Loops, strings and noncommutative geometry

-

Three roads towards unifying gravity with quantum mechanics and explaining the standard model

Bjørn Solheim

Thesis

(Master of science)

Theoretical Physics

Department of Physics, Faculty of Mathematics and Natural Sciences
University of Oslo

November 1, 2016

Copyright © 2016 Bjørn Solheim
PUBLISHED BY BJØRN SOLHEIM
All rights reserved.
First printing, November 2016

Contents

Preface i

I Doing physics beyond the experimental reach

1 Epistemology 1
1.1 Introduction 1
1.2 The general epistemology of physics 2
1.3 The current epistemological challenges 7
1.4 Notes 9

II Foundations of established theories

2 Physical theories in general 13
2.1 Introduction 13
2.2 Probability spaces 14
2.3 Algebraic formulation of mechanics 19
2.4 Mechanics as a probability theory 23
2.5 Dynamics 24
2.6 Composite systems 26
2.7 Notes 27

3 Fixed background theories - QFT 29
3.1 Introduction 30

3.2 Summary of deducing QFT 30
3.3 Isometries in general physical theories 33
3.4 Isometries in quantum theory 34
3.5 Special relativity 37
3.6 Representations of the Poincaré group and $ISL(2,\mathbb{C})$ 37
3.7 From representations to quantum fields 38
3.8 Gauge invariance from Lorentz invariance 41
3.9 Renormalization and RGE 42
3.10 Symmetry breaking with scalar fields 43
3.11 The standard model 43
3.12 Notes 44

4 Background free theories - GR 45
4.1 Introduction 46
4.2 Equivalence of gravitational and inertial forces 46
4.3 Equivalence principles 48
4.4 The Equivalence Principle makes a geometric formulation possible 50
4.5 Diffeomorphism 50
4.6 Background independence 51
4.7 A diffeomorphism invariant action 52
4.8 Dynamics of background free theories 53
4.9 The spin-2 approach to gravity 54
4.10 Notes 55

III Extended theories I - Noncommutative geometry

5 Noncommutative geometry - mathematics 59
5.1 Introduction 60
5.2 Noncommutative topology 61
5.3 Noncommutative measure theory 67
5.4 Vector Bundles 67
5.5 Differential calculus 68
5.6 Connections on modules 71
5.7 Spin manifolds 72

6 Noncommutative geometry - physics overview 77
6.1 Introduction 78
6.2 Notation 79
6.3 Inner fluctuations 81
6.4 Gauge structure 85
6.5 Unimodularity 87

6.6 Action principles 88
6.7 Deriving the action 89
6.8 Deriving the finite algebra 91
6.9 A quantum geometric derivation of the algebra 94
6.10 Notes 94

7 Noncommutative geometry - physics details 95
7.1 Introduction 95
7.2 Useful formulas 96
7.3 Notation 97
7.4 Some general preparations 97
7.5 U(1) gauge theory 99
7.6 QED 105
7.7 Electroweak theory 108
7.8 The standard model 116
7.9 The Pati-Salam model 128
7.10 Notes 129

IV Extended theories II - Loop quantum gravity

8 Loop quantum gravity 133
8.1 Introduction 133
8.2 Classical theories with constraints 135
8.3 Quantization of theories with constraints 148
8.4 The Ashtekar formulation of classical GR 153
8.5 The graph formulation of GR 164
8.6 Quantization of the graph formulation 173
8.7 Implementing constraints 177
8.8 Dynamics in LQG 185
8.9 The classical limit 186
8.10 Black hole entropy and entanglement entropy 186
8.11 Cosmology 187
8.12 LQG in higher dimensions 193
8.13 Notes 194

V Extended theories III - String theory

9 String theory 199
9.1 Introduction 200
9.2 Generalities 200

9.3 A historical prelude 202
9.4 Introduction to perturbative string theory 206
9.5 Classical bosonic string theory 207
9.6 Quantized bosonic string theory 211
9.7 Classical superstring theories 213
9.8 Heterotic string theories 218
9.9 Quantized superstring theories 218
9.10 String interactions 222
9.11 Conformal field theory 225
9.12 Foundations of perturbative string theory 226
9.13 D-Branes 227
9.14 String dualities 229
9.15 Lower dimensional string theory 231
9.16 String compactifications 233
9.17 Brane worlds 238
9.18 Moduli stabilization 239
9.19 String theory landscape 240
9.20 Gravity in string theory 241
9.21 Black hole thermodynamics 243
9.22 Gauge/gravity duality 245
9.23 Entanglement entropy 252
9.24 Notes 259

VI Extended theories - Evaluation

10 Evaluation and conclusion 265
10.1 Evaluation of NCG 266
10.2 Evaluation of LQG 269
10.3 Evaluation of ST 274
10.4 Concluding remarks 280

VII Appendices

Appendices 285

Postscript - Ideas for future research 287

A Logic and sets 293
A.1 Formal systems 293
A.2 Logic 294
A.3 Axiomatic set theory 295

A.4 Relations and functions 297
A.5 Mathematics from axioms 298
A.6 Notes 300

B Category theory 301
B.1 Introduction 301
B.2 Basic concepts 302
B.3 Functors 304
B.4 Notes 305

C Algebra 307
C.1 Algebraic structures 307
C.2 Graph theory 312
C.3 Clifford algebras and spinors 312
C.4 Graded algebras 314
C.5 Notes 315

D Measure theory and probability 317
D.1 Measure theory and integration 317
D.2 Probability theory 318
D.3 Notes 319

E Topology 321
E.1 Basic topology 321
E.2 Further concepts 324
E.3 Derived topological spaces 326
E.4 Notes 327

F Differential geometry 329
F.1 Differential topology 329
F.2 Vectors 331
F.3 Differential operations 333
F.4 Symplectic manifolds 334
F.5 Bundles 336
F.6 Connections 338
F.7 Riemannian manifolds 338
F.8 Complex manifolds 339
F.9 Various topics 339
F.10 Notes 340

G Lie groups and Lie algebras 341
G.1 Lie groups and Lie algebras 341

G.2 Representation theory 342
G.3 Harmonic analysis 343
G.4 Specific groups and algebras 344
G.5 Notes 344

H Functional analysis ... 345
H.1 Basic concepts 345
H.2 Sequences and completeness 346
H.3 Infinite dimensional vector spaces 347
H.4 Operators and norms 348
H.5 Algebras with involutions 349
H.6 Differential operators 353
H.7 Various results 354
H.8 KK-theory 354
H.9 Differential algebras 355
H.10 Functional analysis 357
H.11 Notes 358

I Algebraic topology ... 359
I.1 Homology, cohomology, homological algebra concepts 359
I.2 Notes 366

J Other math ... 367
J.1 Various 367
J.2 Almost periodic functions 367
J.3 Fredholm operators 368
J.4 Real and complex structure 368
J.5 Notes 369

Quotes ... 369

References ... 372

Index ... 405

Preface

"Only one need absorbs me: I must win clarity else I cannot live; I cannot bear life unless I can believe that I will achieve it."

Husserl

Physics today is both supremely impressive and profoundly unsatisfying. It is extremely impressive in the sense that we can in principle accurately predict nature all the way from the level of quarks to the observable universe. It is conceptually very unsatisfying in the sense that it builds on two incompatible theories, namely the general theory of relativity and quantum field theory[1]. This thesis is an independent review and critical evaluation of three leading attempts to create improved theories, that extend the empirical success of current theories, while eliminating their conceptual problems.

The three theories we review are noncommutative geometry (NCG), loop quantum gravity (LQG) and string theory (ST). We center the text around two questions of great importance for theoretical physics: "How can we combine gravity with quantum theory?" and "How can we explain the structure of the standard model?". Throughout the analysis we pay particular attention to the conceptual foundation and mathematical underpinnings of each theory. The three approaches are currently quite different with regards to which of the two questions they are mostly focused on, but they all have the potential and ambition to address them both.

To properly understand why these two question are so important, and why they are so difficult to answer, requires some preparation. The two questions we focus on have their roots in the structures of general relativity (GR) and quantum field theory (QFT). We believe that we must first clarify the foundations of GR and QFT before we can approach our three reviews of extended theories in the proper perspective.

Based on this we spend chapter 1–4 making an analysis of the current foundations of physics. This analysis is based on numerous other works, but it provides what we believe to be a compact yet comprehensive basis aimed at facilitating clarity and generalization. In chapter 1 we focus on methodological challenges and clarify why progress in theoretical physics must likely come

[1] In the form of the standard model

from an increased understanding of foundational issues and by utilizing the idea of conceptual satisfactoriness. In chapter 2 we construct a unified compact algebraic formulation of classical and quantum theories. In chapters 3 and 4 we make a detailed analysis of how the structure of a theory is dictated by being defined on a fixed background (QFT) or being background free (GR).

In chapters 5–9 we then review the theories we have selected and we use chapter 10 to give an evaluation of how well the three approaches are able to answer the two questions listed above. Sections 10.1–10.3 evaluate the individual theories while section 10.4 summarizes what we have uncovered in these evaluations as well presenting our final conclusions.

It is integral to our approach to consider all three theories in a single thesis as we are convinced that it is the distilled total knowledge collected in these approaches that has the potential to go beyond the forty year impasse[2] in theoretical physics. It is therefore important to try and find connections between the three theories. In sections 9.21–9.23 and in section 10.4 we explore how the notions of holography and entanglement entropy are promising avenues towards finding common ground between these three approaches, and how they also represent important keys to answering the two questions of this thesis. In a postscriptum to the thesis we make some radical suggestions that goes even further in this direction.

There are few new equations and new calculations in this thesis. We have made a choice to focus on the conceptual and foundational side of physics. The potential contribution of this work lies in the conceptual analysis and in the presentation. We try to be swift and succinct but not to the extent of being readable only to experts. An effort has been made to guide the reader by chapter summaries, introductions and diagrams. An appendix with mathematical details is also provided to serve as a quick information repository for the reader. Since tools from multiple mathematical areas are needed we have chosen to put most definitions and finer details in the appendix, so as not to disrupt the flow of the main text too much.

