local behaviour* of the random spacetime. Therefore the global properties of this model are of interest. Apart from the construction of such random spacetimes *the aim of these lectures is the analysis of their global properties*.

The stationarity assumption $\mathbf{A_3}$ is the starting point for this. It implies by means of *the zero-infinity law* that a typical realization μ of the spacetime consists of an infinity of simplices. Since spacetimes are locally finite, they thus are infinitely extended. Moreover the question of the ergodic behaviour of the Regge action in a bounded region Λ of E is meaningful. To be more precise, we consider *the total normalized Regge action in* Λ

$$\frac{1}{|\Lambda|} \cdot \sum_{a \in b\mu \cap \Lambda} g_{\mathcal{R}}(a, \mu)$$

and ask for its asymptotic behaviour if $\Lambda \nearrow +\infty$ ($b\mu$ denotes the set of barycentres of the simplices of μ).

One of the main results will be that the limit $\mathfrak{r}$, the so-called *specific Regge action*, exists in a precise sense, is translation invariant and satisfies $P(\mathfrak{r}) = -P_0(g_{\mathcal{R}})$. This last equation says that the expected specific Regge action is, up to the factor -1, the expectation of the Regge action with respect to the Palm distribution of P. This last quantity can be evaluated in principle.

6.2 The Zero-Infinity Law of Stochastic Geometry and the Cluster Process

6.2.1 Basic Concepts

We present as a general frame the scheme of Ripley [15], who proposed a purely measure theoretical approach to point process theory on countably separated, σ-bounded spaces in which random simplicial complexes are constructed.

The starting point is a measurable space $(X, \mathcal{B})$ in which configurations of points will be realized, which are locally finite in the sense that only a finite number of points hit each member of a class $\mathcal{B}_0$ of 'bounded' sets. Following Ripley we assume that $\mathcal{B}$ contains all singleton subsets and that $\mathcal{B}_0$ is a non-empty subset of $\mathcal{B}$ which is hereditary, i.e.

$$(B \in \mathcal{B}_0, C \in B \cap \mathcal{B} \Rightarrow C \in \mathcal{B}_0),$$

closed under finite unions and σ-bounded. The latter means that there exists an increasing sequence $X_1, X_2, \ldots$ in $\mathcal{B}_0$ covering X. Such spaces will be called here as bounded, measurable spaces.

We assume also that $(X, \mathcal{B}, \mathcal{B}_0)$ is countably separated, i.e. there exists a countable π-system $\tilde{\mathcal{B}}_0$ in $\mathcal{B}_0$ separating the points of X; given any finite subset of X, there are disjoint members of $\tilde{\mathcal{B}}_0$ such that each member of the finite set is contained in precisely one of these disjoint sets.

As Ripley has shown one can develop a general theory of point processes in such spaces $(X, \mathcal{B}, \mathcal{B}_0)$, which we call *phase spaces* for short. Examples of phase spaces are all separable metric spaces (X, d) where $\mathcal{B}$ is the Borel σ-field and $\mathcal{B}_0$ the collection of metrically bounded Borel sets. For $\tilde{\mathcal{B}}_0$ take the collection of all balls with centres from a dense countable subset and rational radii. In particular, all polish spaces and all locally compact, second countable Hausdorff topological spaces are phase spaces.

Given a phase space $(X, \mathcal{B}, \mathcal{B}_0)$, we then consider random locally finite measures μ on X, which are defined as follows: $\mathcal{M} = \mathcal{M}(X)$ denotes the set of all measures μ on $(X, \mathcal{B})$ which are locally finite, i.e. μ is finite on $\mathcal{B}_0$. Note that such measures are σ-finite because X is σ-bounded.

$\mathcal{M}$ is endowed with the σ-field $\mathcal{F}$ generated by all variables $\zeta_B : \mu \mapsto \mu(B)$ $(B \in \mathcal{B}_0)$. Important measurable subsets are

$$\begin{aligned}
\mathcal{M}^{\cdot\cdot} &= \{\mu \in \mathcal{M} \mid \mu(B) \in \mathbb{N}_0 \quad \forall\, B \in \mathcal{B}_0\} \text{ (point measures)},\\
\mathcal{M}^{\cdot} &= \{\mu \in \mathcal{M}^{\cdot\cdot} \mid \mu\{x\} \leq 1 \quad \forall\, x \in X\} \text{ (simple point measures)},\\
\mathcal{M}^{\cdot}_{k+1} &= \mathcal{M}^{\cdot} \cap \{\zeta_X = k+1\}, k \geq 0.
\end{aligned}$$

Here $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. The traces of $\mathcal{F}$ in these spaces are denoted by $\mathcal{F}^{\cdot\cdot}$, $\mathcal{F}^{\cdot}$, $\mathcal{F}^{\cdot}_{k+1}$. Elements $\mu \in \mathcal{M}^{\cdot}$ are often identified with locally finite subsets of X. The elements of $\mathcal{M}^{\cdot}_{k+1}$ thus are the subsets of X of cardinality $k+1$.

A *random measure* in X is a probability measure P on $(\mathcal{M}, \mathcal{F})$; we then write $P \in \mathcal{PM}$ for short. Thus P is a law which realizes a locally finite measure μ in X. Probabilities P on $(\mathcal{M}^{\cdot\cdot}, \mathcal{F}^{\cdot\cdot})$ resp. $(\mathcal{M}^{\cdot}, \mathcal{F}^{\cdot})$ are called *point processes* resp. *simple point processes* in X, and write $P \in \mathcal{PM}^{\cdot\cdot}$ resp. $P \in \mathcal{PM}^{\cdot}$, etc.

Then ζ_B, $B \in \mathcal{B}$, are random variables on $(\mathcal{M}, \mathcal{F}, P)$ representing the random measure of B. The first moment measure of P is defined by $\nu_P^1(B) = P(\zeta_B)$, $B \in \mathcal{B}$, where $P(\zeta_B)$ denotes the expectation of ζ_B with respect to P. ν_P^1 is also called the *intensity* of P. Intuitively, $\nu_P^1(B)$ is the *expected random measure of* B.

If in X there is defined a group of measurable transformations

$$T_x : X \to X, y \mapsto y - x,$$

written symbolically as translations, they induce a group of measurable transformations

$$T_x : \mathcal{M} \to \mathcal{M}, \mu \mapsto \mu - x := \text{image of } \mu \text{ under } T_x.$$

These in turn induce a group of translations $(T_x)_x$ on $\mathcal{PM} : T_x : P \mapsto T_x P :=$ image of P under T_x.

A random measure P on X then is called *stationary* if $P = T_x P$ for all $x \in X$; we then write $P \in \mathcal{P}_0\mathcal{M}$. (All this remains valid if $\mathcal{M}$ is replaced by $\mathcal{M}^{\cdot\cdot}$ resp. $\mathcal{M}^{\cdot}$, etc.)

6.2.2 Cluster Properties and the Zero-Infinity Law of Stochastic Geometry

We now take $\mathbb{R}^d$ $(d \geq 1)$, the d-dimensional Euclidean space, as the basic phase space together with its group of translations $(T_a)_{a \in \mathbb{R}^d}$. Given $k \geq 0$ a *k-cluster property* in $\mathcal{M}^{\cdot}$ is any measurable subset $D \subseteq \mathcal{M}^{\cdot}_{k+1} \times \mathcal{M}^{\cdot}$. We shall use the following terminology: If $(x, \eta) \in D$ then x is called a *cluster for* η of type D; if in addition $x \subseteq \eta$ then x is a cluster of type D *in* η. A cluster property D is called *stationary* if

$$((x, \eta) \in D, a \in \mathbb{R}^d \Rightarrow (x - a, \eta - a) \in D).$$

Example 1 Given some $r > 0$ we define a 0-cluster property D_r by means of

$$((a, \eta) \in D_r \text{ iff } \eta(\dot{B}_r(a)) = 0).$$

Here $B_r(a)$ is the open d-ball centred in a with radius r and $\dot{B}_r(a) = B_r(a) \setminus \{a\}$ its perforation. It is easy to show that this is a stationary cluster property in $\mathcal{M}^{\cdot}$.

Thus a point $a \in \mathbb{R}^d$ is a cluster for the configuration $\eta \in \mathcal{M}^{\cdot}$ iff η has no points in $\dot{B}_r(a)$, the perforated r-ball centred in a.

We associate to a given stationary cluster property D in $\mathcal{M}^{\cdot}$ the variable

$$cd_D : \mathcal{M}^{\cdot} \to \mathbb{N}_0 \cup \{+\infty\}, \eta \mapsto \sum_{x \subseteq \eta} 1_D(x, \eta).$$

cd_D counts the clusters of type D in η. It is measurable and invariant under translations T_a, $a \in \mathbb{R}^d$.

The following fundamental result, the so-called zero-infinity law, is the basis of all later constructions. The proof given here is due to Krickeberg [6].

Theorem 1 *(Zero-infinity law of stochastic geometry)*
If D is a stationary k-cluster property in $\mathcal{M}^{\cdot}$ and $P \in \mathcal{P}_0\mathcal{M}^{\cdot}$, then

$$P\{0 < cd_D < +\infty\} = 0.$$

Proof Suppose $P\{0 < cd_D < +\infty\} > 0$. Then $\tilde{P} := P(\cdot \mid 0 < cd_D < +\infty)$ is a well-defined simple, stationary point process in $\mathbb{R}^d$, i.e. $\tilde{P} \in \mathcal{P}_0\mathcal{M}^{\cdot}$. Now transform $\tilde{P}$ into a translation-invariant probability on $\mathbb{R}^d$ by means of

$$\chi : \mathcal{M}^{\cdot} \to \mathbb{R}^d, \eta \mapsto \frac{1}{cd_D(\eta)} \cdot \sum_{x \subseteq \eta} 1_D(x, \eta) \cdot b(x).$$

Here $b(x)$ is the barycentre of x. Note that χ is $\tilde{P}$-a.e. well defined and measurable. Thus we obtain a contradiction. Translation-invariant probabilities do not exist on $\mathbb{R}^d$. **qed**

This theorem immediately implies the

Corollary 1 *If D is a stationary k-cluster property in $\mathcal{M}^{\cdot}$ and $P \in \mathcal{P}_0\mathcal{M}^{\cdot}$ such that*

$$P\{cd_D \geq 1\} > 0, \tag{3}$$

then the law $P_D := P(\cdot \mid cd_D \geq 1)$ is concentrated on $\mathcal{M}^{\cdot}_D = \{cd_D = +\infty\}$.

As a consequence $P_D = \frac{1}{Z_D} \cdot 1_{\mathcal{M}^{\cdot}_D} \cdot P$, where $Z_D := P(\mathcal{M}^{\cdot}_D) > 0$. P_D is a simple stationary point process in $\mathbb{R}^d$ realizing configurations of particles possessing an infinity of D-clusters.

6.2.3 Cluster Properties and the Zero-Infinity Law for Marked Configurations

The phase space now will be $\mathbb{R}^d \times I$, where $I \subseteq]0, +\infty[$ is an interval. Its elements $z = (a, r)$ represent balls centred in a with radius r. $q : \mathbb{R}^d \times I \to \mathbb{R}^d$, $z \mapsto a$, denotes the projection onto the centre and $r : z \to r$ to its radius.

We consider the following collection of configurations of such balls in $\mathbb{R}^d$:

$$\mathcal{M}^{\cdot}_I = \left\{ \nu = \sum_{a\in\eta} \delta_{(a,r_a)} \mid \eta \in \mathcal{M}^{\cdot}, r_a \in I \right\}$$

and are interested in cluster properties in this set. It is obvious that a k-cluster property D in $\mathcal{M}^{\cdot} = \mathcal{M}^{\cdot}(\mathbb{R}^d)$ induces a k-cluster property in $\mathcal{M}^{\cdot}_I$ by means of the measurable projection

$$\Pi : \mathcal{M}^{\cdot}_{I,k+1} \times \mathcal{M}^{\cdot}_I \to \mathcal{M}^{\cdot}_{k+1} \times \mathcal{M}^{\cdot}, (z, \nu) \mapsto (qz, q\nu)$$

as follows: $\mathcal{D} := \Pi^{-1}(D)$.

The natural group $(T_a)_{a\in\mathbb{R}^d}$ of translations in $\mathbb{R}^d \times I$ is defined by $T_a : (b, r) \mapsto (b - a, r)$. They induce translations T_a in $\mathcal{M}^{\cdot}_I$ by

$$T_a : \nu \to \nu - a := \sum_{b\in\eta} \delta_{(b-a,r_b)}.$$

Let $\mathcal{P}_0\mathcal{M}^{\cdot}_I$ denote the stationary laws on $\mathcal{M}^{\cdot}_I$.