The methodological insights of chapter 1 dictate the approach of later chapters. Based on the belief that the amount of relevant empirical data will be sparse, we focus on foundational issues, consistency, and the potential for generalizations. Phenomenology is the essential starting point, and remains the ever important epistemological basis of physics, but phenomenological knowledge alone should not be considered the desired endpoint of physics. Seeking deeper explanations for phenomenologically successful theories sets a higher target for physics, but this also provides new tools. We believe that using the criteria of an intellectually and conceptually satisfying explanation, despite sounding very human-centric and subjective, can actually be a prime guidance tool towards finding a more unified formulation of the foundations of physics, that again can lead to a new synthesis that takes us beyond the current understanding.

In essence, the text is about how to find answers to two challenging questions in a situation where there is little help to be found from new empirical input. From searching within we shall find the means to go beyond, and from exploring the deepest reaches of current knowledge we shall discover paths to the next level of understanding.

Author contributions

The concept, the analysis and the framing of this thesis are my original work. All chapters are written with the intent of providing a fresh and up-to-date presentation of mostly rapidly evolving subjects. With the exception of chapter 7, no other chapter leans heavily on any particular references. Chapter 1 is an independent contribution by the author, though it hopefully just encodes a view of physics that most physicists can agree on. I know of no reference that presents anything close to the full content of chapter 2. Chapter 3 adopts several parts of logic from Wald, Weinberg and Arkani-

[2]Fundamental theoretical physics has made progress in the latest decades but the established core of theoretical physics has not changed since the standard model was accepted forty years ago.

Hamed but the presentation is different. No individual part of chapter 4 is strictly original, but to my knowledge such an analysis is not presented elsewhere. Chapter 5, 6, 8 and 9 is my attempt at an independent review of central parts of noncommutative geometry, loop quantum gravity and string theory with a view towards conceptual foundations and the two formative questions of the thesis. Chapter 7 leans substantially on the work of von Suijlekom and his group (which again extend and complete work by Connes, Marcoli and Chamseddine), but I have tried to improve further on the clarity of the arguments. Chapter 10 is an original analysis not based on any particular reference where we also present ideas for further research. In a separate postscriptum section we present an extended set of suggestion for future research in the form of a template theory.

Acknowledgements

I wish to thank my supervisor Olav Syljuåsen for his help and support. I also wish to thank J. Lye and P.D. Prester for helpful comments on chapter 9. I thank N.F., H.E. and J.S. for their comments. A special thanks goes to K.D. for invaluable assistance. I also extend my gratitude to all the others who have supported me during the preparation of this thesis.

Bjørn Solheim, Oslo 2016

Abbreviations

AC - Almost commutative (manifold or spectral triple)
AdS - Anti de Sitter (space) ADM - Arnowitt, Deser and Misner (gravitational action, GR formulation)
BPS - Bogomol'nyi-Prasad-Sommerfield (limit or bound)
BSM - Beyond standard model
CDP - Cluster decomposition principle
CFT - Conformal field theory
CLC - Connes-Lott-Chamseddine (models)
CM - Classical mechanics
CY - Calabi-Yau (manifolds)
DGA - Differential graded algebra
DIFF - Diffeomorphisms
dS - de Sitter space
EEP - Einstein equivalence principle
EFT - Effective field theory
EH - Einstein-Hilbert (action)
EW - Electroweak (theory)
F-LOST - Fleischhack-Lewandowski-Oklow-Sahlman-Thiemann (uniqueness theorem)
FLRW - Friedmann-Lemaitre-Robertson-Walker (cosmologies)
FODC - First-order differential calculus
GNS - Gelfand-Naimark-Segal (construction)
GR - General relativity
GUT - Grand unified theory
GWEP - Gravitational weak equivalence principle
HF - Holonomy-flux (algebra)
INV. - Invariance
IR - Infrared (i.e. low energy)
IRREP - Irreducible representation
JLB - Jordan-Lie-Banach algebra
LQG - Loop quantum gravity
LQC - Loop quantum cosmology
MSSM - Minimal supersymmetric standard model
mSUGRA - minimal Supergravity NCG - Noncommutative geometry
QCD - Quantum chromodynamics
QED - Quantum electrodynamics
QFT - Quantum field theory
QM - Quantum mechanics
QP - Quantum physics
QG - Quantum gravity
RG - Renormalization group
RGE - Renormalization group equation
SCFT - Superconformal field theory
SM - Standard model (of particle physics)
SEP - Strong equivalence principle
SR - Special relativity
ST - String theory
UV - Ultra violet (i.e. high energy)
WdW - Wheeler de Witt (equation)

WEP - Weak equivalence principle
WIMP - Weakly interacting massive particle
YM - Yang-Mills (theory)
ZF - Zermelo-Fraenkel (axiomatic set theory)
ZFC - ZF + axiom of choice

Chapter summaries

1 – Doing physics beyond the experimental reach – Epistemology

Physics must be empirically correct but should also be clear, consistent and conceptually satisfying. The number of new decisive experiments beyond current experimental limits will be few. Theory development cannot rely on experiments alone but must instead be driven by focusing on making things clear, consistent and conceptually satisfying. A clear, consistent and satisfying formulation of current physics is likely to be an important key to completing physics in empirically sparse domain of quantum gravity. Rigorous mathematical axiomatizations form a useful parametrization of the theory space in the same way as model building tools help to explore the parameter space of particle physics.

2 – Foundations of established theories – Physical theories in general – CM and QM

Classical and quantum systems can be described by a common three-part framework of observables, states and dynamics. In mathematical terms this consists of a *-algebra (representing observables), a set of linear algebra functionals (representing physical states), and a one-parameter family of algebra automorphisms (representing time-evolution or dynamics). Classical and quantum systems are distinguished solely by the commutativity or noncommutativity of the algebra of observables. Both quantum and classical theories can be seen as probability theories with the added structure of algebra automorphisms to represent dynamics. The fundamental use of probabilities, and the reliance on formulas like the Born rule, are common to both approaches. Incompatible observables, the uncertainty principle and entanglement are quantum properties that only occur when using a noncommutative algebra of observables.

3 – Foundations of established theories – Physical theories in static spacetime – QFT

QFTs are quantum theories defined on a fixed background spacetime. Such theories derive their structure from representations of the background space isometry group. The isometry group of Minkowski space is the Poincaré group. Finding irreducible representations of this group leads almost directly to Yang-Mills theories with spin-1/2 fermions and scalar fields. Gauge redundancies are an inevitable consequence of implementing massless spin-1 Poincaré representations as vector fields. Spontaneous breaking of gauge symmetries by scalar fields is a naturally occurring possibility in such models. In the low energy EFT limit, all non-renormalizable interactions are suppressed. There is a substantial number of free parameters and one specific set of parameter values defines the standard model.

4 – Foundations of established theories – Physical theories in dynamic spacetime – GR

The equivalence principle leads us to seek a background free description of gravity in terms of dynamic geometry. The associated symmetry of a background free description is the group of diffeomorphisms. General relativity (GR) is the unique lowest order approximation of such a theory. Classical systems like GR defined on a dynamic spacetime derives its basic structure from implementing the diffeomorphism group of spacetime in its algebra. The redundancy in this description leads to a theory defined solely by constraints. The Hamiltonian of such a system does not define a time parameter, and all Dirac observables are constant. Dynamics cannot be expressed as time-evolution and must instead be defined by using relational methods. Relational methods express dynamics as the correlated changes in two or more families of observables.

5 – Extended theories I – Noncommutative geometry – Math

Both physical systems and geometric spaces can be described by algebraic methods. Classical systems and ordinary continuum space can be described by commutative algebras. In physical systems the transition from a commutative algebra to a noncommutative algebra signifies the transition form classical to quantum. For spaces the corresponding transition leads from ordinary differential geometry to noncommutative (differential) geometry (NCG). NCG is a significant generalization of the concept of space and spacetime. Ordinary topological spaces, vector bundles, measure spaces, one-forms, and connections can all be given a commutative algebraic formulation. These spaces and structures can then be generalized to the noncommutative case. We study the constructions of such noncommutative spaces and structures in this chapter. The most important construction we present is the generalization from (commutative) Riemannian manifolds to noncommutative spin manifolds. These new spaces are defined by spectral triples. Tensor products of triples will form the basis of the physical models of chapter 6 and 7.

6 – Extended theories I – Noncommutative geometry – Physics Overview

Canonical spectral triples define classical spin manifolds, while finite spectral triples define zero-dimensional finite spaces. A tensor product of a canonical triple with a finite triple defines an almost commutative space. When used as a model for physics we will refer to such spaces as Connes-Lott-Chamseddine (CLC) models. CLC models allow us to derive GR and spontaneously broken Higgs-Dirac-Yang-Mills theory, with minimal gravity-matter coupling. The algebra of the finite triple determines the details of the resulting gauge theory, and is an essential input to the CLC models. Two set of arguments lead to the prediction of a (practically) unique finite triple for four-dimensional CLC models from first principles. The algebra for this triple is $M_2(\mathbb{H}) \oplus M_4(\mathbb{C})$. This algebra leads to the Pati-Salam model with $SU(2) \oplus SU(2) \oplus SU(4)$ gauge symmetry. It contains the $U(1) \oplus SU(2) \oplus SU(3)$ standard model. CLC models provide a compact geometric explanation of several parts of the standard model.

7 – Extended theories I – Noncommutative geometry – Physics Details

In this chapter the details of the derivations of various Connes-Lott-Chamseddine (CLC) models are described. We construct the CLC versions of QED, the electroweak theory and the standard model. We also briefly review the Pati-Salam model.

8 – Extended theories II – Loop quantum gravity – Foundations and applications

LQG is both a specific quantization of GR, and a general set of methods for quantizing connection based theories in a background free manner. LQG uses a formulation of GR based on connections smeared along one-dimensional curves. The classical theory is defined solely by constraints. Following Dirac's method for quantizing fully constrained systems, we first quantize the unconstrained theory, and then implement the constraints after quantization. LQG leads to a mathematically rigorous picture of quantum geometry, where area and volume operators can be defined and take discrete values. In LQG there is no spacetime at the fundamental level. Spacetime is an emergent phenomenon, most likely appearing as a specific phase in the classical macroscopic limit of the theory. As there is no time variable, the dynamics of the theory must be expressed by relational constructions. Application of LQG to cosmology yields a bounce model where the Big Bang singularity is eliminated.

9 – Extended theories III – String theory

Perturbative string theory can be seen as a generalization from one-dimensional world-line QFT to two-dimensional world-sheet QFT. Except for spacetime, the string is the only fundamental entity in the theory. Different particles with different spins are all realized as states of the string. The string spectrum contains an infinite set of states including a massless graviton spin-2 state. The graviton state gives rise to a UV finite spin-2 theory of gravity. When supersymmetry is added, string theory can be consistently quantized in a 10-dimensional spacetime (subject to some consistency requirements on the spacetime). String theory incorporates both gauge bosons and chiral fermions. This can happen in several different ways. Compactifications and intersecting branes are the two leading approaches for making contact with four-dimensional physics. The dynamics of the theory is defined by on-shell perturbative S-matrix elements. The definition of the "free theory" is sufficient for obtaining this perturbative expansion. The theory is believed to be finite order-by-order to all orders. The AdS/CFT conjecture states that string theory on $AdS_5 \times S^5$ is dual to a four-dimensional CFT. This is known as gauge/gravity duality. Entanglement entropy is a central aspect of how gauge theory states and spacetime geometries are related by such dualities.

10 – Evaluation of Extended theories

In this chapter we evaluate noncommutative geometry, loop quantum gravity and string theory. We evaluate them with respect to the two questions "How can we unify gravity with quantum theory?" and "How can we explain the standard model?". In addition we evaluate their foundational structure.

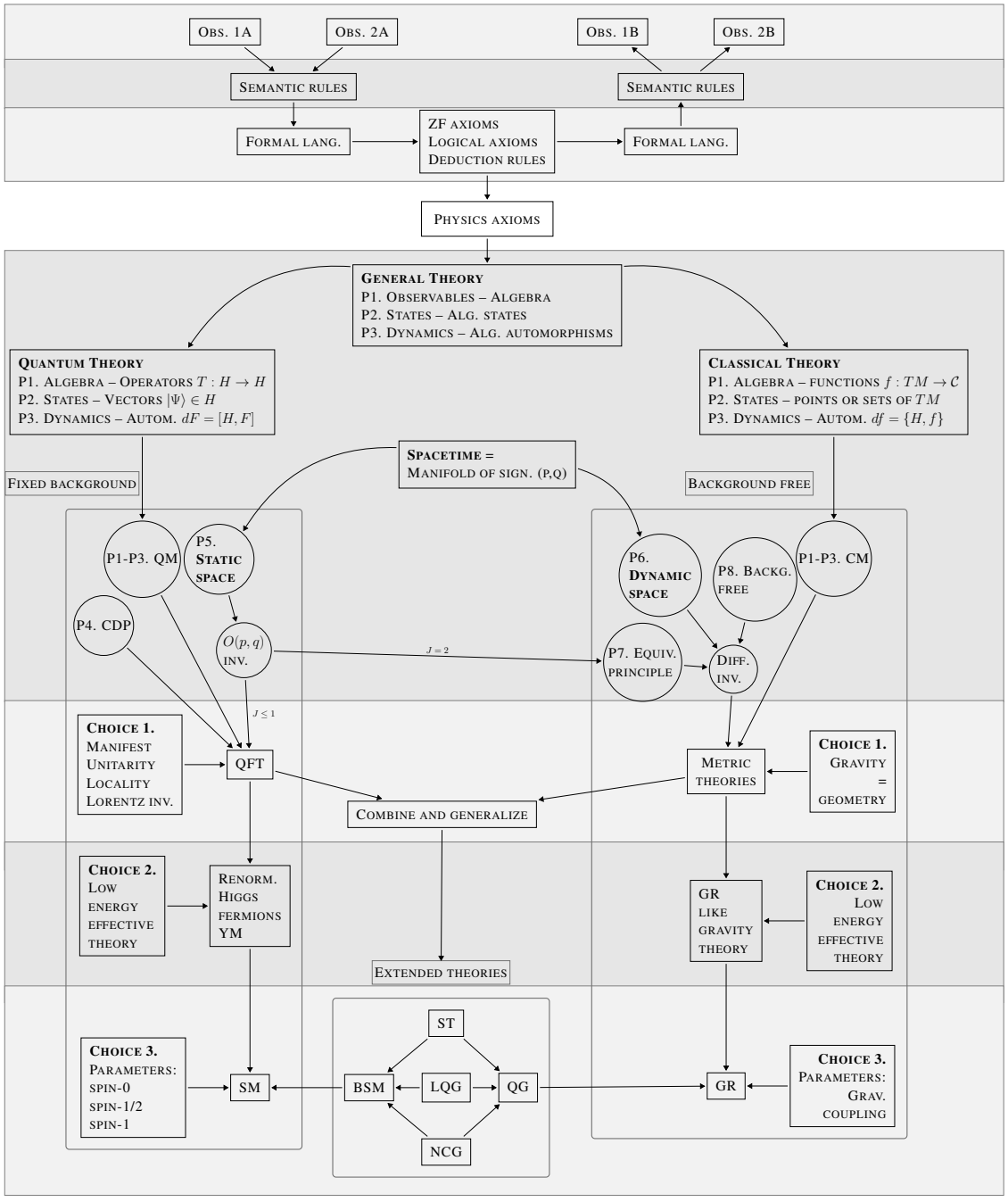


Figure 1: Overview of the text. Abbreviations: P = Principle, Obs. = Observation, Alg. = Algebra, Autom. = Automorphisms, SM= Standard model, YM = Yang-Mills, Grav. = Gravitational, GR = General relativity, TM = Tangent bundle of M, O(p,q) = Orthogonal group, CDP = Cluster decomposition principle, LQG = Loop quantum gravity, NCG = Noncommutative geometry, Diff. = Diffeomorphism, Sign. = Signature, QM = Quantum mechanics, CM = Classical mechanics, ST=String theory, QG = Quantum gravity, BSM = Beyond standard model theories, J represents spin

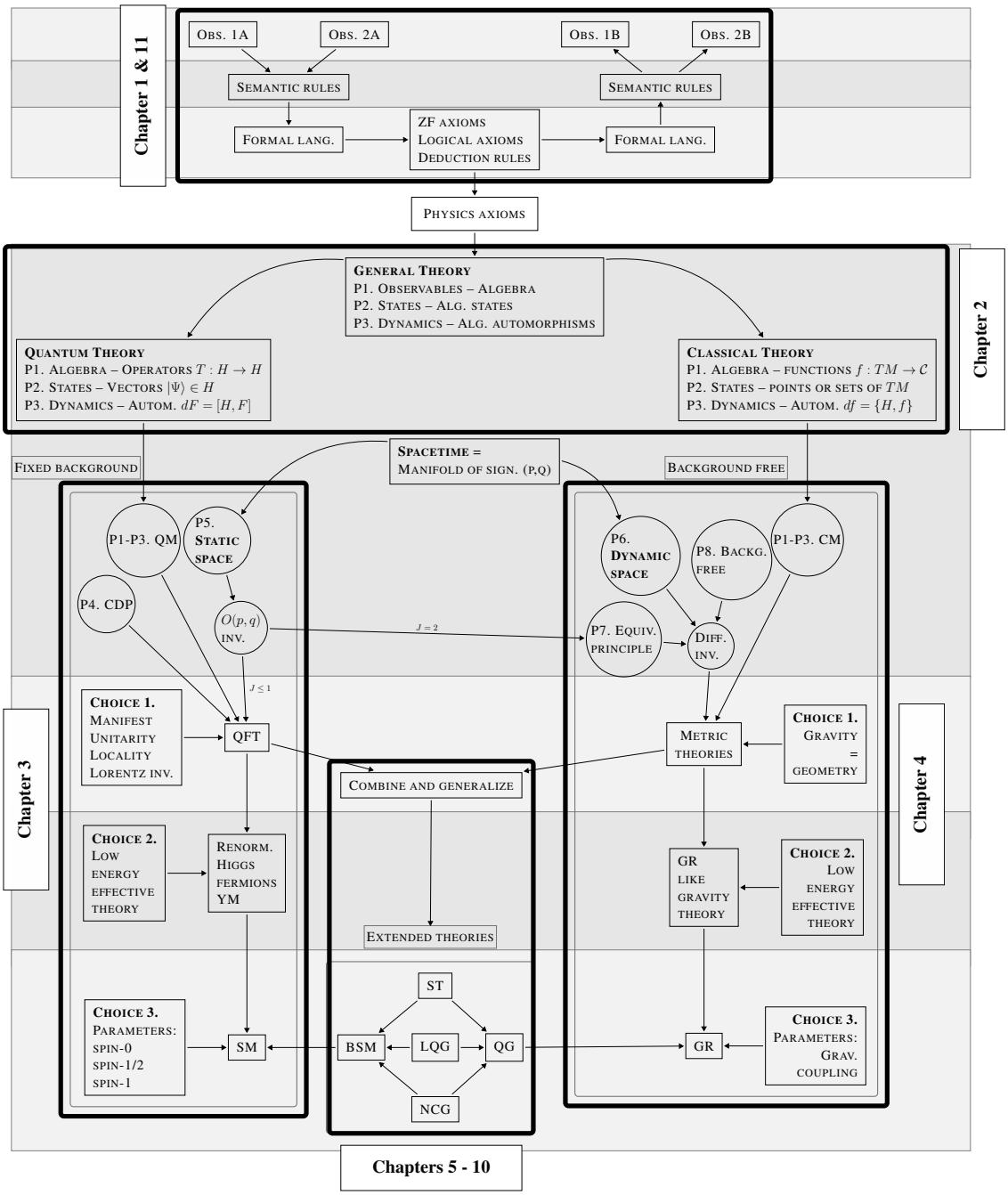


Figure 2: Overview of the text with chapter assignments. See figures 1.1, 2.2, 3.2 and 4.1 for enlarged versions of the content of the chapter boxes.

Doing physics beyond the experimental reach

1. Epistemology

> [in reference to the LHC null result at ICHEP 2016]... you can't trust the judgment of scientists whose future funding depends on their continued optimism.
>
> Hossenfelder

Summary

Physics must be empirically correct but should also be clear, consistent and conceptually satisfying. The number of new decisive experiments beyond current experimental limits will be few. Theory development cannot rely on experiments alone but must instead be driven by focusing on making things clear, consistent and conceptually satisfying. A clear, consistent and satisfying formulation of current physics is likely to be an important key to completing physics in empirically sparse domain of quantum gravity. Rigorous mathematical axiomatizations form a useful parametrization of the theory space in the same way as model building tools help to explore the parameter space of particle physics.