If D is a stationary k-cluster property in $\mathcal{M}^{\cdot}(\mathbb{R}^d)$, then $\mathcal{D}$ is a stationary k-cluster property in $\mathcal{M}^{\cdot}_I$. Now it is straightforward that the following $0-\infty$-law and its corollary hold true.

Theorem 2 *(Zero-infinity law (for marked configurations))*
If D is a stationary k-cluster property in $\mathcal{M}^{\cdot}(\mathbb{R}^d)$ and $P \in \mathcal{P}_0\mathcal{M}^{\cdot}_I$, then

$$P\{0 < cd_{\mathcal{D}} < +\infty\} = 0.$$

Corollary 2 *If in the situation of the 0∞-law we know that the image $P^* = qP$ of P under $q : \nu \to q\nu$ satisfies*

$$P^*\{cd_D \geq 1\} > 0, \tag{4}$$

then $P\{cd_{\mathcal{D}} \geq 1\} > 0$; and $P_{\mathcal{D}} := P(\cdot \mid cd_{\mathcal{D}} \geq 1) \in \mathcal{P}_0\mathcal{M}^{\cdot}_I$ with

$$P_{\mathcal{D}}\{cd_{\mathcal{D}} = +\infty\} = 1.$$

Notation: $\mathcal{M}^{\cdot}_{I,\mathcal{D}} = \{cd_{\mathcal{D}} = +\infty\}$.

6.2.4 The Cluster Process

We shall now construct basic objects of these lectures, namely point processes in the phase space of clusters. In order to do this we have to impose another condition on a cluster property. A k-cluster property D in $\mathcal{M}^{\cdot} = \mathcal{M}^{\cdot}(\mathbb{R}^d)$ is called locally finite iff for any $\eta \in \mathcal{M}^{\cdot}_D$ the point measure

$$\mu := \sum_{x \subseteq \eta} 1_D(x, \eta) \cdot \delta_x$$

is locally finite on the space $X := \mathcal{M}^{\cdot}_{k+1}$ in the following sense:

$$\mu(\mathcal{G}_\Lambda) < +\infty \text{ for any } \Lambda \in \mathcal{B}_0(\mathbb{R}^d), \tag{5}$$

where

$$\mathcal{G}_\Lambda = \{x \in X \mid x \cap \Lambda \neq \emptyset\}.$$

We remark that X is a nice phase space if endowed with the σ-fields $\mathcal{B}(X)$ generated by all $\mathcal{G}_\Lambda$, $\Lambda \in \mathcal{B}_0(\mathbb{R}^d)$, and with the collection of bounded sets

$$\mathcal{B}_0(X) = \{B \in \mathcal{B}(X) \mid B \subseteq \mathcal{G}_\Lambda \text{ for some } \Lambda \in \mathcal{B}_0(\mathbb{R}^d)\}.$$

Instead of X we are mainly interested in the space $X_I := \mathcal{M}^{\cdot}_{I,k+1}$ of marked clusters. If D is locally finite in $\mathcal{M}^{\cdot}(\mathbb{R}^d)$ then so is $\mathcal{D}$ in the following obvious sense. For any $\nu \in \mathcal{M}^{\cdot}_{I,\mathcal{D}}$ the point measure

$$\kappa := \sum_{z \subseteq \nu} 1_{\mathcal{D}}(z, \nu) \cdot \delta_z$$

is locally finite on the space X_I, in the sense that

$$\kappa(\mathcal{G}_{I,\Delta}) < +\infty \text{ for any } \Delta \in \mathcal{B}_0(\mathbb{R}^d \times I), \tag{6}$$

where

$$\mathcal{B}_0(\mathbb{R}^d \times I) = \{\Delta \in \mathcal{B}(\mathbb{R}^d \times I) \mid \Delta \subseteq \Lambda \times I \text{ for some } \Lambda \in \mathcal{B}_0(\mathbb{R}^d)\}$$

and

$$\mathcal{G}_{I,\Delta} = \{z \in X_I \mid z \cap \Delta \neq \emptyset\}.$$

Again X_I is a nice phase space if $\mathcal{B}(X_I)$ is generated by the $\mathcal{G}_{I,\Delta}$ and if

$$\mathcal{B}_0(X_I) = \{B \in \mathcal{B}(X_I) \mid B \subseteq \mathcal{G}_{I,\Delta} \text{ for some } \Delta \in \mathcal{B}_0(\mathbb{R}^d \times I)\}.$$

Now we are able to study point processes on X_I: Given a locally finite, stationary k-cluster property D in $\mathcal{M}^{\cdot}(\mathbb{R}^d)$, define the measurable transformation

$$\gamma_D : \mathcal{M}^{\cdot}_{I,\mathcal{D}} \to \mathcal{M}^{\cdot}(X_I), \nu \mapsto \kappa = \sum_{z \subseteq \nu} 1_{\mathcal{D}}(z, \nu) \cdot \delta_z.$$

If $P \in \mathcal{P}_0\mathcal{M}^{\cdot}_I$ is a given stationary (marked) point process in $\mathbb{R}^d \times I$ concentrated on configurations from $\mathcal{M}^{\cdot}_I$ and satisfying condition (4), then $P_{\mathcal{D}} \in \mathcal{P}_0\mathcal{M}^{\cdot}_I$ and thus the following process in X_I is well defined

$$Q_D := \gamma_D P_{\mathcal{D}}.$$

Here the right-hand side denotes again the image of $P_{\mathcal{D}}$ under γ_D. We call Q_D the *cluster process* of P of type D. More explicitly one has

$$Q_D(\varphi) = \frac{1}{P(\mathcal{M}^{\cdot}_{I,\mathcal{D}})} \cdot \int_{\mathcal{M}^{\cdot}_{I,\mathcal{D}}} \varphi \circ \gamma_D \, dP, \varphi \geq 0 \text{ measurable.} \tag{7}$$

It is obvious that Q_D is stationary with respect to the group of translations induced by $(T_a)_{a \in \mathbb{R}^d}$ in X_I and denoted in the same way.

To summarize, we have the following.

Theorem 3 *Let D be a locally finite, stationary k-cluster property in $\mathcal{M}^{\cdot}(\mathbb{R}^d)$ and $P \in \mathcal{P}_0\mathcal{M}^{\cdot}_I$ such that*

$$qP\{cd_D \geq 1\} > 0. \tag{8}$$

Then $Q_D \in \mathcal{P}_0\mathcal{M}^{\cdot}(X_I)$, and Q_D is concentrated on configurations consisting only of an infinity of $\mathcal{D}$-clusters.

For later use we modify the cluster process slightly. We transform it by means of the following augmentation transformation:

$$\alpha : \mathcal{M}^{\cdot}(X_I) \to \mathcal{M}^{\cdot}(\hat{X}_I), \kappa \mapsto \operatorname{supp} \sum_{z \in \kappa} \sum_{0 \neq y \subseteq z} \delta_y =: \mu.$$

Here $\hat{X}_I = \bigcup_{l=1}^{k+1} \mathcal{M}^{\cdot}_{I,l}$, and supp denotes the support of a point measure. Again as above $\hat{X}_I$ can be made into a nice phase space, carrying the group $(T_a)_a$ of translations induced by $\mathbb{R}^d$. $\alpha(\kappa) = \mu$ adds to κ, which consists of marked simplices, all their faces of lower dimension, each of them being added only once.

Denote by $\hat{Q}_D$ the image of Q_D under α, i.e. $\hat{Q}_D = \alpha Q_D$. This is a stationary point process in $\hat{X}_I$ whose configurations μ are built on simplices with all their faces and have the characteristic property of a simplicial complex, i.e.

$$(x \in \mu, y \subseteq x \Rightarrow y \in \mu).$$

Denote this subset of $\mathcal{M}^{\cdot}(X_I)$ of simplicial complexes by Γ and by $\mathcal{F}_{\Gamma}^{\cdot}$ its canonical σ-field. $\hat{Q}_D$ is called the *random simplicial complex* of P of type D.

Until now we saw already one example of a cluster property, which is stationary and obviously locally finite. But we have no example of a point process P satisfying the assumptions of the theorem. Such an example is given by

6.2.5 The Poisson Process P_ρ

Let $(X, \mathcal{B}, \mathcal{B}_0)$ be a phase space and $\rho \in \mathcal{M}(X)$.

We first assume that $0 < \rho(X) < +\infty$. The *Poisson process* P_ρ is constructed as follows: Select first an integer n at random according to the Poisson law with parameter $\rho(X)$ and then distribute n independent random points in X, each one according to the law $\rho/\rho(X)$. The corresponding point process P_ρ is given by

$$P_\rho(\varphi) = \exp(-\rho(X)) \cdot \sum_{n=0}^{\infty} \frac{1}{n!} \cdot \int_{X^n} \varphi(\delta_{x_1} + \cdots + \delta_{x_n})\rho(dx_1) \ldots \rho(dx_n)$$

($\varphi \geq 0$ and measurable). It is easy to see that the *reduced Campbell measure* of P_ρ, defined by

$$C^!_{P_\rho}(h) := \int_{\mathcal{M}^{\cdot\cdot}} \int_X h(x, \mu - \delta_x)\mu(dx) P_\rho(d\mu), \;\; h \geq 0 \text{ measurable,}$$

is the product measure $\rho \otimes P_\rho$ on $X \times \mathcal{M}^{\cdot\cdot}$; and this property characterizes P_ρ. To say this in another way: Given ρ as above, a point process $P \in \mathcal{PM}^{\cdot\cdot}(X)$ is equal to P_ρ iff

$$C^!_P = \rho \otimes P. \tag{9}$$

This basic equation obviously makes sense for any $P \in \mathcal{PM}^{\cdot\cdot}$ if ρ is infinite; and in fact a famous theorem of Mecke [10] states that given any $\rho \in \mathcal{M}(X)$, $\rho \neq 0$, then (9) has a unique solution P_ρ in $\mathcal{PM}^{\cdot\cdot}$ called the *Poisson process in X with intensity* ρ. This name is well adapted to its properties: For any $B \in \mathcal{B}_0(X)$ ζ_B has a Poisson distribution with parameter $\rho(B)$, and $\zeta_{B_1}, \ldots, \zeta_{B_k}$ are independent whenever $B_1, \ldots, B_k \in \mathcal{B}_0(X)$ are disjoint.

One can show that $P_\rho \in \mathcal{PM}^{\cdot}$ iff ρ is *diffuse*, i.e. $\rho\{x\} = 0$ for any $x \in X$. It follows directly from (9) that $\nu^1_{P_\rho} = \rho$, i.e. the expected random point configuration of P_ρ is given by ρ.

We now consider the phase space $E = \mathbb{R}^d \times I$ endowed with the diffuse measure $\rho = z \cdot \lambda \otimes \tau$, where $z > 0$, λ is a Lebesgue measure on $\mathbb{R}^d$ such that the unit cube has measure 1, and τ is a *finite* measure on I, $\neq 0$.

It is easy to see that P_ρ is a stationary point process in E, i.e. $P_\rho \in \mathcal{P}_0\mathcal{M}^{\cdot\cdot}(E)$, using Mecke's characterization of a Poisson process. In the same way one obtains

$$qP_\rho = P_{q\rho} \text{ with } q\rho = z \cdot \tau(I) \cdot \lambda.$$

This implies the

Lemma 1 $P_\rho \in \mathcal{P}_0\mathcal{M}^{\cdot}_I$.

Proof Since $q\rho$ is diffuse, one knows that qP_ρ is concentrated on $\mathcal{M}^{\cdot} = \mathcal{M}^{\cdot}(\mathbb{R}^d)$, i.e. $P_\rho\{q \in \mathcal{M}^{\cdot}\} = 1$. On the other hand $\{q \in \mathcal{M}^{\cdot}\} = \mathcal{M}^{\cdot}_I$. **qed**

To summarize: If we want to use theorem 3 for some given cluster property D and the Poisson process P_ρ, we have to verify only the condition

$$P_{q\rho}\{cd_D \geq 1\} > 0. \tag{10}$$

Example 2 For the cluster property D_r condition (10) is true. For a proof let $b \in \mathbb{R}^d$, $\varepsilon, r > 0$, and consider the event $A_b = \{\zeta_{B_\varepsilon(b)} = 1, \zeta_{B_{\varepsilon+r}(b)\setminus B_\varepsilon(b)} = 0\}$ that the ball $B_\epsilon(b)$ contains exactly one point of the configuration and $B_{\varepsilon+r}(b) \setminus B_\varepsilon(b)$ is void.

It is evident that $A_b \subseteq \{cd_{D_r} \geq 1\}$. On the other hand, $P_{q\rho}(A_b) = z \cdot \tau(I) \cdot \lambda(B_\varepsilon(A)) \cdot \exp(-z \cdot \tau(I) \cdot \lambda(B_{\varepsilon+r}(b))) > 0$. Thus the pair (D_r, P_ρ) satisfies the conditions of the theorem.