1.1 Introduction

The autumn of 2016 the Large Hadron Collider (LHC) at CERN in Switzerland has just reported that its first year running at 13TeV has confirmed all predictions of the standard model, and found no evidence of new physics (see e.g. [1] or the listing [2]). Placing a bet that LHC will not discover anything more than the Higgs particle, might now get reasonable odds from quite a few theoretical physicists (see e.g. [3]). Some five or ten years ago not many would have predicted the continued uncontested rein of the standard model. Lights are by no means out for new physics at the LHC, but so far there are no hints in this direction.

The possible continued triumph of unmodified standard model physics would be an amazing turn of events. If this were to happen, theoretical physics faces the Herculean challenge of extending the most accurate and comprehensive theory we have ever had, and furthermore to do so without any major input from experiment. In fact, independent of any results at the LHC scale, many believe that the most pressing challenge beyond the standard model lies within quantum gravity. Quantum gravity is the attempt to produce a coherent whole of the two current pillars of physics, the standard model (SM) and general relativity (GR). Quantum gravity is not easy to approach by experimental means. The essential secrets of quantum gravity are quite possibly located at the Planck scale, some 16 orders of magnitude beyond our current experimental reach.

Faced with such a challenge, which is brought on by its own success, it is no wonder that theoretical physics has become a bewildering place to set one's feet. Bold speculations and hair raising ideas abound. Discussions suggesting the existence of several extra invisible dimensions, and of hundreds of unknown new particles are just the tiny top of the iceberg. Despite suggestions to the contrary, this is all natural and necessary, and not a sign that physicist have gone of the deep end. Nevertheless, the question of how to move forward in such a situation is not an easy one. It is not obvious how one can define and evaluate if physics is making progress. What questions could we possibly ask of extensions of a theory that is so uncompromisingly successful? What methods should we use to find such extensions, and how should we check the answers that we come up with?

What ever the right direction for theory to move is, we can surely agree that we must foray into the unknown with new tools. Like every other vibrant subject, physics needs to continually rebuild itself. When rebuilding a subject it is important to include all the sound building blocks that are already in place. To achieve this we will analyze physics in a manner somewhat analogous to Descartes analysis of reality, in search of "cogito statements" for physics. In section 1.2 we describe the general business order of physics, and analyze how this dictates the structure of physics. In section **??** we look at the specifics of the current situation and suggest how the methods of theoretical research can adapt to these conditions.

1.2 The general epistemology of physics

> If, as one believes, all mathematics reduces to the mathematics of logic, and all physics reduces to mathematics, what alternative is there but for all physics to reduce to the mathematics of logic? Logic is the only branch of mathematics that can "think about itself."
>
> Misner et al.

In this section we attempt to analyze the knowledge structure of physics. We try to place observation, prediction, deduction in their proper role within the system. While upcoming sections advocate some changes to physics methodology, this section is almost anti-revolutionary in its content. The section is meant to substantiate the following claims: It is *not* surprising that the world is partially predictable, it is in fact essential for any sort of order and life in the universe. It is *not* mysterious that mathematics does so well in describing our world. In fact *any* useful description must involve structures containing a formal language and deduction rules, that are of the same character as those found in mathematics. In addition to the purely formal mathematical structures we need to give meaning to the formal expressions. Semantic rules are an essential and irreducible part of physics, since there are no a priori semantic connections between mathematics and observations. Building on these assertions we will conclude that physics, as we define it here, simply must be a mathematical system with a set of axioms and deductive rules that enables us to make predictions, and a suitable set of semantic rules to allow correspondences to be made between formal language statements and empirical observations.

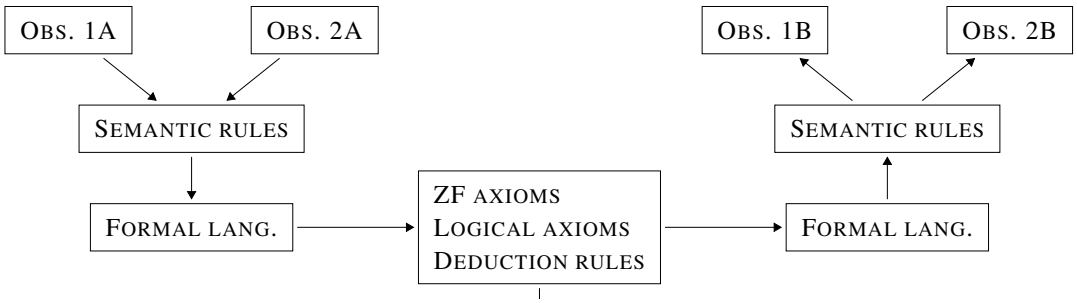


Figure 1.1: The basic epistemological structure of physics. It uses two or more correlated input observations 1A and 1B, and two or more correlated output observations 2A and 2B. This is meant to suggest that fundamental physical observations are observations of two (or more) correlated properties. Time is not a necessary part of this basic empirical setup. Semantic rules translate these correlated observations into formal language statements. The deductive apparatus of the formal system along with relevant axioms transform the formal language encoded initial observation into a formal language encoded prediction. The formal language encoded prediction is then transformed by (reverse) semantic rules into a set of correlated observations. Abbreviations: see figure 1.

Let us now go through the above claims and try to substantiate them. We start by formulating the task that physics is based on.

The primary task of physics is using observations of the present to make predictions of the future. However, we want physics to do more. We want to make these predictions in a manner that is independent of any one person's special abilities and intuitions. We want physics not just to make any kind of predictions, but make predictions that are intersubjective, repeatable and improvable. Having establish what physics must do, we can now move on to discuss *how* to perform such a task.

A central clue to meeting these demands is the implementation of the demand of intersubjectivity. To make our observation and predictions open to reproduction and improvement, we need to transfer them out of our individual experiences. This process involves three different stages (see figure 1.1). The first stage involves encoding our sensory experiences in a (written) language. The second stage involves expressing the knowledge behind our predictions in a set of language based prediction rules. The third stage is the reverse of the first stage, translating language expression back into sensory experiences.

The first stage could simply be called "writing down what we see". In this first stage we must transfer sense experience onto a format that is objectively available. Creating written language expression, that represent our sensory experiences, is the only known method to accomplish this. Note that it is not sufficient to simply store the experiences in the form of drawings, photos and videos, we need a format that is enables us to use formal deduction methods to be able to proceed. To be useful for physics the sensory data that we encode in formal language must be in the form of correlated observations. We cannot just encode "I see a ball", we must encode "I see a ball in front of the window correlated with the sun being directly above the church spire." We demand that useful observation must have a certain specificity with regards to being correlated with other observations. Our everyday experience suggest that we need four correlation "dimensions". We encode these correlation in the form of tuples. We do not make any assumptions of whether this constitutes localization in some hypothetical time and space continuum.

The second step concerns predicting future phenomena based on the input from step one. To achieve the goal of predicting phenomena, we need rules to allow us to transform the encoded initial observations to a set of encoded predictions. This step manipulates the initial formal language expression, and produces a new formal language expression. With this new expression in hand, we are ready for the third step.

The third step is the reverse of the first step, it transforms these encoded predictions back into (descriptions of) sensory experiences. This final step allows us to evaluate our prediction steps and if necessary to improve on them. To further explore the details of these three steps we continue to focus on what is necessary to achieve intersubjectivity.

As humans what we can reliably and strictly agree on is quite a limited set of experiences. We seem to at least be able to agree on whether or not two symbolic expression are identical, and we can also agree on the results of well defined formal manipulations of such symbols. Or in a more useful formulation, we can agree on whether two strings of letters are identical, and which strings results from applying some strictly defined transformations to a given string of letters. To achieve intersubjective, repeatable and improvable prediction the use of such symbols and such strictly defined transformations is essential.

Symbols alone would not be of much use unless we can relate sequences of symbols to our experiences. Indeed, in most cases the use of symbols is just an extension of the basic idea behind a spoken language. In a spoken language we make an arbitrary correspondence between a given sequence of sounds and a different unrelated sensory experiences. We must therefore assume that we can setup a collection of semantic rules to allow ourselves to describe parts of human experience with strings of letters in a well defined manner. As in the spoken language, these rules will be arbitrary, but once defined will represent well defined rules of meaning. We decide that say, seeing a ball lying on the ground, will be encoded by some string of letters.

The three steps above is a general description of central pieces of a formal system, extended with a set of semantic rules for translating to and from experiences. If we agree that the business of physics is inter-subjective, repeatable, improvable predictions, we can conclude that this means using a formal system with some semantic rules. The best known and most useful example of a formal system is mathematics. Mathematics is a quite general formal system, and many of its constructions are specifically tailored to encode some aspect of the human experience. Natural numbers were based on our experience with counting, and its algebraic rules were closely modeled on what we observed when counting things in nature. The final construction of the natural numbers is of course quite a substantial generalization of its more humble empirical origin. In the same manner a function could be said to encode the basic idea of correlated observations of two experiences like the position of the sun and the sea level height in a port. In the end it is not important whether all mathematical structures can be related to natural experiences, it is enough to know that many of them are so inspired. In principle mathematics can be used for pure mathematics without introducing any semantic rules, but for physics purposes we need such rules. A simple example of a rule of meaning is that the formula $x^2 + y^2 = 1$ in mathematics could correspond to visual experience of seeing a circle. In general such rules seem to be possible and we shall not discuss problems of semantics any further.

Transformation rules in a formal system are often called deduction rules. There could potentially be more than one possible set of deduction rules. We could have considered the deduction rules of the formal system as a separate question. This may not be strictly necessary since we can express system with alternate deduction rules within mathematics. Quantum logic and topos theory are examples of systems (or subsystems), modeled within mathematics, which employ different deduction rules [4][5]. We note that in deductive part found in step 2, the deductive framework is not primarily changed by changing the rules of deduction, but instead by changing the set of axioms that are used. These axioms encode our knowledge of how the world works. We shall assume that what we have described is a necessary foundation for doing physics. Before going on to other matters we include a section with a more detailed discussion of the nature of observations and their relationship with predictions.