As a consequence the associated D_r-cluster process, denoted by Q_{ρ,D_r}, is well defined. It is called *Poisson exclusion process* by Mürmann [11] and realizes an infinity of hard $\frac{r}{2}$-balls in $\mathbb{R}^d$, with centres marked by positive numbers.

6.3 Poisson–Delaunay Surfaces with Intrinsic Random Metric

The above example D_r of a cluster property already yields an interesting cluster process which realizes geometric objects, namely hard balls. We now present our main example.

6.3.1 The Delaunay Cluster Property

First some notations: We consider the set $\mathcal{A}^{\cdot}_d$ of all $x \in \mathcal{M}^{\cdot}_{d+1}$ $(= \mathcal{M}^{\cdot}_{d+1}(\mathbb{R}^d))$ which are *affinely independent*. (This means that any subset of $k+1$ points, $1 < k \leq d+1$, is not contained in a linear subspace of dimension $< k$.) Elements of $\mathcal{A}^{\cdot}_d$ represent simplices in $\mathbb{R}^d$.

Given $x \in \mathcal{A}^{\cdot}_d$, any $(d-1)$-face y of x (i.e. $y \subseteq x$ such that $\dim y := cd\ y - 1 = d - 1$) generates a hyperplane H_y together with closed half-spaces $H^+_y(x)$ and $H^-_y(x)$, where $H^+_y(x)$ is the one containing x. Furthermore, $K(x)$ denotes the circumball of x, i.e. the smallest closed ball containing x. x is a subset of $S(x)$, the $(d-1)$-sphere of $K(x)$. Denote $\dot{K}(x) := K(x) \setminus x$.

We need also the following subset $\mathcal{C}^{\cdot}$ of configurations η spanning the whole space $\mathbb{R}^n$ in the following sense:

$$\eta \in \mathcal{C}^{\cdot} \text{iff} \, \langle \eta \rangle := \bigcup_{y \subseteq \eta \text{ finite}} \langle y \rangle = \mathbb{R}^n.$$

Here $\langle y \rangle$ denotes the convex hull of y. By a theorem of Carathéodory the condition $\eta \in \mathcal{C}^{\cdot}$ is equivalent to saying that not all points of η are contained in some half space. Then we define the following d-cluster property D_{el}:

$$(x, \eta) \in D_{\mathrm{el}} \text{ iff } x \in \mathcal{A}^{\cdot}_{d}, \eta \in \mathcal{C}^{\cdot}, \eta(\dot{K}(x)) = 0. \tag{11}$$

Thus we are interested in clusters x for configurations $\eta \in \mathcal{C}^{\cdot}$, which represent simplices and possess the *Delaunay property* $\eta(\dot{K}(x)) = 0$, which goes back to Delaunay [4]. This property means that the point configuration η has no points in $\dot{K}(x)$.

The stationarity of D_{el} is obvious, whereas its measurability is not. (For all measurability questions see [5].) As a consequence of the $0{-}\infty$-law, taking P_ρ, $\rho = z \cdot \lambda \otimes \tau$, as the underlying law, the counting variable $cd_{\mathcal{D}_{\mathrm{el}}}$ assumes P_ρ – almost surely only the values 0 and ∞.

Lemma 2 $P_{q\rho}\{cd_{D_{\mathrm{el}}} \geq 1\} > 0$.

Proof (Write D resp. $\mathcal{D}$ instead of D_{el} resp. $\mathcal{D}_{\mathrm{el}}$.) Recall that $q\rho = z \cdot \tau(I) \cdot \lambda$. *First* we show that $P_{q\rho}(\mathcal{C}^{\cdot}) = 1$. This follows from the Borel–Cantelli lemma, the argument being given for $d = 2$ for simplicity. Choose an equilateral triangle t in $\mathbb{R}^2$ with barycentre 0 and place small balls $B_\varepsilon(a)$, $\varepsilon > 0$, in each corner $a \in t$. Take $\varepsilon > 0$ so small that the whole collection

$$B_\varepsilon(n \cdot a), \; n \geq 1,$$

of balls is disjoint. Given n, consider the event 'each ball $B_\varepsilon(a)$ contains exactly one element of the configuration', i.e.

$$A_n(t) := \{\zeta_{B_\varepsilon(n \cdot a)} = 1 \text{ for any } a \in t\}.$$

By construction these events are independent with respect to $P_{q\rho}$. Moreover, by translation invariance of $q\rho$

$$P_{q\rho}(A_n(t)) = \prod_{a \in t} q\rho(B_\varepsilon(a)) \cdot \exp(-q\rho(B_\varepsilon(a))), n \geq 1;$$

these probabilities are strictly positive and do not depend on n. Consequently $\Sigma_{n \geq 1} P_{q\rho}(A_n(t)) = +\infty$. And by the Borel–Cantelli lemma it follows that

$$P_{q\rho}\left(\lim_{n\to\infty}\sup A_n(t)\right) = 1.$$

Since $\lim_{n\to\infty} \sup A_n(t) \subseteq \mathcal{C}^{\cdot}$, the assertion follows.

It remains to show that $P_{q\rho}\{cd_{D'} \geq 1\} > 0$, where D' is defined by

$$(x, \eta) \in D' \text{ iff } x \in \mathcal{A}^{\cdot}_d,\, \eta(\dot{K}(x)) = 0.$$

This follows by a similar argument: Choose t as above and $\varepsilon > 0$ small enough such that

$$E_n(t) = \{x \in \mathcal{M}^{\cdot}_3 \mid x(B_\varepsilon(a)) = 1 \text{ for any } a \in t\} \subseteq \mathcal{A}^{\cdot}_2.$$

Then choose a ball B containing all $B_\varepsilon(a)$, $a \in t$, and let $B' := B \setminus \cup_{a\in t} B_\varepsilon(a)$. Then the event $\{\zeta_{B'} = 0, \xi_{B_\varepsilon(a)} = 1 \text{ for any } a \in t\}$ is contained in $\{cd_{D'} \geq 1\}$ and has positive probability with respect to $P_{q\rho}$. **qed**

As a consequence of Theorem 3 the associated cluster process $Q_{\rho,D_{\mathrm{el}}}$ of P_ρ of type D_{el} is well defined. We call this process the *Poisson–Delaunay process*, $\mathcal{PD}$-process for short.

The following result is well known. A proof can be found in Chap. 6 of the German edition of Schneider/Weil [16].

Lemma 3 *For P_ρ – almost any $\nu \in \mathcal{M}^{\cdot}_{I,\mathcal{D}}$ the random cluster measure $\kappa = \gamma_D(\nu)$, having $Q_{\rho,\, D_{\mathrm{el}}}$ as its distribution, has the following support properties:*

(1) If $z, z' \in \kappa$ are distinct with $\langle x\rangle \cap \langle x'\rangle \neq \emptyset$ (where $x = qz$, $x' = qz'$), then $x \cap x' \neq \emptyset$, $x \setminus x'$ and $x' \setminus x$ are nonempty and situated on opposite sides of some hyperplane containing $x \cap x'$.
(2) Every $(d-1)$ face of an element $z \in \kappa$ is shared by another element $z' \in \kappa$.
(3) κ is a tesselation of $\mathbb{R}^d$; in particular one has

$$\bigcup_{x\in q\nu:(x,q\nu)\in D} \langle x\rangle = \mathbb{R}^d.$$

Thus a $\mathcal{PD}$-process typically realizes tesselations built on Delaunay triangles whose vertices are marked by strictly positive radii.

We remark as an aside: If we add more restrictions to the definition of D_{el}, then $\kappa = \gamma_{D_{\mathrm{el}}}(\nu)$ may show a qualitatively different geometrical and topological behaviour. To be more precise: Given two parameter $0 < r < R < +\infty$, consider the following cluster property in $\mathcal{M}^{\cdot}$:

$$(x, \eta) \in D_{\mathrm{el}}(r, R) \text{ iff } (x, \eta) \in D_{\mathrm{el}},\, \eta \in \mathcal{M}^{\cdot}_{(r)},\ \operatorname{diam} x < R.$$

Here $\eta \in \mathcal{M}^{\cdot}_{(r)}$ iff $\eta \in \mathcal{M}^{\cdot}$ s.th. $(a, b \in \eta, a \neq b \Rightarrow \|a - b\| \geq r)$ and diam x denotes the diameter of x being defined as the diameter of $K(x)$. Note that $r = 0$ and $R = +\infty$ corresponds to the former case.

Now (x, η) is an admissible pair iff $\eta \in \mathcal{C}^{\cdot}$ is built on hard $\frac{r}{2}$-balls and x is a Delaunay cluster for η of diameter $< R$. It is evident that $D_{\mathrm{el}}(r, R)$ is stationary and locally finite. As above one shows easily that condition (4) resp. (8) is satisfied.

It is intuitively clear that now the cluster measures κ, realized by $Q_{\rho, D_{\mathrm{el}}(r,R)}$, may have holes (and thus boundaries) if the parameters r, R are chosen in a right way. In this case and if $d = 3$ or $d = 4$ it would be interesting to consider the corresponding surface process. Here we shall not follow this line of thinking, which has been developed by Matzutt in [9].

6.3.2 Poisson–Delaunay Surfaces

We assume from now on $d = 2$ for simplicity. As above the underlying point process in $\mathbb{R}^2 \times I$ is P_ρ with $\rho = z \cdot \lambda \otimes \tau$. And the cluster property given by $D = D_{\mathrm{el}}$. P_ρ is concentrated on configurations $\nu \in \mathcal{M}_I^{\cdot}$ possessing an infinity of Delaunay triangles, here denoted by t. Consider then the random simplicial complex $\hat{Q}_{\rho,D}$ of P_ρ of type D.

We now consider its realizations μ under a new geometric point of view: If

$$t = \sum_{a \in x} \delta_{(a, r_a)}, \ x \in \mathcal{A}_2^{\cdot},$$

is a marked triangle of μ we now interpret the marks r_a in such a way that they define a metric on the edges e of $\alpha(t)$, the augmented t. We give two examples for such an interpretation:

Example 3 (See Thurston [17].) For a given marked triangle t the length of an edge $e = \delta_{(a,r_a)} + \delta_{(b,r_b)}$ is given by $r_a + r_b$. A Euclidean realization of t is given by a configuration of the centres of three balls of radii $r_a, a \in x$, where each two of them touch one another. Such a configuration of balls exists and is unique up to isometry. We speak of the Thurston metric on the edges.

Example 4 (See Reshetnyak [14].) Here the length of an edge $e = \delta_{(a,r_a)} + \delta_{(b,r_b)}$ is $r_a + r_b$ if $d(a, b) \geq \pi$, ($d =$ Euclidean distance). If $d(a, b) < \pi$ then

$$\mathrm{length}(e) = \left(r_a^2 + r_b^2 - 2 \cdot r_a \cdot r_b \cdot \cos\, d(a, b)\right)^{1/2}.$$

This is the length of side BC of a planar triangle ABC, for which length$(A, B) = r_a$, length$(A, C) = r_b$ and angle$(B, A, C) = d(a, b)$. Again any marked triangle t has a Euclidean realization, unique up to isometry.

In this way one can endow the configurations μ of $\hat{Q}_{\rho,D}$ with an intrinsic random metric, which itself depends on chance and which is no longer flat. It is an interesting open question when such random surfaces have an Euclidean realization altogether.

We call $\hat{Q}_{\rho,D}$ the *Poisson–Delaunay surface* with intensity ρ. We know already that $\hat{Q}_{\rho,D} \in \mathcal{P}_0\Gamma$. Moreover, $\hat{Q}_{\rho,D}$ is of first order with respect to the barycentre since P_ρ has this property.

Finally one has

Lemma 4 *$\hat{Q}_{\rho,D}$ is mixing, i.e.*

$$\lim_{|a|\to+\infty} \hat{Q}_{\rho,D}(A \cap T_a B) = \hat{Q}_{\rho,D}(A) \cdot \hat{Q}_{\rho,D}(B),\ A, \mathcal{B} \in \mathcal{F}_{\dot{\Gamma}}.$$

This implies immediately the ergodicity of $\hat{Q}_{\rho,D}$, i.e. $\hat{Q}_{\rho,D}(A) \in \{0, 1\}$ for any translation-invariant event $A \in \mathcal{F}_{\dot{\Gamma}}$.

Proofs of these assertions are contained in [16] (Theorem 6.4.2 of the German edition) resp. [5].