Observations and predictions

We have argued that the empirical part of physics consists of writing down correlated observations, usually in the form of number tuples. One example could be registering the number of pendulum swings, and corresponding grid positions of a moving particle. We as humans are confined to making a quite small number of localized observations of correlated phenomena. Based on these we try to predict a much bigger set of potential measurements, and then compare some of those predictions with what we actually observe. A major point in all this, is not just the sequence of steps involved in this process, but also the fact that the knowledge that goes into the predictions must be "compact" in the sense of allowing us to know and predict "more" than we measure. From just observing two points on a particle trajectory we want to predict the complete infinite set of points on the trajectory. The prediction of infinitely many potential measurements using a concise set of prediction rules is part of the essence of physics. Just setting up large tables of mindless correlations would not constitute physics in the sense that we desire. Sets of such tuples produced by observations will have the properties of (isolated parts of) functions (potentially multi-valued). This means that a natural way to formulate predictions will be in the form of (complete) functions. This means that we must find mathematical structures that can determine whole functions based on only a small and finite set of points. In general any kind of equations that determine functions from a few select points would be suitable. Differential equations are certainly one primary example of this. This suggests that differential equations (or difference equation or potentially integral equations) might play a central role in any description of nature. We make no assumptions on what total set of possible tools exists (or can be made) to support this kind of information decompression. We simply state that the relevant tools must allow this decompression from discrete and finite data to something representing much more information.

Data types and probabilities

Our raw set of actual measurements could in principle be of several types. We could use ideas from Stevens classification into nominal, ordinal, interval and ratio data [6]. We shall not go into such generalities, but instead assume that we are using tuples consisting of elements from some of the common number systems. We assume that this will be a general enough tool to capture any physical information. This broad assumption does not mean that which number system we choose is a trivial or irrelevant choice. Describing the position of a one-dimensional system as an element of $\mathbb{N}$ versus using $\mathbb{R}$ does make quite a difference. In this regard we should not assume that even though $\mathbb{R}$ is a a very useful number system for tuples describing low-energy physical systems, that this is an appropriate choice at all energy scales [4].

Finally, we need to add one minor but very important extension to our system. There are many indications that we need to use probabilities to properly describe some of our observations (see any textbook on probability or quantum physics). By probabilities we shall either mean a measurement of relative frequency of outcomes, or a measurement of averages, all from some assumed similar preparation. These type of measurement are in principle not much different from regular measurements. We just need to register a reasonably large set of occurrences, and their corresponding frequency (or averages). (We are likely to need some concept of identically prepared systems or something similar.) Thus, when including probabilities, our fundamental model will be that observations are finite set of tuples of correlated averages from identically prepared systems. Notice that this does not refer to the existence of any special time parameter, they are simply correlated measurements of different observables. Nor does it assume any sharply defined (i.e. not based on averaging) values of observables. This framework is a priori timeless (or at least agnostic about time) and based on averages (or relative frequencies) of multiple observations.

Physics reasoning

We will now say some words about reasoning in physics and related subjects. This is an important area to clarify since, as we have argued above, we believe it to be a crucial feature of the methods employed in physics. Mathematics, physics and logic have not always seen eye-to-eye as to what constitutes proper deductive reasoning. Through various historical periods the viewpoints and insights voiced by practitioners in these subject areas have differed with regards to such matters. We consider the relation between mathematics and logic first. We claim that the differences between mathematics and (formal) logic, which is claimed to exists by some mathematicians, are actually just superficial differences [7]. For instance, mathematicians currently still use indications of proofs as proofs, rather than actual proofs. This is highly efficient, and reasonably resistant to errors in cases where proofs are of moderate complexity. However, the checking of complex proofs (or proof indications) is very time consuming and demanding, and complex proofs require years of collective effort before they are accepted as correct. Very complex proofs are virtually impossible to check manually, and alternative methods must eventually emerge. We suggest that the necessary techniques are those found in formal logic proofs. This is essentially nothing new, it just represents a formalization of the informal mathematical proofs that are used today.

It is actually not too difficult to transfer the process of proof checking to the formal domain, and at the same time implement automated proof verification by computers. This process is currently progressing somewhat slowly because of practical issues like disagreements on the proper choice of representation language and so on. In the near future it seems likely that all mathematical theorems and definition will be encoded in computer databases, and new results will require a computer verified proof before being admitted in the database (see [8] and other papers found in[9] for an updated perspective). Journals will likely only publish results with computer verified proofs, and referees can focus on evaluating the significance and presentation of the results, before allowing publication. Even though mathematicians are likely to still use informal proofs in various phases of their work, as well as in teaching, we consider that such a transition to formal proofs would be a definitive demonstration that mathematical reasoning is correctly regarded as a branch of formal logic. This leaves the question of physics reasoning versus mathematics reasoning.

In a quick sidebar, we note that our discussion is agnostic with regards to any Platonic ontological existence claims of mathematical objects. The strictly formalist position advocated above, while possibly compatible with Platonic ideas, does not lend any support to such ideas. Nor does our setup for physics endow any ontological equivalence between "the real world" and mathematics. As argued before, mathematics embodies, through being a suitable general purpose formal system, the structures needed for any meaningful physics, and its success is simply a reflection of the fact that physics is possible at all, not evidence of any mysterious force within mathematics.

Now on to reasoning within physics. The goal of physics (in our view) is to achieve a mathematical description of nature, and thus when completed all (correct) physics arguments will be isomorphic to an argument in mathematics. There will be no (essential) difference between physics reasoning and mathematics reasoning. At the moment this complete description is not available, and there will in many cases be a separate form of physics reasoning distinct from mathematical reasoning. Such reasoning will use a mixture of ideas partly from the purely mathematical model, and partly from the semi-mathematical intuitive understanding that a physicist might have about a physical system. This kind of reasoning is probably essential for progress in many cases. The challenge that this kind of reasoning presents, is that it is not purely logical. A physicist might for instance engage in reasoning related to the sum (a concept which assumes convergence) of an infinite series which is known to be non-convergent, this introduces statements that from a logical perspective are contradictory. In the strict logical sense one could, from these contradictions, deduce any possible mathematical statement (see Note on completeness 1.4). This means that any such physical reasoning must at all later steps be limited by some sense of "being reasonable". This

again means that the possible deductions from a statement are a combination of logic, restricted by the physicist's intuition about what must be considered "unacceptable" inferences.

Login that is "sensibly-limited" is also used in mathematics. Many mathematics books profess to only using so called naive set theory. As is well known, one can easily derive contradictions (like Russel's paradox) in naive set theory (see [10]), and again from this contradiction one can derive any desired statement. In the strict logical sense this is of course methodological suicide, but when supplied with "agreed upon" limitations, it is a functioning system.

A system using "sensibly-limited" logic can be quite powerful within some domains. These domains are typically domains that we already know very well, and those we have useful intuitions about. The domain of theoretical physics beyond the current experimental limit, is an area where reliable intuition is hard to build, mainly as a result of the absence of data. This makes "sensibly-limited" logic a less powerful tool. Since we suggest that an important approach for the developments ahead will be analysis of the deep structure of physics, a more hard-line view on the role of formal reasoning might be required. For instance, declaring a set of axioms to be a new proper set of axioms for quantum theory, can only realistically be verified within a quite strict mathematical framework. Such axioms will be of little value if they are only vaguely linked to the already established quantum framework. The situation is not black-or-white, as many semi-rigorous arguments do capture the essence of the rigorous argument, but it suggests that a shift towards more rigorous methods in theoretical physics might be indicated.

1.3 The current epistemological challenges

> Given the lack of experimental evidence that would point us towards fundamentally new principles, we should do our best to understand the full implications of the principles we already have!
>
> Baez

Can we get new data?

Physics is, and must remain, empirically founded. The question is not if a new theory must be tested by how well it agrees with observations, it must be. The question is whether we can expect a continuous stream of new empirical data sufficient to *drive* physics beyond its current limits. Making new observation often requires building new instruments that extend the available energies or sensitivities. We need telescopes that can see further than before, and microscopes that can detect smaller things than before. The best microscopes are high-energy colliders. They are built using acceleration by radio-frequency (RF) techniques, and bending by superconducting magnets. Both technologies are believed to be close to their maximum potential. Bigger colliders can still be built, but will require building larger and larger tunnels. This involves increasingly larger sums of money, that must be put together by big international cooperative efforts. The combined time for preparing a collaboration, securing funding, agreeing on a design, building and comissioning can easily reach 30 years or more. It is indeed a question of whether there will ever be built a collider beyond LHC with the current type of collider technologies. New collider techniques like plasma acceleration will surely come along, but the timeframe of building a collider substantially superseding LHC energies is still 30+ years. This suggest that we must be cautious in hoping for a strictly data driven progress in this domain.

On the other hand, there are some possibilities for non-collider based measurements. Proton decay searches in the upcoming Mega-Kamiokande is one example of a measurement that reaches far beyond the LHC energies. LIGO and other gravitational wave experiments, as well as BICEP

and ICECUBE are other examples of instruments that probe new domains. Hopefully, many more ingenious techniques will be invented to investigate the unknown reaches of the physical world. Indeed, all these comments are not to suggest a negative outlook on future empirical results, just that the speed of obtaining such data probably will be too slow to be the driving force behind creating theories beyond the standard model.

What to do without new data

> ...if some of the constants of nature, like the masses of the quarks and leptons, are environmental, it means that these are much more loosely connected to the fundamental mathematics that underlies physics. There is an essential randomness to them, not unlike the weather. This is not 'normal physics,' and it is not what anyone wants, but unfortunately there is significant evidence that this is the case.
>
> Polchinski

We now move on to consider specific additional challenges the present situation in physics entails. We have already stated that the major aspects of the current situation are that we have very successful theories, and that new experiments, even just slightly beyond the current reach, will be very expensive and far into the future, and finally that the major challenges are believed to lie in a domain, not just slightly beyond, but far outside current experimental reach.