6.3.3 Scholion: The Voronoi Cluster Property

Here we present shortly the alternative construction of Delaunay tesselations by means of Voronoi tesselations as one can find them in the book of Schneider/Weil [16]. Given $a \in \mathbb{R}^d$ and $\eta \in \mathcal{C}^{\cdot}$ the associated *Voronoi polytope* is defined by

$$V(a, \eta) = \{b \in \mathbb{R}^d \mid d(b, a) \leq d(b, \eta)\}.$$

Let $\mathcal{E}\ V(a, \eta)$ denote the set of its extreme points. The Voronoi cell belonging to (a, η) is given by $y + \delta_a$, the set y augmented by a, where $y = \mathcal{E}\ V(a, \eta)$. The *Voronoi cluster property* is then defined by the following measurable subset D_v of $\mathbb{R}^d \times \mathcal{M}^{\cdot}_f \times \mathcal{M}^{\cdot}$:

$$((a, x, \eta) \in D_v \text{ iff } \eta \in \mathcal{C}^{\cdot}, a \in x, x - \delta_a = \mathcal{E}\ V(a, \eta)).$$

We call then $x - \delta_a$, the set x without its element a, the *Voronoi cluster* for η centred in a. The Voronoi configuration belonging to η is $\eta^* = \operatorname{supp} \sum a \in \eta\, b \in \mathcal{E} \sum V(a, \eta)\delta_b$.

D_v is stationary and locally finite. Moreover, one can show that the counting variable

$$cd_{D_v} : \eta \to \sum_{a\in\eta} \sum_{a\in x\subseteq\eta^*+\delta_a} 1_{D_v}(a, x, \eta)$$

satisfies $P_{z\cdot\lambda}\{cd_{D_v} \geq 1\} > 0$. Thus given $\rho = z\cdot\lambda\otimes\tau$, the associated *Voronoi cluster process* $Q = Q_{\rho,D_v}$ is well defined as the image of P_{ρ,D_v} under the measurable transformation

$$\gamma_{D_v} : \nu \mapsto \kappa = \sum_{(a,t)\in\nu} \sum_{(a,t)\in z\subseteq\nu^*+\delta_{(a,t)}} 1_{\mathcal{D}_v}((a, t), z, \nu) \cdot \delta_{(q(z-\delta_{(a,t)}),t)}.$$

Thus Q realizes configurations κ of marked Voronoi cells. One can show that Q-a.e. κ is a marked tesselation of $\mathbb{R}^d$.

We are now in the position to recover in this context again the *Delaunay cluster process*. To be able to do this augment the realizations κ of Q by all vertices of the cells, i.e. let $\tilde{Q}$ denote the image of Q under $\alpha : \kappa \to \kappa + \kappa_{(0)} =: \mu$.

The *Delaunay cluster property* is then defined by the following measurable subset $\mathbb{D}_{\mathrm{el}}$ of $(\mathbb{R}^d \times I) \times \mathcal{M}^{\cdot}_f(X) \times \tilde{\Gamma}$:

$$\left(((a,t), \nu, \mu) \in \mathbb{D}_{\mathrm{el}} \text{ iff } \kappa_{(0)} \in \mathcal{C}^{\cdot}, (a,t) \in \kappa_{(0)}, \nu = \sum_{(a,t)\in y \in \kappa} \delta_y \right).$$

Here X denotes the collection of pairs (x,t), x representing the extreme points of a polytope and $t > 0$; $\tilde{\Gamma}$ denotes the collection of all μ. One can show that $\tilde{Q}$-a.a. μ tesselates $\mathbb{R}^d$, and that each Delaunay cluster consists of $d+1$ polytopes. (See [16] for a proof.)

6.4 Ergodic Behaviour of $\mathcal{PD}$-Surfaces

Until now we gave a construction of stationary random surfaces. Next we study its curvature and other metrical properties. For this we need the notions of

6.4.1 Palm Measures and Palm Distributions

We recall Mecke's celebrated approach to Palm measures (see Mecke [10]). Remember that $d = 2$.

Let $P \in \mathcal{P}_0\Gamma$. (Recall also that Γ is the set of configurations of simplicial complexes, built on marked triangles in $\mathbb{R}^2$, augmented by all their edges and vertices.) Consider the barycentre defined by

$$b : \mathcal{A}^{\cdot} \to \mathbb{R}^2, b(z) = \text{barycentre of } qz,$$

where $\mathcal{A}^{\cdot}$ denotes the collection of all affinely independent marked subsets of $\mathbb{R}^2$. b is a measurable transformation. The intensity measure of P with respect to the barycentre is defined by

$$\nu^1_{bP}(B) := \int_{\Gamma} b\mu(B) P(d\mu),\ B \in \mathcal{B}(\mathbb{R}^2).$$

We assume that $\nu^1_{bP} \in \mathcal{M} \setminus \{0\}$, i.e. ν^1_{bP} is not the trivial measure and

$$\nu^1_{bP}(\Lambda) < +\infty \text{ for any } \Lambda \in \mathcal{B}_0(\mathbb{R}^2). \tag{12}$$

This implies that bP is a simple point process in $\mathbb{R}^2$. Moreover bP is stationary. Thus ν^1_{bP} is translation invariant, so that it is the multiple of the Lebesgue measure: $\nu^1_{bP} = z' \cdot \lambda$, $z' > 0$.

In this situation one can define the *Palm measure* of P with respect to the barycentre for measurable $\varphi \geq 0$ by

$$P^0(\varphi) := \int_\Gamma \int_{F_0} \varphi(\mu - a) b\mu(da) P(d\mu).$$

Here F_0 denotes the cube in $\mathbb{R}^2$, centred in 0 with $\lambda(F_0) = 1$.

Observe that $P^0(\Gamma) = \nu^1_{bP}(F_0) < +\infty$. Therefore P^0 is a finite measure on Γ. Moreover, P^0 is concentrated on $\Gamma^0 = \{\mu \in \Gamma \mid 0 \in b\mu\}$, the collection of all simplicial complexes possessing 0 as a barycentre. We denote in the sequel by $z_0(\mu)$ the unique cell in $\mu \in \Gamma^0$ possessing 0 as a barycentre and call it the *typical cell* of μ.

The *Palm distribution* is given by

$$P_0 = \frac{1}{\nu^1_{bP}(F_0)} \cdot P^0.$$

Note that $P_0 \in \mathcal{P}\Gamma^0$, i.e. P_0 is a random simplicial complex. Let $\Gamma_{(k)} = \{\mu \in \Gamma \mid 0 \in b\mu_{(k)}\}$, $k = 0, 1, 2$. Here $\mu_{(k)}$ is the subconfiguration of μ consisting of all k-dimensional cells of μ. Note that $\Gamma^0 = \Gamma_{(0)} \cup \Gamma_{(1)} \cup \Gamma_{(2)}$.

We call a point process $P \in \mathcal{P}_0\Gamma$ satisfying condition (12) a *stationary random surface*. We remark that the Poisson – Delaunay surface $\hat{Q}_{\rho,D_{\mathrm{el}}}$ belongs to this class.

6.4.2 An Ergodic Theorem for Intrinsic Metric Quantities of Stationary Random Surfaces

Let P be a stationary random surface. Assume that we are given some

$$g \in \mathcal{L}^1(P^0), \tag{13}$$

i.e. g is integrable with respect to P^0.

Interpretation: Given $\mu \in \Gamma^0$, let $z_0(\mu)$ be the typical cell of μ, i.e. the unique cell of μ with barycentre 0. $g(\mu)$ then should be imagined as a metrical aspect of the typical cell.

We are interested in the variables

$$\varphi_\Lambda(\mu) = \int_\Lambda g(\mu - a) b\mu(da),\ \Lambda \in \mathcal{K},\ \mu \in \Gamma,$$

where $\mathcal{K}$ denotes the collection of convex, compact subsets of $\mathbb{R}^2$, including $\emptyset$.

Lemma 5 $(\varphi_\Lambda)_{\Lambda\in\mathcal{K}}$ *is a family contained in* $\mathcal{L}^1(P)$ *having the following properties of a so-called spatial process:*

(1) $\varphi_\emptyset \equiv 0$*,*
(2) (additivity) for $\Lambda_1, \Lambda_2 \in \mathcal{K}$ *disjoint*

$$\varphi_{\Lambda_1\cup\Lambda_2} = \varphi_{\Lambda_1} + \varphi_{\Lambda_2},$$

(3) (invariance) for any $\Lambda \in \mathcal{K},\ b \in \mathbb{R}^2$

$$\varphi_{\Lambda-b} \circ T_b = \varphi_\Lambda.$$

(The proof is an easy exercise.)

We finally observe that the integrability condition (13) implies that for any $\Lambda \in F_0 \cap \mathcal{B}(\mathbb{R}^2)$ and $\mu \in \Gamma$

$$|\varphi_\Lambda|(\mu) \leq \sum_{a\in b\mu\cap F_0} |g|(\mu - a) =: Y,$$

where $Y \in \mathcal{L}^1(P)$. In this situation we have the following *ergodic theorem* (Nguyen/Zessin [12]):

Theorem 4 *If* P *is a stationary random surface satisfying condition (13), then*

$$\lim_{n\to\infty} \frac{1}{\nu^1_{bP}(F_n)} \cdot \int_{F_n} g(\mu - a) b\mu(da) = \frac{1}{\nu^1_{bP}(F_0)} \cdot E_P(\varphi_{F_0} \mid \mathcal{J}) \tag{14}$$

P*-a.e. and in* $\mathcal{L}^1(P)$*. Here* $\mathcal{J}$ *denotes the sub-*σ*-field of* $\mathcal{F}_{\dot\Gamma}$ *consisting of all events, which are translation invariant. And* F_n *denotes the cube in* $\mathbb{R}^2$*, centred in* 0*, with edge length* $2n+1$*. In particular the limit on the right-hand side of (14) is invariant under translations and*

$$E_P\left(\frac{1}{\nu_{bP}(F_0)} \cdot E_P(\varphi_{F_0} \mid \mathcal{J})\right) = P_0(g).$$

Important for us will be the

Corollary 3 *If under the conditions above* P *is also ergodic, i.e.* $\mathcal{J}$ *contains only events of probability* 0 *or* 1*, then*

$$\lim_{n\to\infty} \frac{1}{\nu^1_{bP}(F_n)} \cdot \int_{F_n} g(\mu - a) b\mu(da) = P_0(g)$$

P*-a.e. and in* $\mathcal{L}^1(P)$*.*

6.4.3 Curvature Properties of the Poisson–Delaunay Surface

The underlying stationary random surface is now given by $P = \hat{Q}_{\rho,D}$, where D is the Delaunay cluster property D_{el}. We now give examples of functionals g satisfying the assumptions of the theorem, which are of central importance for quantum gravity.

Example 5 (Curvature) By means of

$$g_c(\mu) = \begin{cases} 2\pi - \sum_{z_0(\mu)\in t\in\mu_{(2)}} \beta(z_0(\mu), t), \text{ if } \mu \in \Gamma_{(0)} \\ 0, \text{else} \end{cases} \tag{15}$$

one defines the deficit angle of μ at its typical cell $z_0(\mu)$, if this is a vertex. Here $\beta(a, t)$ denotes the angle of the triangle t in its vertex a. (Recall that $\mu_{(2)}$ is the subconfiguration of μ consisting of all triangles.) Independently of the choice of the intrinsic metric g_{c} satisfies the estimate

$$(2 - d(\mu)) \cdot \pi < g_{\mathrm{c}}(\mu) < 2\pi, \mu \in \Gamma_{(0)}, \tag{16}$$

where $d(\mu)$ denotes the number of edges of the typical vertex of μ. The associated variable

$$\mathcal{C}_\Lambda(\mu) = \sum_{a\in b\mu\cap\Lambda} g_{\mathrm{c}}(\mu - a), \Lambda \in \mathcal{K},$$

the total sum of deficit angles for the vertices of μ in Λ, is called the curvature of μ in Λ. The integrability condition (13) is satisfied by the following:

Lemma 6 $P^0(|g_{\mathrm{c}}|) \leq 6 \cdot \pi \cdot \nu^1_{bP}(F_0)$.

Proof Using the estimate (16) one obtains

$$P^0(|g_{\mathrm{c}}|) \leq \pi \cdot \int_\Gamma \int_{F_0} d(\mu - a) b\mu(da) P(d\mu) = \pi \cdot P_0(d) \cdot \nu^1_{bP}(F_0).$$

Now it is well known that $P_0(d) = 6$. (See [16], Theorem 6.2.12 of the German edition.) **qed**

Recall that the Poisson – Delaunay surface is mixing and thus ergodic. As a consequence of the corollary of the ergodic theorem we see that

$$\lim_{n\to\infty} \frac{1}{\nu^1_{bP}(F_n)} \mathcal{C}_{F_n}(\mu) = P_0(g_{\mathrm{c}})$$

P-a.s. and in $\mathcal{L}^1(P)$.