One can attack this problem in the same manner as the historical record of the last 100 years suggests. We can postulate some new gauge group, add new particles, and derive new experimental consequences to be found at some higher energy collider to be built in the year 20XX. But, even apart from the time frame, there are subtle signs that this should not be the major avenue pursued. Most importantly because, even if such an extension should turn out to be correct, it is unlikely that such an extension alone will lead us any closer to a solution of the more fundamental questions that many believe physics should now turn towards. Physics is now at its highest level of phenomenological success ever. Yet, almost five decades has passed since the completion of the standard model, without any increase in the *established* fundamental knowledge of physics.

We should certainly build all the super high-powered accelerators we can afford, but we suggest that progress toward the ultimate domains of theoretical physics will most likely come by achieving a deeper understanding of current physics. Formulating a satisfying and clear account of quantum mechanics, the process of change, GR and the standard model could open up a clearer path to how these theories can be generalized. What are the meaningful generalizations of current theories? This question might have a very limited set of answers. Secondly, being able to reproduce current physics puts enormous constraints on any extended theory, and this will also be essential to determine whether any given extended theory is the correct one. If we are to stay within at least a semi-empirical domain of theory testing, the extended theory must shine a substantial amount of light on current physics. The environmental ideas of the multiverse paradigm, described in later paragraphs of this section, could put a serious and permanent dent in this last ambition.

The ultimate question, of the sort that looks beyond phenomenology, would be to ask why physical theories looks the way they do. We do not quite know what kind of an answer can be imagined for such a question. It could be some sort of self-explanatory theory, that not just explains current physics, but also explains why it is the only solution that can be applicable. It would need to somehow break the cycle of being an explanation that demands another explanation, that demands one more and so on. To give an example of this kind of question, we could ask: Why is the world quantum mechanical? By this we mean to not just state that it is quantum mechanical, and state the rules of the quantum game, but explain why these *must* be the rules of the game.

A question slightly below this ultimate level concerns the origin of the so called free parameters of physical theories. If we accept the premise that the world is well described by gauge theory

and GR, there are still very many free parameters that need to be chosen. The number of such parameters is much greater than the commonly listed free parameters describing masses, mixings and couplings (see note 1.4). Those parameter-lists only appear in their "traditional" form after we have made quite a few significant choices. How should we for instance count the number of parameters to describe which gauge groups we should use? There is an infinite number of possible "single" gauge groups, and as we are allowed to combine several groups, they also come in an equally infinite set of possible combinations. The same would apply to choosing what charges or representations we assign to fermions.

There are two diametrically opposite approaches to this parameter problem. We can refer to them as the “multiverse” solution and the “unique world” solution. The multiverse solution posits that the explanation of the free parameters is similar to our explanation of the various solar-planetary distances of our solar systems [11]. These distances are considered to be "environmentally" determined, that is they are considered to be random picks from a set of billions of billions of similar systems with different numerical values for these distances. The "unique world" solution on the other hand, claims that the actual parameter values of a theory like the standard model are actually unique, or at least a selection from a very limited set.

Some physicists feel that the multiverse solution is an unscientific explanation since it might be impossible (for localized observers) to test it. It is not hard to empathize with this viewpoint, but we are not sure this is a correct assessment, regardless of whether we can find a practical way to test the theory or not. An analogy with the historical case of Kepler might be informative in this respect. In his time, Kepler tried very hard to explain the solar-planetary distances of our solar systems by a "unique-world" type of explanation. Most people would say Kepler’s efforts were misguided, and that a "multiverse" type of explanation is appropriate for this problem. We can use this example to put a different perspective on the observable and not observable distinction by setting up a thought experiment. For the Kepler problem we can easily imagine a situation where our solar system looks just as it does today, and that like today, it is not one of kind, but actually one of a billion-billion other solar systems, but that the expansion of the universe had (already) driven all other solar systems out of our empirical reach. (This scenario would entail placing our solar system as a lone island outside of bound systems like galaxies that are not pulled apart by cosmic expansion.) We suggest that most physicists would agree that the "multiverse"-type explanation would still be the correct one, even though we would probably never be able to directly confirm it.

Summing up, we most likely will not be able to just "scatter" our way to a full theory of nature. Therefore we should also put a lot of effort into forging theories with explanatory clarity, consistency and "satisfactoriness", as well as being compatible with the extensive set of facts we already know. This "behind the scenes" approach provides some extra tools that might be powerful enough to nudge our understanding the final step of the way.

1.4 Notes

Completness

In any complete logical system it is true that any contradiction leads to being able to deduce any statement. In an incomplete systems all non-independent propositions follows from a contradiction. Propositional logic and (pure) first-order logic are complete systems. However Peano arithmetic, Zermelo-Fraenkel (ZF) set theory are not complete systems. As Gödel uncovered these systems have an infinite set of independent statements. Independent means that neither the statement nor its negation are provable from the axioms. The axiom of choice and the continuum hypotheses are example of propositions that are independent of the ZF axioms. Some express Gödel by saying that there are true statements (in mathematics) which are not provable. However this statement depends on a definition of (mathematical) truth beyond the formal logical one, and it is not universally

accepted that such a definition of truth has any meaning.

Counting of parameters

The notion of how to count (free) parameters of a QFT like the standard model is involved. Certainly we have the well known parameters which figure on all such listings and discussions. Frequently the question of the gauge groups and representations are disregarded. Is the choice of gauge groups a discrete parameter? In some sense it is, but as long as models are based on empirical input we can really only conclude that unseen gauge groups must have smaller couplings than we can currently detect, or couplings that only show up at higher energies. This makes the choice of gauge groups into an infinite dimensional problem. This applies almost equally with respect to fermion representations. Except for anomaly cancellation there are no a priori restrictions on these representations.

II Foundations of established theories

2. Physical theories in general

> Some principle uniquely right and uniquely simple must, when one knows it, be also so compelling that it is clear the universe is built, and must be built, in such and such a way, and that it could not possibly be otherwise.
>
> Misner et al.

Summary

Classical and quantum systems can be described by a common three-part framework of observables, states and dynamics. In mathematical terms this consists of a *-algebra (representing observables), a set of linear algebra functionals (representing physical states), and a one-parameter family of algebra automorphisms (representing time-evolution or dynamics). Classical and quantum systems are distinguished solely by the commutativity or noncommutativity of the algebra of observables. Both quantum and classical theories can be seen as probability theories with the added structure of algebra automorphisms to represent dynamics. The fundamental use of probabilities, and the reliance on formulas like the Born rule, are common to both approaches. Incompatible observables, the uncertainty principle and entanglement are quantum properties that only occur when using a noncommutative algebra of observables.

2.1 Introduction

In all of physics there is a common three-part structure of observables, states and dynamical equations [12][13][14]. In this chapter we argue that these three concepts can be described for any system by giving the observables in the form of an algebra, the states in the form of a set of linear functionals on the algebra, and the dynamical equations by a one-parameter set of automorphisms

of the algebra. We will see that these structures are equivalent to a generalized probability theory with dynamics [13]. We therefore argue that probability theory is a suitable general framework to formulate physical theories. This involves seeing the traditional sharp states of classical mechanics as limiting cases of probability distributions. In short we argue for the propositions that classical mechanics equals commutative probability theory, and quantum mechanics equals noncommutative probability theory. This represents a unified, compact, and in our view also very useful, perspective on these theories. A key step to be able to build this common physics framework is to re-express probability theory using an algebraic formulation of measure theory. This algebraic probability theory is subsequently generalized to a noncommutative version [15][16]. Rules such as the Born rule of probabilities applies equally to both the classical and quantum case.

As a side-note to later chapters we mention that the topics of this chapter form some of the historically earliest appearances of what is now known as noncommutative geometry [16]. Indeed these structures where used by Heisenberg, von Neumann, Born and Dirac before the word noncommutative measure theory and noncommutative geometry were even invented. Seen from a historical perspective, noncommutative measure theory is the paradigmatic example of noncommutative geometry [17]. In this chapter we treat noncommutative measure theory as a distinguished part of noncommutative geometry because of its essential role in expressing the foundations of quantum mechanics.

The essential difference between the classical and quantum systems lies in the commutativity or noncommutativity of the algebra. Except for this we approach the classical and quantum system in the exact same manner. To make both the unity of our approach and the decisive difference as clear as possible, we eliminate all unnecessary extra structures. This means that we eliminate references to things like phase spaces, Hilbert spaces, operators, and eigenvalues in the fundamental definition of the systems. This is a substantial improvement in terms of conceptual clarity. At later stages it will turn out that the more traditional formulations of physics emerges naturally from the abstract formulation without requiring any additional assumptions. From our commutative algebraic description of classical systems we can deduce a description based on functions on phase space. Likewise, in a noncommutative setting, a realization using operators on a Hilbert space appears natural (see chapter 5 for more details on these two statements). This also represents a conceptual improvement as the commonly used formulation appear as natural consequences of a more unified and compact framework.

In this chapter we consider classical and quantum mechanics in the meaning of abstract general systems. We will pay little attention to how to actually determine which specific algebra corresponds to any specific physical system, and therefore also talk little about the process of quantization, i.e. how we can find the relevant quantum algebra from a given classical algebra. We will treat quantization in chapter 8.