Within this example consider the special case of an equilateral $\mathcal{PD}$-surface. This is obtained if the underlying measure τ is a Dirac measure, say δ_{r_0}, and if we choose the Thurston metric of example 3 on the edges of the realizations. This choice produces surfaces composed of equilateral triangles of edge lengths $2 \cdot r_0$. The only source of randomness is then the number of neighbours of a given vertex. All angles $\beta(z_0(\mu), t)$ are equal to $\frac{\pi}{3}$. Nevertheless there is curvature given by

$$g_c(\mu) = 2\pi - \frac{\pi}{3} \cdot d(\mu), \mu \in \Gamma_{(0)}.$$

Thus $g_c(\mu) = 0$ iff $d(\mu) = 6$, negative iff $d(\mu) < 6$ and positive if $d(\mu) > 6$. It is obvious that d is not constant because the origin of randomness is the Poisson process $P_{z \cdot \lambda \otimes \delta_{r_0}}$. We can even compute $P_0(g_c)$:

$$P_0(g_c) = 2\pi - \frac{\pi}{3} \cdot P_0(d) = 0.$$

As a consequence the equilateral $\mathcal{PD}$-surface is an example of a random surface which is asymptotically flat in the sense that its specific curvature $c(\mu) := \lim_{n \to \infty} \frac{1}{\nu^1_{bP}(F_0)} \mathcal{C}_\Lambda(\mu)$ is 0 P-a.e. *(end of the example 5).*

Example 6 The surface measure is defined by

$$g_s(\mu) = \begin{cases} m(z_0(\mu)), & \mu \in \Gamma_{(2)} \\ 0, & \text{else} \end{cases}.$$

Here $m(z_0(\mu))$ is the intrinsic measure induced by the intrinsic metric of the typical cell of μ if it is a triangle. The associated random variable

$$M_\Lambda(\mu) = \sum_{a \in b\mu \cap \Lambda} g_s(\mu - a), \mu \in \Gamma,$$

is then the surface measure of μ in Λ.

Lemma 7 *If τ is supported by a bounded interval I, then*

$$P^0(|g_s|) \leq C \cdot \nu^{1,2}_{bP}(F_0)$$

for some positive constant C. Here

$$\nu^{1,2}_{bP}(F_0) = P^0(\Gamma_{(2)}).$$

In the equilateral situation

$$P_0(g_s) = m(r_0) \cdot \frac{\nu^{1,2}_{bP}(F_0)}{\nu^1_{bP}(F_0)},$$

where $m(r_0)$ *is the intrinsic measure of an equilateral triangle of edge length* $2 \cdot r_0$. *As a consequence, if* τ *is supported by a bounded interval, then*

$$\lim_{n\to\infty} \frac{1}{\nu^1_{bP}(F_n)} \cdot M_{F_n} = P_0(g_s)$$

P-a.s. and in $\mathcal{L}^1(P)$ *(end of the example 6).*

6.5 The Two-Dimensional Regge Model of Pure Quantum Gravity

Let $P \in \mathcal{P}_0\Gamma$ satisfy condition (12). Thus P is a stationary random surface whose vertices are marked by positive numbers.

In a seminal paper from 1961 Regge proposed in [13] to consider the following functional which is a discrete version of the so-called *Einstein-Hilbert action* of the theory of gravitation:

$$g_{\mathcal{R}} := -[g_c - g_s].$$

We assume that

$$g_{\mathcal{R}} \in \mathcal{L}^1(P^0). \tag{17}$$

Observe that this is now an assumption on P, because $g_{\mathcal{R}}$ is given in advance. We saw that a $\mathcal{PD}$-surface is an example if τ is supported by some bounded interval. Let

$$\mathcal{R}_\Lambda(\mu) = \sum_{a\in b\mu\cap\Lambda} g_{\mathcal{R}}(\mu - a),\ \mu \in \Gamma.$$

The two-dimensional *Regge model of pure quantum gravity* is given by

$$(\Gamma, \mathcal{F}_{\Gamma}^{\cdot}, P, g_{\mathcal{R}}),$$

where $P \in \mathcal{P}_0(\Gamma, g_{\mathcal{R}})$, i.e. P is a stationary random surface satisfying (17). In the context of the theory of gravitation realizations $\mu \in \Gamma$ are called *spacetimes*.

We are in the situation of the ergodic theorem. Thus

$$\lim_{n\to\infty} \frac{1}{\nu_{bP}(F_n)} \cdot \mathcal{R}_{F_n} = \mathfrak{r} \qquad P\text{−a.e. and in } \mathcal{L}^1(P),$$

that is, the specific Regge action $\mathfrak{r}$ exists, is stationary and satisfies

$$P(\mathfrak{r}) = -[P_0(g_c) - P_0(g_s)].$$

6.6 Comments and Final Dreams

One of the central open problems then is to characterize the *ground states*. These states minimize the functional

$$\chi : \mathcal{P}_0(\Gamma, g_{\mathcal{R}}) \to \mathbb{R},\ P \mapsto P(\mathfrak{r}) = P_0(g_{\mathcal{R}}).$$

It is believed that such states P have the property that in P_0 almost all μ satisfy the *equations of gravitation.*

This program should be realized in the physically relevant spacetime dimension $d = 3 + 1$ (see [2]).

I add some comments on the relations of the ideas presented here and the work of Ambjørn et al. [3]. In a certain sense we make here a first step in realizing one part of the program formulated in [3]. The authors assume that they are using ‘a correct measure on the set of geometries’. This assumption is realized above: We construct a random spacetime whose realizations are ‘piecewise linear geometries’ to use again their formulation. To be more precise: In the two-dimensional case we construct a probability measure which allows ‘to sum over geometries’ called Poisson–Delaunay surface. Whether this random spacetime has the desired properties, postulated in [3], has to be investigated.

We terminate these lectures with these basic questions.

Comments on some existing rigorous results from stochastic geometry related to quantum gravity: A complete presentation of the above results, including all proofs and technical details like measurability questions and constructions of the main objects, can be found in the more general framework of Thurston processes in the Diplom thesis of Kaiser [5]. (See also [19] on which [5] is based.)

One starting point of activities within the mathematical community of stochastic geometry seems to be the Les Houches lectures of Ambjørn [1] and the book of Ambjørn et al. [3]. There the authors presented for the first time a general and systematic exposition of the mathematical programme behind the numerous activities in this field which took place during the 1990s and formulated the main open mathematical problems. This initiated for instance the work of Vadim Malyshev and his collaborators [7, 8] as well as my own activities [18–20] which are connected with [5, 9].

Within the community of stochastic geometry this programme has not yet found the resonance which it deserves. Until now random curved spacetimes are not in the centre of attention there. Mainly random geometric objects like flat Delaunay or Voronoi tesselations are considered. Curvature does not play any role.

Acknowledgements I am grateful to Otto Kaiser and Kai Matzutt for several illuminating discussions. I also thank the referee for his remarks which led to a substantial improvement of this chapter.

References

1. J. Ambjørn, "Quantization of Geometry", in: J. Dalibard, J.M. Raimond, J. Zinn-Justin, (eds.), Les Houches, Session LIII, 1990. *Systemes Fondamentaux en Optique Quantique/Fundamental Systems in Quantum Optics*. Elsevier Science Publishers B.V. (1995).
2. J. Ambjørn, B. Durhuus, T. Jonsson, *Quantum Geometry*, Cambridge University Press, Cambridge (1997).
3. J. Ambjørn, J. Jurkiewicz, R. Loll, "Quantum gravity as sum over spacetimes", Lect. Notes Phys. **807**. Springer, Heidelberg (2010).
4. B. Delaunay, *Sur la sphère vide*, Bull. Acad. Sci. URSS VI, Class. Sci. Math. Nat., 793–800 (1934).
5. O. Kaiser, *Das ergodische Verhalten der Lipschitz-Killing-Krümmung stationärer Thurston-Prozesse*, Diplomarbeit, Universität Bielefeld (2008).
6. K. Krickeberg, *Processus ponctuels en statistique*, École d'été de St. Flour X-1980, Lect. Notes Math **929**, (1982) 206–313.
7. V.A. Malyshev. *Probability related to quantum gravity. Planar gravity*, Russ. Math. Survey **54** (4), (1999) 685–728.
8. V.A. Malyshev. A.A. Yambartsev, A.A. Zamyatin, *Two-dimensional Lorentzian models*, Moscow Math. Journ. **1** (3), (2001) 439–456.
9. K.M. Matzutt, *Konstruktion zufälliger, lokal endlicher Mosaike, insbesondere Laguerrescher,* Diplomarbeit, Universität Bielefeld (2006).
10. J. Mecke, *Stationäre zufällige Maße auf lokalkompakten Abelschen Gruppen*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **9**, (1967) 36–58.
11. M.G. Mürmann, *Poisson point processes with exclusions*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **43**, (1978) 23–37.
12. X.X. Nguyen, H. Zessin, *Punktprozesse mit Wechselwirkung*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **37**, (1976) 91–126.
13. T. Regge, *General relativity without coordinates*, Nuovo Cimento **19**, (1961) 558–571.
14. Yu.G. Reshetnyak, *Two-dimensional manifolds of bounded curvature*, in: Reshetnyak, Yu.G. (ed.), *Geometry IV*, Encyclopaedia of Mathematical Sciences, vol. 70, Springer, Berlin (1993).
15. B.D. Ripley, *Locally finite random sets: Foundations for point process theory*, Annals of Probability **4**, (1976) 983–994.
16. R. Schneider, W. Weil, *Stochastic and Integral Geometry*, Springer, Berlin (2008). German edition: *Stochastische Geometrie*, Teubner, Stuttgart (2000).
17. W.P. Thurston, *The geometry and topology of three manifolds*, Electronic version 1.1, http://www.msri.org/publications/books/gt3m/ (2002).
18. H. Zessin, *Specific index and curvature of random simplicial complexes*, Izv. Nats. Akad. Nauk Armenii Mat. **37**(1), (2002) 70–88.
19. H. Zessin, *The Thurston Process*, Preprint Nr. 171, FSPM, University of Bielefeld, Germany (2005).
20. H. Zessin, *The Gauß-Bonnet theorem for stationary simplicial complexes*, Izv. Nats. Akad. Nauk Armenii Mat. **38** (3), (2003) 75–82.

Chapter 7
Steps Towards Quantum Gravity and the Practice of Science: Will the Merger of Mathematics and Physics Work?

B. Booß-Bavnbek

Abstract I recall general tendencies of the mathematization of the sciences and derive challenges and tentative obstructions for a successful merger of mathematics and physics on fancied steps towards *quantum gravity*. This is an edited version of the opening words to an international workshop *Quantum Gravity: An Assessment*, Holbæk, Denmark, 17–18 May 2008. It followed immediately after the Quantum Gravity Summer School, see `http://QuantumGravity.ruc.dk`

7.1 Regarding the Need and the Chances of Unification

Reading the literature, preparing and attending the Quantum Gravity Summer School, editing of this volume, and discussing with co-editors and contributors provoked the following claims and questions: (1) The natural laboratory for quantum gravity is the universe, of which we have no control. (2) Nevertheless, we have to accept the challenge, and also a new feature: soon a lifetime will be insufficient to verify a new important theory, and only our successors will be able to prove or disprove it. By the way, it took 49 years just for the Casimir effect, let alone Hawking radiation or string theory. (3) Do we really need to unify gravity with the other interactions? Some physicists say "No". However, if gravity stays in isolation, physicists must be afraid that gravitational physics as a subject will die soon. (4) Most innovative work in quantum gravity is balancing on a knife's edge between abstract mathematics and fresh views on physics concepts. A mathematician, however, may have many remaining unanswered questions, both regarding the claims of physics relevance of the most innovative work and regarding the mathematical clearness and reliability of various new concepts and calculations.

In the following, I recall common knowledge on modelling, mathematization, and science history to put the declared "New Paths Towards Quantum Gravity" in a common frame in spite of their scattering and heterogeneity. Some of the following considerations were published in [12] in condensed form before the Summer School

B. Booß-Bavnbek (✉)
NSM/IMFUFA, Roskilde Universitet, P.O. Box 260, DK-4000 Roskilde, Denmark,
`booss@ruc.dk, http://milne.ruc.dk/~Booss`

Booß-Bavnbek, B.: *Steps Towards Quantum Gravity and the Practice of Science: Will the Merger of Mathematics and Physics Work?*. Lect. Notes Phys. **807**, 339–354 (2010)
DOI 10.1007/978-3-642-11897-5_7

as a kind of *platform* for the assessment of our endeavour. For a comparison with mathematization in other frontier fields I refer also to [9] and the recent [11].

7.2 The Place of Physics in John Dee's *Groundplat of Sciences and Artes, Mathematicall* of 1570

The use of mathematical arguments, first in pre-scientific investigations, then in other sciences, foremost in medicine and astronomy and in their shared border region astrology, has been traced way back in history by many authors from various perspectives, Bernal [6], Høyrup [19], and Kline [21].