2.2 Probability spaces

> ...the fact that we do not really understand, in a coherent and conceptual way, what that most successful theory of physics called "quantum mechanics" tells us about Nature, represents an intellectual scandal.
>
> Frolich and Schubnel

Probability is an essential ingredient to be able to formulate quantum mechanics, while classical mechanics is most often presented without any probabilistic ingredients. However, in classical *statistical* mechanics the probability aspect is central (see e.g. [18]). This distinction stems from the assumption in (non-statistical) classical mechanics that one can in principle determine the state with infinite precision, and that therefore no probabilities are needed when specifying a state. We note in

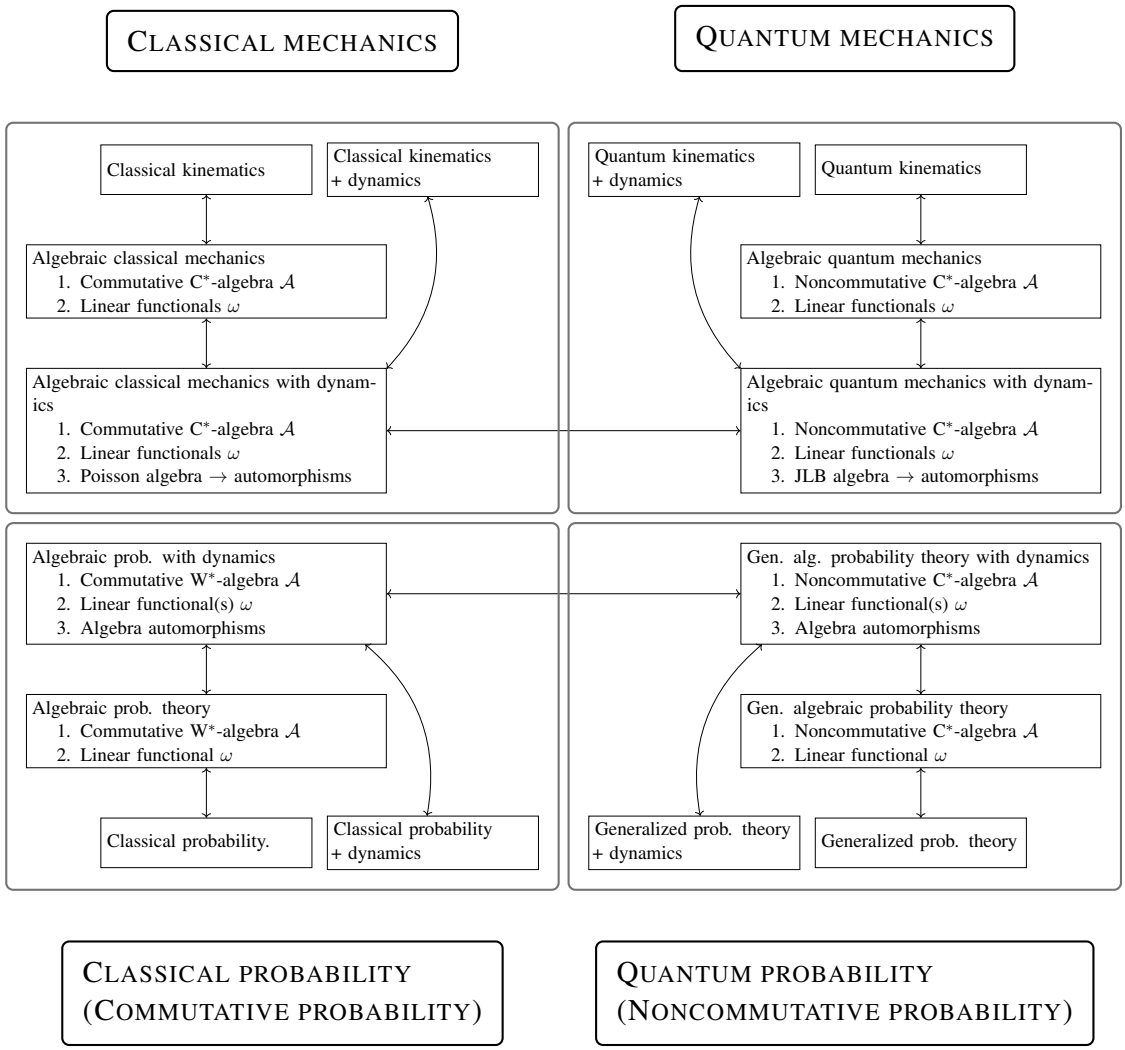


Figure 2.1: The diagram shows how the four theories of classical mechanics, quantum mechanics, classical probability and quantum probability are small (but highly significant) variations of the same structure. The two-part structures represent pure probability theories or the kinematical aspects of physical theories. When we add dynamics to this two-part structure, we get the three-part structure which represent probability theories with dynamics, or the kinematical and dynamical part of physical theories.

opposition to this that even when doing regular point mechanics, without the intention of doing statistical physics, the assumption of being able to classify a state with infinite precision is never realized. Instead it is therefore more realistic to suggest that any state of classical mechanics only determines a probability distribution on the set of point states. We therefore decide to regard the single point states as highly idealized limiting cases of the formalism used in statistical mechanics, and choose to see the statistical formulation as the more fundamental. This makes the connection with quantum theory much more transparent. This is not to be regarded as an ad hoc trick to achieve this goal, but as a suggestion that this formulation should be the natural starting point for any formulation of mechanics. It is the use of point states that should be considered to be a trick to achieve a particularly simple description. To the extent that one needs the idealized point states these should be derived as limiting states. To see how probability theory fits together with mechanics, we start by reviewing classical probability theory in a form suitable to our purposes.

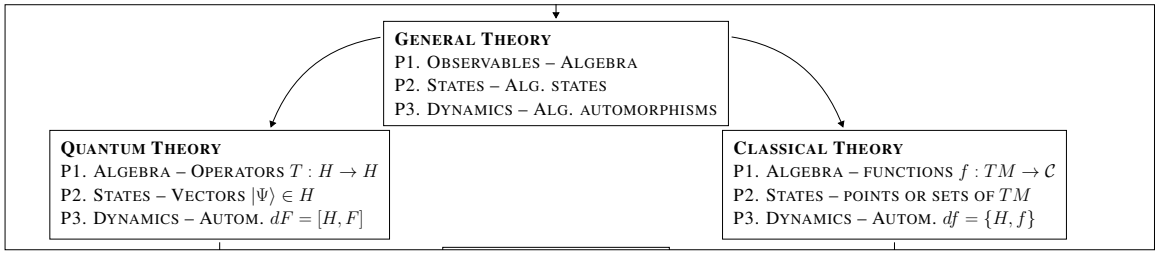


Figure 2.2: The basic structure of physical systems is the three-part structure of observables, states and dynamical equations. This can be represented by an algebra, algebra states and a one-parameter group of algebra automorphisms. This applies for both classical and quantum systems. Abbreviations: see figure 1.

Commutative probability spaces

Hilbert suggested that one should find an axiomatic basis for probability [19], and Kolmogorov did so in 1933 [20]. The classical Kolmogorov axioms for probability is captured in the idea of a probability space (which is simply a measure space of measure 1) [21]. A measure space (probability space) (X, σ, μ) is defined by a set of subsets of X, the σ-algebra σ, and the measure μ, which is a positive real valued function on this σ-algebra (see chapter D for a fuller description of probability theory). The set X is a member of σ so it is implicitly defined by σ, and can be left out of the definition and we are left with the pair (σ, μ).

The total set X represents elementary event, like the six possible outcomes when rolling a dice. The σ-algebra σ represents events in the sense of subsets of possible outcomes. One such event could be defined by the statement "rolling a dice and getting an even number of dots". This would equivalent to the subset $2, 4, 6$. The positive real valued measure μ gives a probability between 0 and 1 for each event, for instance for a model of a fair dice we would have

$$\mu\left([2, 4, 6]\right) = 1/2. \tag{2.2.1}$$

A probability spaces was defined above through subsets and a specific function on these subsets. We shall soon see that we can also define a probability space using algebraic methods. These two very different but equivalent definitions of the same object is a type of correspondence which we will see many instances of.

We digress slightly to note some general aspects of such a correspondence. We will typically have on the one side a space, defined in terms of a set of points with some additional requirements on it, and usually also some extra structures defined on it. On the other side of the correspondence, we will have an algebra satisfying some algebraic properties, and also in this case possibly with some other structures defined on it. We will then try to establish that there is a complete correspondence between theses two different descriptions. We describe this by saying that there is a functor between the category of spaces of the relevant form on the one hand, and the category of algebras of the relevant form on the other hand (see chapter B for an introduction on categories and functors). To complete the mapping between the descriptions we also need another functor going in the opposite direction.

The algebra on the one side of such a functor will often be given by a set of functions on the space defining the other side. So to achieve the algebraic version of probability theory let us then first find the relevant set of functions on set of possible outcomes X. A function $f\colon X \to Y$ from one measure space (X, Ω_X, σ_X) to another (Y, Ω_Y, σ_Y) is called a measurable function if the inverse image of a measurable set of Y is a measurable set of X. The set of measurable functions (also known as random variables) with values in $\mathbb{R}$ or $\mathbb{C}$, with respect to the Borel measure on $\mathbb{R}$ or $\mathbb{C}$, form a closed algebra of functions under the usual addition, multiplication and scalar multiplication in $\mathbb{R}$ (or $\mathbb{C}$). We can define a subset of this algebra, which is more suitable for our

purposes, by limiting the elements to be essentially bounded measurable functions (see appendix D), which we write as

$$\mathcal{A} = L^{\infty}(X, \sigma, \mu). \tag{2.2.2}$$

Integrating such measurable functions with respect to the given measure defines a linear functional ω on the algebra, and we write

$$\omega(f) = \int f d\mu,\ f \in \mathcal{A}. \tag{2.2.3}$$

Since we have used properties given directly by the definitions of the probability space to define the algebra and the functionals, we infer that the probability space determines this algebra and the functional uniquely. This then defines a functor from the category of measure spaces to the category of "algebras with a functional" (see definition 2.2.3). To be precise, to actually show that this is actually a functor we need to also show that morphism and mapped to morphism. We skip this step to keep the discussion brief (see [22]).

Interestingly this also works in the opposite direction, and we now start to look for a functor in the reverse direction. The first step is to have a precise description of category that is the target of the functor we already described. We state the relevant information in a theorem.

Theorem 2.2.1 Let (Ω, Σ, μ) be a probability space, then $\mathcal{A} \coloneqq L^{\infty}(\Omega)$ is a commutative von Neumann algebra, and $\omega \colon f \mapsto \int f d\mu$ is a faithful normal state on $\mathcal{A}$.

Proof. See [16, p5] for a proof. ■

A functor can now be made that, from a given algebra and a functional, is able to construct a probability space (up to sets of measure 0). We label the measure space that result from identifying sets that differ by sets of measure 0 by $(X, \tilde{\sigma}, \tilde{\mu})$. We can reconstruct this (slightly modified) probability space $(X, \tilde{\sigma}, \tilde{\mu})$ from the algebraic definition as stated by the following theorem.

Theorem 2.2.2 If $\mathcal{A}$ is a commutative von Neumann algebra, and ω is a faithful normal state on $\mathcal{A}$, then there exists a a probability space (Ω, Σ, μ) , with $\mathcal{A} = L^{\infty}(\Omega)$, and $\omega = \mu$.