Globally speaking, they all agree on three mathematization tendencies:

1. The progress in the individual sciences makes work on ever more complicated problems possible and necessary.
2. This accumulation of problems and data demands conscious, planned, and economic procedures in the individual sciences, i.e. an increased emphasis on questions of methodology.
3. This increased emphasis on questions of methodology is as a rule associated with the tendency of mathematization.

All of this applies generally. In detail, we find many various pictures. In his *Groundplat of Sciences and Artes, Mathematicall* of 1570 [14], the English alchemist, astrologer, and mathematician John Dee, the first man to defend the Copernican theory in Britain and a consultant on navigation, pointed out, in best Aristotelian tradition, that it is necessary in the evaluation of mathematization to pay strict attention to the specific characteristics of the application area in question. He postulated a dichotomy between the *Principall* side, pure mathematics, and the *Deriuative* side, i.e. applied mathematics and mathematization. He then classified the applications of pure mathematics according to objects treated:

- *Ascending Application in thinges Supernaturall, eternall and Diuine*,
- *In thinges Mathematicall: Without farther Applications*,

and finally, on the lowest and most vulgar plane in the Aristotelian scheme,

- *Descending Application in thinges Naturall: both Substantiall & Accidentall, Visible & Inuisible & c.*.

Now that history has excluded matters divine from mathematics, we can with some justification ask whether later generations may regard with equal amusement and astonishment the fact that in our time there are a large number of professional mathematicians and physicists, who are completely satisfied with spending their entire lives working in the second, inner mathematical level and who persistently refuse to descend to vulgar applications.

The panorama of the individual sciences and the role that mathematics had to play in them was perfectly clear to John Dee. In our time the matter is somewhat more complex. For a class on quantum gravity, I cannot point out a geodetically perfect picture of today's landscape of mathematization nor of precise border lines between the mathematics and physics addressed. I must treat the matter rather summarily. A summary treatment may have the advantage that in comparison among the mathematization progress in different branches of physics, common problems on one hand and on the other hand special features of the here advocated new paths towards quantum gravity can be seen more clearly.

In the following, I shall restrict myself to the study of dead nature in *physics*, the field which has the highest degree of mathematization on any chosen scale, both quantitatively and qualitatively. To put things in relief, I shall occasionally touch upon the investigation of living matter in *medicine*, the field where one might expect the greatest mathematization advances in our century, and confront our highly speculative branch of mathematical physics with the treatment of financial issues and decision making for commerce and production in *economics*, a field of questionable scientific state that, beyond well founded actuary estimations, lacks unambiguous results and convincing clear perspectives regarding mathematization.

7.3 Delimitation Between Mathematics and Physics

The intimate connection between mathematics and physics makes it difficult to determine the theoretical relevance of mathematics and obscures the boundary between genuinely physical thought and observation on one side and the characteristically mathematical contribution on the other side. Recall Hilbert's perception of probability theory as a chapter of physics in his famous 6th Problem [18]:

> 6. Mathematical Treatment of the Axioms of Physics. The investigations on the foundations of geometry suggest the problem: *To treat in the same manner, by means of axioms, those physical sciences in which already today mathematics plays an important part; in the first rank are the theory of probabilities and mechanics*.

To say it mildly, as Gnedenko did in his comments to the Russian edition of 1969: *Today this viewpoint* (to consider probability theory as a chapter of physics) *is no longer so common as it was around the turn of the century, since the independent mathematical content of the theory of probabilities has sufficiently clearly showed since then*.... With hindsight and in view of the still challenging foundational problems of quantum mechanics, however, we may accept that parts of mathematics and physics can be interlaced in a non-separable way.

Another famous example of that inextricable interlacement is provided by the Peierls–Frisch memorandum of 1940 to the British Government: suggested by the codiscoverer of fission Otto Frisch, the physicist Rudolph Peierls, like Frisch a refugee in Britain, made the decisive feasibility calculation that not tons (as – happily – erroneously estimated by Heisenberg in the service of the Nazis) but only about 1 kg (later corrected to 6 kg) of the pure fissile isotope U_{235} would be needed to make the atomic bomb. Was it mathematics or physics? It may be worth mentioning that Peierls was a full professor at the University of Birmingham since 1937 and became joint head of *mathematics* there [16]. Theoretical physics in Britain is often in mathematics. As a matter of fact, physics in our sense did not exist as a single science before the 19th century. There were well-defined *experimental physics* comprising heat, magnetism, electricity, and colour, leaving mechanics in mathematics, see [19, p. 493].

In spite of that intermingling, physics can provide a ready system of categories to distinguish different use of mathematics in different modelling situations. Perhaps, the situation can be best compared with the role of physics in general education. After all, physics appears as the model of mathematization: there is no physics without mathematics – and, as a matter of fact, learning of mathematics is most easy in a physics context: calculation by letters; the various concepts of a function (table, graph, operation) and its derivatives and anti-derivatives; differential equations; the concept of observational errors and the corresponding estimations and tests of hypotheses; Brownian movements; all these concepts can be explained context-free or in other contexts (where some of the concepts actually originated), but they become clearest in the ideally simple applications of physics, which are sufficiently complicated to see the superiority of the mathematization as compared to feelings, qualitative arguments, discussions, convictions, imagination – but simple enough to get through.

7.4 Variety of Modelling Purposes

It may be helpful to distinguish the following modelling purposes:

7.4.1 Production of Data, Model-Based Measurements

Clearly, the public associates the value of mathematical modelling foremost to its *predictive* power, e.g. in numerical weather prediction, and its *prescriptive* power, e.g. in the design of the internal ballistics of the hydrogen bomb; more flattering to mathematicians, the *explanatory* power of mathematization and its contribution to *theory development* yield the highest reputation within the field. However, to the progress of physics, the *descriptive* role, i.e. supporting model-based measurements in the laboratory, is – as hitherto – the most decisive contribution of mathematics. Visco-elastic constants and phase-transition processes of glasses and other soft materials cannot be measured directly. For high precision in the critical region, one measures electric currents through a “dancing” piezoelectric disc with fixed potential and varying frequency. In this case, solving mathematical equations from the fields of electro-dynamics and thermo-elasticity becomes mandatory for the design of the experiments and the interpretation of the data. In popular terms, one may speak of a *mathematical microscope*, in technical terms of a *transducer* that becomes useful as soon as we understand the underlying mathematical equations.

7.4.2 Simulation

Once a model is found and verified and the system's parameters are estimated for one domain, one has the hope of doing computer calculations to predict what new experiments in new domains (new materials, new temperatures, etc.) should be made and what they might be expected to show. Rightly, one has given a special name of honour to that type of calculations, *computer simulations*: as a rule, it requires to run the process on a computer or a network of computers under quite sophisticated conditions: typically, the problem is to bring the small distances and time intervals of well-understood molecular dynamics up to reasonable macroscopic scales, either by aggregation or by Monte Carlo methods – as demonstrated by Buffon's needle casting for the numerical approximation of π.

One should be aware that the word *simulation* has, for good and bad, a connotation derived from NASA's space simulators and Nintendo's war games and jukeboxes. Animations and other advanced computer simulations can display an impressive beauty and convincing power. That beauty, however, is often their dark side: simulations can show a deceptive similarity with true observations, so in computational fluid dynamics when the numerical solution of the Bernoulli equations, i.e. the linearization of the Navier–Stokes equations for laminar flow, displays eddies characteristic for the non-linear flow. The eddies do not originate from real energy loss due to friction and viscosity but from hardly controllable hardware and software properties, the chopping of digits, thus providing a *magic realism*, as coined by Abbott and Larsen [1, 2]. In numerical simulation, like in mathematical statistics, results which fit our expectations too nicely must awake our vigilance instead of being taken as confirmation.

7.4.3 Prediction

As shown in the preceding section, there is no sharp boundary between description and prediction. However, the quality criteria for predictions are quite simple: do things develop and show up as predicted? So, for high-precision astrology and longitudinal determination in deep-sea shipping, the astronomical tables of planetary movement, based on the outdated and falsified Ptolemaic system (the *Resolved Alfonsine Tables*) and only modestly corrected in the *Prutenic Tables* of 1551, were, until the middle of the 17th century, rightly considered as more reliable than Kepler's heliocentric *Rudolphine Tables*, as long as they were more precise – no matter on what basis, see Steele [31, p. 128].

Almost unnoticed, we have had a similar revolution in weather prediction in recent years: the (i) analogy (*synoptic*) methods of identifying a similarly looking weather situation in the weather card archives to base the extrapolation on it were replaced by almost pure (ii) numerical methods to derive the prediction solely from the thermodynamic and hydrodynamic basic equations and conservation laws, applied to initial conditions extracted from the observation grid. "Almost" because the uncertainty of the interpolation of the grid and the high sensitivity of the evolution equations to initial conditions oblige to repeated runs with small perturbations and human inspection and selection of the most "probable" outcome like in (i). That yields sharp estimates about the certainty of the prediction for a range of up to 10 days. In 9 of 10 cases, the predictions are surprisingly reliable and would have been impossible to obtain by traditional methods. However, a 10% failure rate would be considered unacceptable in industrial quality control.

In elementary particle physics, the coincidence of predictions with measurements is impressive, but also disturbing. I quote from Smolin [30, pp. 12-13]:

> Twelve particles and four forces are all we need to explain everything in the known world. We also understand very well the basic physics of these particles and forces. This understanding is expressed in terms of a theory that accounts for all these particles and all of the forces except for gravity. It's called *the standard model of elementary-particle physics* – or the standard model for short. ... Anything we want to compute in this theory we can, and it results in a finite number. In the more than thirty years since it was formulated, many predictions made by this theory have been checked experimentally. In each and every case, the theory has been confirmed.
>
> The standard model was formulated in the early 1970s. Except for the discovery that neutrinos have mass, it has not required adjustment since. So why wasn't physics over by 1975? What remained to be done?
>
> For all its usefulness, the standard model has a big problem: It has a long list of adjustable constants. ...

We feel pushed back to the pre-Keplerian, pre-Galilean, and pre-Newtonian cosmology built on ad hoc assumptions, displaying clever and deceptive mathematics-based similarity between observations and calculations – and ready to fall at any time because the basic assumptions are not explained.

Perhaps the word *deceptive* is inappropriate when speaking of description, simulation, and prediction: for these tasks, *similarity* can rightly be considered as the highest value obtainable, as long as one stays in a basically familiar context. From a semiotic angle, the very similarity must have a meaning and is indicating something; from a practical angle, questions regarding the epistemological status can often be discarded as metaphysical exaggerations: who cares about the theoretical or ad hoc basis of a time schedule in public transportation – as long as the train leaves on time!

7.4.4 Control

The prescriptive power of mathematization deserves a more critical examination. In physics and engineering we may distinguish between (a) the feasibility, (b) the efficiency, and (c) the safety of a design. A *design* can be an object like an airplane or a circuit diagram for a chip, an instrument like a digital thermometer, TV set, GPS receiver or pacemaker, or a regulated process like a feed-back regulation of the heat in a building, the control of a power station or the precise steering of a radiation collimator in modulated breast cancer therapy. Mathematics has its firm footing for testing (a) in thought experiments, estimations of process parameters, simulations, and solving equations. For testing (b), a huge inventory is available of mathematical quality control and optimization procedures by variation of key parameters.

It seems to me, however, that (c), i.e. safety questions, raise the greatest mathematical challenges. They appear differently in (i) experience-based, (ii) science-based and (iii) science-integrated design. In (i), mathematics enters mostly in the certification of the correctness of the design copy and the quality test of the performance. In (ii), well-established models and procedures have to be modified and re-calculated for a specific application. Experienced physicists and engineers, however, seldom trust their calculations and adaptations. Too many parameters may be unknown and pop-up later: therefore, in traditional railroad construction, a small bridge was easily calculated and built, but then photogrammetrically checked when removing the support constructions. A clash of more than δ_{crit} required re-building. Similarly, even the most carefully calculated chemical reactors and other containers under pressure and heat have their prescribed "Soll-Bruchstelle" (supposed line of fracture) in case that something is going wrong.

The transition from (ii) to (iii) is the most challenging: very seldom one introduces a radically new design in the physics laboratory or engineering endeavour. But there are systems where all components and functions can be tested separately though the system as a whole can only be tested in situ: a new design of a diesel ship engine; a car, air plane, or space craft; a new concept in cryptography. In all these cases, one is tempted to look and even to advocate for mathematical proofs of the safe function according to specification. Unfortunately, in most cases these "proofs" belong rather to the field of fiction than to rigorous mathematics. For an interesting discussion on "proofs" in cryptography (a little remote from physics) see the debate between Koblitz and opponents in [22] and follow-ups in the *Notices of the American Mathematical Society*.

An additional disturbing aspect of science-integrated technology development is the danger of a loss of transparency. Personally, I must admit, I am grateful for most black-box systems. I have no reason to complain when something in my computer is hidden for my eyes, as long as everything functions as it shall or can easily be re-tuned. However, for the neighbourhood of a chemical plant (and the reputation of the company) it may be better not to automatize everything but to keep some aspects of the control non-mathematized and in the hands of the service crew to avoid de-qualification and to keep the crew able to handle non-predictable situations.

A last important aspect of the prescriptive power of the mathematization is its formatting power for thought structure and social behaviour. It seems that there is not so much to do about it besides being aware of the effects.

7.4.5 Explain Phenomena

The noblest role of mathematical concepts in physics is to explain phenomena. Einstein did it when *reducing* the heat conduction to molecular diffusion, starting from the formal analogy of Fick's law with the cross section of Brownian motion. He did it also when *generalizing* the Newtonian mechanics into the special relativity of constant light velocity and again when *unifying* forces and curvature in general relativity.

Roughly speaking, mathematical models can serve physics by reducing new phenomena to established principles; as heuristic devices for suitable generalizations and extensions; and as "a conceptual scheme in which the insights ... fit together" (C. Rovelli). Further below I shall return to the last aspect – the unification hope.

Physics history has not always attributed the best credentials to explaining phenomena by abstract constructions. It has discarded the concept of a ghost for perfect explanation of midnight noise in old castles; the concept of ether for explaining the finite light velocity; the phlogiston for burning and reduction processes, the Ptolemaic epicycles for planetary motion. It will be interesting to see in the years to come whether the mathematically advanced String Theory or the recent Connes–Marcolli reformulation of the Standard Model in terms of spectral triples will undergo the same fate.

7.4.6 Theory Development

Finally, what has been the role of mathematical concepts and mathematical beauty for the very theory development in physics? One example is Johann Bernoulli's purely aesthetic confirmation of Galilean fall law $s = g/2t^2$ among a couple of candidates as being the only one providing the same equation (shape) for his brachistochrone and Huygens' tautochrone [7, p. 395]:

> Before I end I must voice once more the admiration that I feel for the unexpected identity of Huygens' tautochrone and my brachistochrone. I consider it especially remarkable that this coincidence can take place only under the hypothesis of Galilei, so that we even obtain from this a proof of its correctness. Nature always tends to act in the simplest way, and so it here lets one curve serve two different functions, while under any other hypothesis we should need two curves.

Another, more prominent example is the lasting triumph of Maxwell's equations: a world of radically new applications was streaming out of the beauty and simplicity of the equations of electro-magnetic waves!

However, not every mathematical, theoretical, and empirical accumulation leads to theory development. Immediately after discovering the high-speed rotation of the Earth around its own axis, a spindle shape of the Earth was suggested and an infinitesimal tapering towards the North pole confirmed in geodetic measurements around Paris. Afterwards, careful control measurements of the gravitation at the North Cap and at the Equator suggested the opposite, namely an ellipsoid shape with flattened poles. Ingenious mathematical mechanics provided a rigorous reason for that. Gauss and his collaborator Listing, however, found something different in their control. They called the shape *gleichsam wellenförmig* and dropped the idea of a theoretically satisfactory description. Since then we speak of a *Geoid*. For details see Listing [25] and the receent Torge [32, p.3].

7.5 "The Trouble with Physics"

That is the title of an interesting and well-informed polemic by Lee Smolin against String Theory and present main stream physics at large. He notices a *stagnation* in physics, *so much promise, so little fulfillment* [30, p. 313], a predominance of anti-foundational spirit and contempt for visions, partly related to the mathematization paradigm of the 1970s, according to Smolin: *Shut up and calculate*.

Basically, Smolin may be right. Børge Jessen, the Copenhagen mathematician and close collaborator of Harald Bohr, once suggested to distinguish in sciences and mathematics between periods of *expansion* and periods of *consolidation*. Clearly physics had a consolidation period in the first half of the 20th century with relativity and quantum mechanics. The same may be true for biology with the momentous triumph of the DNA disclosure around 1950, while, to me, the mathematics of that period is characterized by an almost chaotic expansion in thousands of directions. Following that way of looking, mathematics of the second half of the 20th century is characterized by an enormous consolidation, combining so disparate fields like partial differential equations and topology in index theory, integral geometry and probability in point processes, number theory, statistical mechanics and cryptography. A true period of consolidation for mathematics, while – at least from the outside – one can have the impression that physics and biology of the second half of the 20th century were characterized merely by expansion, new measurements, new effects – and almost total absence of consolidation or, at least failures and vanity of all trials in that direction.

Indeed, there have been impressive successes in recent physics, in spite of the absence of substantial theoretical progress in physics: perhaps the most spectacular and for applications most important discovery has been the High Temperature Superconducting (HTS) property of various ceramic materials by Bednorz and Müller [4, 5] - seemingly without mathematical or theoretical efforts but only by systematic combinatorial variation of experiments – in the tradition of the old alchemists.

The remarkable advances in fluid dynamics, weather prediction, oceanography, climatic modelling are mainly related to new observations and advances in computer power while the equations have been studied long before.

Nevertheless, I noticed a turn to theory among young experimental physicists in recent years, partly related to investigating the *energy landscapes* in material sciences, partly to the re-discovery of the *interpretational* difficulties of quantum mechanics in recent quantum optics.

7.6 Theory–Model–Experiment

Physics offers an extremely useful practical distinction between *theory*, *model*, and *experiments*. From his deep insight in astronomy, computing, linguistics, and psychology, Peter Naur ridicules such distinctions as "metaphysical exaggeration" in [28]. He may be right. We certainly should not exaggerate the distinction. In this review, however, the distinction helps to focus on differences of the role of mathematics in doing science.

7.6.1 First Principles

By definition, the very core of modelling is mathematics. Moreover, if alone by the stochastic character of observations, but also due to the need to understand the mathematics of all transducers involved in measurements, mathematics has its firm stand with experiments. First principles, however, have a different status: they do not earn their authority from the elegance of being mathematically wrapped, but from the almost infinite repetition of similar and, as well disparate observations connected to the same principle(s). In the first principles, mathematics and physics meet almost on eye level: first principles are also *established* – like mathematics, and are only marginally questioned. To me, the problem with the pretended eternal authority of first principles is that new cosmological work indicates that the laws of nature may also have undergone some development; that some evolutionary relics might have "survived"; and that we had better be prepared to be confronted under extreme experimental conditions, with phenomena and relations which fall out of the range of accredited first principles. The canonical candidate for such a relic is the Higgs particle, whether already observed or not. Participants of the Quantum Gravity Assessment Workshop 2008 will recall Holger Bech Nielsen's contributions.

7.6.2 Towards a Taxonomy of Models

Not necessarily for the credibility of mathematical models, but for the way of checking the range of credibility, the following taxonomy of models may be extremely useful.

The Closing Round Table of the International Congress of Mathematicians (Madrid, August 22-29, 2006) was devoted to the topic *Are pure and applied mathematics drifting apart?* As a panelist, Yuri Manin ([26], see also [27]) subdivided the mathematization, i.e. the way mathematics can tell us something about the external world, into three modes of functioning (similarly Bohle, Booß and Jensen 1983, [8], see also [10]):

1. An *(ad hoc, empirically based) mathematical model* "describes a certain range of phenomena, qualitatively or quantitatively, but feels uneasy pretending to be something more". Manin gives two examples of the predictive power of such models, Ptolemy's model of epicycles describing planetary motions of about 150 BCE, and the standard model of around 1960 describing the interaction of elementary particles, besides legions of ad hoc models which hide the lack of understanding behind a more or less elaborated mathematical formalism of organizing available data.

2. A *mathematically formulated theory* is distinguished from an ad hoc model primarily by its "higher aspirations. A theory, so to speak, is an aristocratic model". Theoretically substantiated models, such as Newton's mechanics, are not necessarily more precise than ad hoc models; the coding of experience in the form of a theory, however, allows a more flexible use of the model, since its embedding in a theory universe permits a theoretical check of at least some of its assumptions. A theoretical assessment of the precision and of possible deviations of the model can be based on the underlying theory.
3. A *mathematical metaphor* postulates that "some complex range of phenomena might be compared to a mathematical construction". As an example, Manin mentions artificial intelligence with its "very complex systems which are processing information because we have constructed them, and we are trying to compare them with the human brain, which we do not understand very well – we do not understand almost at all. So at the moment it is a very interesting mathematical metaphor, and what it allows us to do mostly is to sort of cut out our wrong assumptions. If we start comparing them with some very well-known reality, it turns out that they would not work".

Clearly, Manin noted the deceptive formal similarity of the three ways of mathematization which are radically different with respect to their empirical foundation and scientific status. He expressed concern about the lack of distinction and how that may "influence our value systems". In the words of [10, p. 73]:

> Well founded applied mathematics generates prestige which is inappropriately generalized to support these quite different applications. The clarity and precision of the mathematical derivations here are in sharp contrast to the uncertainty of the underlying relations assumed. In fact, similarity of the mathematical formalism involved tends to mask the differences in the scientific extra-mathematical status, in the credibility of the conclusions and in appropriate ways of checking assumptions and results... Mathematization can – and therein lays its success – make existing rationality transparent; mathematization cannot introduce rationality to a system where it is absent... or compensate for a deficit of knowledge.

Asked whether the last 30 years of mathematics' consolidation raise the chance of consolidation also in phenomenologically and metaphorically expanding sciences, Manin hesitated to use such simplistic terms. He recalled the notion of Kolmogorov complexity of a piece of information, which is, roughly speaking,

> the length of the shortest programme, which can be then used to generate this piece of information... Classical laws of physics – such phantastic laws as Newton's law of gravity and Einstein's equations – are extremely short programmes to generate a lot of descriptions of real physical world situations. I am not at all sure that Kolmogorov's complexity of data that were uncovered by, say, genetics in the human genome project, or even modern cosmology data ... is sufficiently small that they can be really grasped by the human mind.

7.6.3 The Scientific Status of Quantum Gravity as Compared to Medicine and Economics

From the rich ancient literature preserved, see Diepgen [15], Kudlien [23], and, in particular, Jürss [20, 312–315], we can see that the mind-set in Greek medicine already from the fifth century BCE was *ours*: instead of the partition (familiar from earlier and shaman medicine and similar to the mind set preserved, as seen above, in physics until recent times) into an empirical – rational branch (healing wounds) and a religious – magic branch (cure inner diseases), a physiological concept emerged which focused on the patient as an individual organism within a population, with organs, liquids, and tissue, subjected to environmental and dietetic influences and, in principle, open for unconfined investigation of functions, causal relations and the progressive course of diseases. In Hippocratic medicine, we meet for the first time the visible endeavour after a rational surmounting of all problems related to body events.

With a shake of the head, we may read of Greek emphasis and speculations about the body's four liquids or other strange things, like when we recall today the verdict of the medical profession 60 years ago against drinking water after doing sports and under diarrhoea, or their blind trust in antibiotics, not considering resistance aspects at all. Admittedly, we have no continuity of results in medicine, but, contrary to physics, we have an outspoken continuity in mind set: no ghosts, no metaphysical spirits, no fancied particles or relations are permitted to enter our explanations, diagnoses, prevention, cure, and palliation.

Physicists of our time like to date the physics' beginning back to Galileo Galilei and his translation of measurable times and distances on a skew plane into an abstract fall law. Before Galilei – and long time after him, the methodological scientific status of what we would call mechanical physics was quite low as compared with medicine. Physics was a purely empirical subject. It was about precise series of observations and quantitative extrapolations. It was the way to predict planetary positions, in particular eclipse times, the content of silver in compounds, or the manpower required to lift a given weight with given weight arm. It was accompanied and mixed up with all kinds of speculations about the spirits and ghosts at work. We can easily see the continuity of results, of observations and calculations from Kepler and Newton to our time. However, we can hardly recognize anything in their thinking about physics, in the way they connected physics with cosmic music or alchemy or formulated assumptions. We may wonder what later generations will think about our fancied new paths towards quantum gravity.

While a rational point of departure for economics, in particular under the present crisis, can only be a *systems view*, a holistic unifying view in physics like our efforts in quantum gravity have a smell of vanity, "stagnant and stuck" in the words of Baez [3]. One may argue that the time has hardly come for that endeavour – comparable to the felt necessity but still continuing futility of or at least doubts about a holistic all-embracing systems biology programme in medicine.

7.7 General Trends of Mathematization and Modelling

7.7.1 Deep Divide

Regarding the power and the value of mathematization, there is a deep moral divide both within the mathematics community and the public.

On the one side, we have the outspoken science and math optimism of outstanding thinkers: Henri Poincaré's *Nature not only suggests to us problems, she suggests their solution*; David Hilbert's *Wir müssen wissen; wir werden wissen - We must know; we will know* of his Speech in Königsberg in 1930, now on his tomb in Göttingen; or Bertolt Brecht's vision of mathematical accountability in *Die Tage der Kommune* [13] of 1945: "Das ist die Kommune, das ist die Wissenschaft, das neue Jahrtausend... - That is the Commune, that is the science, the new millennium..."). We have astonishing evidence that many mathematization concepts either appear to us as *natural* and *a-priori*, or they use to emerge as clear over time. We have the power and validity of extremely simple concepts, as in *dimension analysis*, *consistency requirements*, and *gauge invariance* of mathematical physics. Progressive movements emphasize science and education in liberation movements and developing countries. Humanitarian organizations (like WHO and UNICEF) preach science and technology optimism in confronting mass poverty and epidemics.

On the other side, deep limitation layers of science and mathematical thinking have been dogged up by Kurt Gödel's *Incompleteness Theorem* for sufficiently rich arithmetic systems, Andrei N. Kolmogorov's *Complexity Theory*, and Niels Bohr's notion of *Complementarity*. Incomprehensibility and lack of regularity continue to hamper trustworthy mathematization. Peter Lax [24, p. 142] writes about the *profound mystery of fluids*, though recognizing that different approaches lead to remarkably coinciding results, supporting reliability.

The abstruseness of the mathematical triumphs of the hydrogen bomb is commonplace. The widespread trust in superiority and invincibility, based on mathematical war technology like high precision bombing, has proved to be even more vicious for warriors and victims than the immediate physical impact of the very math-based weaponry, recently also in Iraq and Afghanistan.

In between the two extremes, Hilbert's optimistic prediction of *clearness* and the sceptical Kafkaesque expectation of increasing *bewilderment* when digging deeper mathematically, we have the optimistic scepticism of Eugene Wigner's *unreasonable effectiveness of mathematics*, but also Jacob Schwartz's verdict against the *pernicious influence of mathematics on science* and Albert Einstein's demand for *finding the central questions* against the dominance of the beautiful and the difficult.

7.7.2 Charles Sanders Peirce's Semiotic View

From the times of Niels Bohr, many physicists, mathematicians, and biologists have been attentive to philosophical aspects of our doing. Most of us are convinced that the frontier situation of our research can point to aspects of some philosophical relevance – if only the professional philosophers would take the necessary time to become familiar with our thinking. Seldom, however, we read something of the philosophers which can inspire us.

The US-American philosopher Charles Sanders Peirce (1839–1914) is an admirable exception. In his semiotics and pragmaticist (he avoids the word "pragmatic") thinking, he provides a wealth of ideas, spread over an immense life work. It seems to me that many of his ideas, comments, and concepts can shed light on the why and how of mathematization. Here I shall only refer some thoughts of Peirce's *The Fixation of Belief* from 1877, see [29].

My fascination of Peirce's text is, in particular, based on the following observations which may appear trivial (or known from Friedrich Engels), but are necessary to repeat many times for the new-modeller:

1. For good and bad, we are all equipped with innate (or spontaneous) orientation, sometimes to exploit, sometimes to subdue. Our innate orientation is similar to the habits of animals in our familiar neighbourhood. We are all "logical machines".
2. However, inborn logic is not sufficient in foreign (new) situations. For such situations, we need methods how to fixate our beliefs. Peirce distinguishes four different methods. All four have mathematical aspects and are common in mathematical modelling.

   Tenacity is our strength not to become confused, not to be blown away by unfounded arguments, superficial objections, misleading examples, though sometimes keeping our ears locked for too long.

   Authority of well-established theories and results is what we tend to believe in and have to stick to. We will seldom drop a mastered approach in favour of something new and unproved.

   Discussion can hardly help to overcome a belief built on tenacity or authority.

   Consequences have to be investigated in all modelling. At the end of the day, they decide whether we become convinced of the validity of our approach (Peirce's Pragmaticist Maxim).
3. The main tool of modelling (i.e. the fixation of belief by mathematical arguments) is the transformation of symbols (signals, observations, *segments of reality*) into a new set of symbols (mathematical equations, *models*, and *descriptions*). The advantage for the modeller, for the person to interpret the signs, is that signs which are hard or humid and difficult to collect in one hand can be replaced by signs which we can write and manipulate.
4. The common mapping cycle *reality* $\rightarrow$ *model* $\rightarrow$ *validation* is misleading. The quality of a mathematical model is not how similar it is to the segment of reality under consideration, but whether it provides a flexible and goal-oriented approach, opening for doubts and indicating ways for the removal of doubts (later trivialized by Popper's falsification claim). More precisely, Peirce claims

- Be aware of differences between different approaches!
- Try to distinguish different goals (different priorities) of modelling as precisely as possible!
- Investigate whether different goals are mutually compatible, i.e. can be reached simultaneously!
- Behave realistically! Do not ask *How well does the model reflect a given segment of the world*? But ask *Does this model of a given segment of the world support the wanted and possibly wider activities/goals better than other models*?

I may add: We have to strike a balance between *abstraction and construction*, *top-down and bottom-up*, and *unification and specificity*. We better keep aware of the variety of *Modelling purposes* and the multifaceted relations between *theory – model – experiment*. Our admiration for the *power of mathematization*, the *unreasonable effectiveness of mathematics* (Wigner) should not blind us for the *staying and deepening limitations of mathematization opposite new tasks*.

References

1. M. Abbott, J. Larsen, J. Tao, *Modelling circulations in depth-averaged flows, Part 1: The accumulation of the evidence*, J. Hydr. Res. **23** (1985), 309–326.
2. M. Abbott, J. Larsen, *Modelling circulations in depth-averaged flows, Part 2: A reconciliation*, J. Hydr. Res. **23** (1985), 397–420.
3. J. Baez, This Week's Finds, Work on quantum gravity has seemed stagnant and stuck for the last couple of years, which is why I've been turning more towards pure math, Week 220 (August 31, 2005), `http://math.ucr.edu/home/baez/week220.html`.
4. J.G. Bednorz, K.A. Müller, "Perovskite-type oxides – The new approach to high-T_c superconductivity", Nobel Lecture, December 8, 1987.
5. G. Ekspong (ed.), *Nobel Lectures in Physics 1981–1990*, World Scientific Publishing Company, Singapore (1993).
6. J. D. Bernal, *Science in History*, vol. 4, Watts, London (1954) 1328p.
7. J. Bernoulli, *Acta Eruditorum* May 1697, 201–224, translated in excerpts in D. Struik (ed.), *Source Book in Mathematics, 1200–1800*, Princeton University Press, Princeton (1986).
8. M. Bohle–Carbonell, B. Booß, J.H. Jensen, "Innermathematical vs. extramathematical obstructions to model credibility", in: X. Avula, (ed.), *Mathematical Modelling in Science and Technology*, Proceedings of the Fourth International Conference (Zürich, August 1983), Pergamon Press, New York (1984), pp. 62–65.
9. B. Booß, *Niakoi tendenzii v matematishaziiata na otdelnite nauki (Some tendencies in the mathematization of the individual sciences, Bulgarian)*. Fiz.-Mat. Spis. B"lgar. Akad. Nauk. **22(55)**(3)(1979), 197–204.
10. B. Booß-Bavnbek, "Against ill-founded, irresponsible modelling", in: M. Niss, et al. (eds.), *Teaching of Mathematical Modelling and Applications*, Ellis Horwood, Chichester (1991), pp. 70–82.
11. B. Booß-Bavnbek, *The mathematization of the individual sciences – revisited, J. Contemp. Math. Anal. (Armen. Acad. Sci.)* **44** (2) (April 2009), 81–96, arXiv:0903.0942v1 [math.HO]. Russian edition: Izvestiya NAN Armenii. Matematika (2009), No. 2, 16–35.
12. B. Booß-Bavnbek, G. Esposito, M. Lesch, "Quantum gravity: unification of principles and interactions, and promises of spectral geometry", *SIGMA (Symmetry, Integrability and Geometry: Methods and Applications)* **3** (2007), pp. 098–127 [arXiv:0708.1705v3 [hep-th]].

13. B. Brecht, *Die Tage der Kommune, in: Stücke 8*, Aufbau-Verlag, Berlin and Weimar, Suhrkamp Verlag, Frankfurt (1992), pp. 243–317, here p. 174.
14. J. Dee, *Mathematicall Praeface to the Elements of Geometrie of Euclid of Megara*, 1570. Kessinger Publishing, Kila (1999), 60p.
15. P. Diepgen, *Geschichte der Medizin. Die historische Entwicklung der Heilkunde und des ärztlichen Lebens*, vol. 1, Walter de Gruyter & Co., Berlin (1949), pp. 67–158.
16. S. Edwards, *Rudolph E. Peierls-Obituary*, Physics Today **49** (2) (February 1996), 74–75.
17. B. Haghighat, A. Klemm, *Topological strings on Grassmannian Calabi-Yau manifolds*, JHEP 0901:029-65, 2009, arXiv:0802.2908v1 [hep-th]
18. D. Hilbert, Mathematische Probleme, Vortrag, gehalten auf dem internationalen Mathematiker-Kongreß zu Paris 1900, *Nachrichten von der Königl. Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-Physikalische Klasse* (1900), 253-297; `http://www.mathematik.uni-bielefeld.de/~kersten/hilbert/rede.html`, with Boris Gnedenko's comments to the 6th problem in: P.S. Alexandrov (ed.), *Problemi Gilberta*, Nauka, Moscow, 1969, German translation in Ostwalds Klassiker der exakten Wissenschaften, vol. 252, Akademische Verlagsgesellschaft, Leipzig (1971), pp. 145–150.
19. J. Høyrup, *From Hesiod to Saussure, from Hippocrates to Jevons–An Introduction to the History of Scientific Thought*, Roskilde 2004; `http://akira.ruc.dk/∼jensh/Publications/FromHesiod_04.pdf`
20. F. Jürss, *Geschichte des wissenschaftlichen Denkens im Altertum*, Akademie-Verlag, Berlin (1982).
21. M. Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, Oxford (1974).
22. N. Koblitz, The uneasy relationship between mathematics and cryptography, *Notices AMS* **54** (8) (September 2007), 972–979; `http://www.ams.org/notices/200708/tx070800972p.pdf`
23. F. Kudlien, *Der Beginn des medizinischen Denkens bei den Griechen, Von Homer bis Hippokrates*, Artemis, Zürich and Stuttgart (1967).
24. P. Lax, *Mathematics and physics*, Bull. Am. Math. Soc. **45** (1) (January 2008), 135–152.
25. J.B. Listing, *Über unsere jetzige Kenntniss der Gestalt und Grösse der Erde*, Nachr. d. Kgl. Gesellsch. d. Wiss. und der Georg August University, Göttingen (5. Februar, No. 3, 1873, pp. 33–98), reprinted September 2007, VDM Verlag, ISBN: 3836422999.
26. Y. Manin, Presentation to the closing round table "Are pure and applied mathematics drifting apart?", in: M. Sanz-Solé, et al. (eds.), *Proceedings of the International Congress of Mathematicians* (Madrid, August 22–30, 2006), European Mathematical Society, Zürich (2007), pp. 762–764, 775–776.
27. Y. Manin, Mathematical knowledge: internal, social and cultural aspects. Commissioned for C. Bartocci, P. Odifreddi (eds.), *Mathematics and Culture*, vol. 2 (2007), math.HO/0703427; also in: *Mathematics as Metaphor: Selected Essays of Yuri I. Manin with Foreword by Freeman J. Dyson*, American Mathematical Society, 2007, pp. 3–26.
28. P. Naur, *Knowing and the Mystique of Logic and Rules*, Kluwer Academic Publishers, Dordrecht (1995).
29. C.S. Peirce, "The fixation of belief", *The Popular Science Monthly* (1878). Reprinted In: Ch. Hartshorne, P. Weiss (eds.), *The Collected Papers of Charles Sanders Peirce*, Vol. V, book 2, Harvard University Press, Cambridge, MA (1934), #358–386, pp. 223–247; `http://milne.ruc.dk/∼Booss/MaMoCalc/WP1_Peirce_Fixation.pdf`. This is the original text by Peirce, but edited for first year students of the Modelling/Calculus Class at Roskilde University.
30. L. Smolin, *The Trouble with Physics*, Allen Lane, Penguin Books, London (2006).
31. J.M. Steele, *Observations and Predictions of Eclipse Times by early Astronomers*, Kluwer Academic Publishers, Dordrecht (2000).
32. W. Torge, *Geodesy*, 3. Completely revised and extended edition, de Gruyter, Berlin-New York (2001).

Index

Booß-Bavnbek, B. et al. (eds.): *Index*. Lect. Notes Phys. **807**, 355–359 (2010)
DOI 10.1007/978-3-642-11897-5