Proof. See [16, p7] for a proof. ■

Except for detaling the mapping of the morphisms, this concludes the construction (in principle) of the functor in the reverse direction. We now have functors in both directions between probability spaces and von Neumann algebras. Note that we had to modify the measure space a bit to make this work, which means that we will not be able to claim they are fully equivalent as categories (see [22]). Even so, this set of functors enables us to give a new definition of a probability space in the purely algebraic domain. We collect the essential algebraic properties in the following definition of a algebraic probability space [15].

Definition 2.2.3 — Commutative algebraic probability space. A commutative algebraic probability space is a pair $(\mathcal{A}, \omega)$ where:

1. The algebra $\mathcal{A}$ is a commutative von Neumann algebra (W*-algebra).
2. The map ω is a normal (faithful) state on $\mathcal{A}$.

This might also be named commutative W*-algebraic probability space to prepare for the eventuality of using other types of algebras. Other types of algebras will be used but we will not use such a detailed terminology to denote them.

We will now interpret this construction by associating words from classical probability theory to elements of this structure. We refer to elements of $\mathcal{A}$ as random variables, and we call ω a probability distribution. Combining the probability distribution ω with a random variable $A \in \mathcal{A}$, results in the expectation value $E(A) = \omega(A)$.

Definition 2.2.4 — Expectation values. The expectation value of some $A \in \mathcal{A}$ is given by $E(A) = \omega(A)$.

To get to the "proper" probability distribution, that gives probabilities instead of expectation values, we need to use a special set of random variables (or observables) that we now define. The set of observables that are valued only in the set $\{0, 1\}$, corresponds to the set of true or false propositions that the theory can make. These observables are given by the set of all projections of the algebra.

Definition 2.2.5 — Projections. Any element $P \in \mathcal{A}$ such that $P^2 = P$ and $P^\star = P$ is called a projection. The set of all projections is denoted by $\mathcal{P}(\mathcal{A})$.

The set of events Σ of a classical probability space is given by exactly this the set of all projections $\mathcal{P}(\mathcal{A}) \subset \mathcal{A}$ of the algebra. Intersections, unions, and complements on the sample space corresponds to the logical operations of "and", "or" and "negation" between the events. On the space of projections this must be represented by slightly different operations. The "and" operation is replaced by the product of projections, like this

$$E_1 \wedge \ldots \wedge E_k \coloneqq P_1 \cdot \ldots \cdot P_k. \tag{2.2.4}$$

The "or" operation is replaced by the sum given by

$$E_1 \vee \ldots \vee E_k \coloneqq P_1 + (1 - P_1)P_2 + \sum_{i=3}^{k} (1 - P_1) \cdot \ldots \cdot (1 - P_{i-1})P_i. \tag{2.2.5}$$

The "negation" operation is given by

$$\neg E_1 \coloneqq (1 - P_1). \tag{2.2.6}$$

The expectation values of projection have a natural interpretation as the probability of the corresponding event. Thus the probability of an event is given by $Prob(P) = \omega(P)$.

Definition 2.2.6 — Born probability. The Born probability of an event represented by the projector $P \in \mathcal{P}(\mathcal{A})$ is given by $Prob(P) = \omega(P) \in [0, 1]$.

In all of the above we have avoided introducing any representation of the algebra $\mathcal{A}$. We have defined projections and spectra without reference to the algebra acting on any vector space. Any von Neumann algebra can in fact be represented as a set of operators on a Hilbert space. We have chosen to avoid this for several reasons. First, it is not necessary to do so, and this makes the essential structure stand out more clearly. Second, it is often not clear when introducing a Hilbert space, whether it is matter of convenience or an essential part. Third, to the extent that it is necessary to introduce a representation it is often not clear what necessary extra choices this entails. We therefore choose to do as much as we are able to with just the algebra itself.

Noncommutative probability spaces

> Quantum mechanics is what you would inevitably come up with if you started from probability theory, and then said, let's try to generalize it so that the numbers we used to call "probabilities" can be negative numbers. As such, the theory could have been invented by mathematicians in the nineteenth century without any input from experiment. It wasn't, but it could have been.
>
> Aronson

In this section we generalize the commutative probability spaces defined above to the noncommutative realm. We will show that this generalization reproduces classical probability theory when made commutative. We will also show that this generalization continues to make sense as an extended theory of probability. The starting definition is identical to the one we used in the previous section with the exception that we now allow for a noncommutative algebra.

Definition 2.2.7 — General algebraic probability space. A general algebraic probability space is a pair $(\mathcal{A}, \omega)$ where

1. The algebra $\mathcal{A}$ is a (not necessarily commutative) von Neumann algebra (W*-algebra)
2. The map ω is a normal (faithful) state on $\mathcal{A}$

In the classical commutative theory we must by necessity demand the algebra to be a (commutative) von Neumann theory to create a theory that is (close to) isomorphic to the one based on measure spaces. In the noncommutative case we could allow other algebras like *-algebras and C*-algebras. In the finite dimensional case there is no difference between these types. Even in the infinite dimensional case there is some overlap. We shall not focus on what differences result from choosing between *-algebras, von Neumann and C*-algebras. We shall simply assume that all these types of algebras can be used to define a probability theory[23].

Starting from the basic definition we need to identify the other necessary concepts of a probability theory. When the algebra is noncommutative we can no longer identify the algebra with an algebra of functions on a set. This also means that there is no point-set to act as a sample space. However the sample space is a rather passive element in probability theory, and its presence is not actually essential for defining probability theory. The key concept that is needed is the idea of an event, and this can still be defined when the algebra is noncommutative. We introduce events in the same manner as above, by defining the set of events Σ to be given by the set of projections $\mathcal{P}(\mathcal{A}) \subset \mathcal{A}$ [24]. If the algebra were to be represented on a Hilbert space, these projections would correspond to closed sub-spaces of the Hilbert space. We define analogs of unions, intersections and complements on this space in the exact same manner as given previously. The (Born) probability of an event represented by the projector $P \in \mathcal{P}$ is given by $Prob(P) = \omega(P) \in [0, 1]$.

We do want to establish that this new structure reproduces the previous commutative structure for the appropriate conditions.

Lemma 2.2.8 A commutative general algebraic probability space is (almost) equivalent to a classical probability space.

Proof. This follows from theorems 2.2.1 and 2.2.2. ■

2.3 Algebraic formulation of mechanics

Classical mechanics

In this section we reformulate classical mechanical systems (including classical statistical mechanics) as commutative algebras (representing observables) with linear functionals (representing states)

(see [12], [13] and [14] for similar formulations). As already suggested, quantum mechanics will be formulated in the same manner. This formulation has a close relation with algebraic probability theory. The initial description will ignore questions of dynamics (see section 2.5 for a discussion of dynamics).

To start off we present the Riesz-Markov representation theorem to establish the correspondence between a functional on $C(X)$, and a measure on X.

Theorem 2.3.1 — Riesz-Markov representation theorem. If ω is a positive linear functional on $C(X)$ (the set of continuous functions from X to $\mathbb{R}$), then there is a unique (Radon) measure, called the Riesz-Markov measure μ_ω on X, such that $\omega(f) = \int f \, d\mu_\omega$ for all $f \in C(X)$.

Proof. See [21, p.212]. ■

Let us first set up the traditional description of classical mechanics in the Hamiltonian version (see e.g. [25, chapter 24], [26, chapter 1], [27], [28] and [29]). The ingredients we need are states and observables. Underlying both states and observables is the concept of a phase space. Phase space defines the range of values that position and momentas can take. The phase space is assumed to be compact (for simplicity). We represent states as probability measures on the phase space, and we represent observables as (continuous or smooth) complex functions on this manifold. We do not need a symplectic structure for now since we are not discussing dynamics.

Definition 2.3.2 — Phase space. The phase space Ph of a single particle living in m-dimensional space is a $2m$-dimensional real compact smooth manifold.

Definition 2.3.3 — Phase space state. A state is a probability measure on Ph. A pure state is a state defined by Dirac measures (point measures). I.e. a pure state corresponds to a point $m \in Ph$. A mixed state is a state that is not pure.

Definition 2.3.4 — Algebra of observables . The algebra of observables of a classical system with a phase space Ph is the set $C(Ph)$ (or $C^\infty(Ph)$). We define involution as complex conjugation, and pointwise addition and multiplication in the usual manner, making it into an *-algebra. With the norm given by the supremum norm we get a C*-algebra. (Note that we are assuming that Ph is compact.)

This completes the listing of elements involved in the traditional description of Hamiltonian systems. The next task is to to come up with an equivalent algebraic definition.
Let us assume in the algebraic approach that we are given a commutative algebra $\mathcal{A}$ and a linear functional ω on $\mathcal{A}$. By the Gelfand-Naimark theorem we can reconstruct a topological space from the algebra of complex functions on it (see section 5.2 and theorem 5.2.17). Thus if we are given an algebra $\mathcal{A}$ that is isomorphic to the algebra of functions on some phase space Ph, we can then reconstruct Ph. We do assume that the algebra is sufficient to also reconstruct the phase space manifold structure. (This assumption is not without problems, we discuss some of them in the notes section of this chapter.) How can we use the linear functional? By the Riesz-Markov representation theorem, any linear functional on $C(X)$ corresponds to a measure μ on X. We can therefore derive a phase space state from a linear functional (and vice versa). The kinematics of a classical systems in algebraic terms is summarized by the next definition (see [14] [13] [12]).

Definition 2.3.5 — Algebraic classical system. An algebraic classical system is

1. A commutative *-algebra $\mathcal{A}$

2. A full set of normal (faithful) states $\omega\colon \mathcal{A} \to \mathbb{C}$

This definition is compact and uncomplicated. It may seem strange that this actually forms a basis for descriptions of a very general set of physical systems. However, it is really not all that different from the usual starting point of classical statistical mechanics, using phases space, phase space functions, and probability measures. In fact it has a simple translation back into the traditional language. The algebra describes both the set of phase-space functions and the phase space. This includes the topology (and potentially the differential structure) of the phase space manifold. A single functional describes a single physical state given by a probability measure on the phase space. The set of all possible functionals describes the set of all possible states.

Quantum mechanics

Quantum mechanics (QM) can be formulated algebraically exactly like we did with classical mechanics.[13][30] [12][5] And just like we did away with the phase space in the classical case, we will do without the Hilbert space in the quantum case. We exclude dynamics for now. It is of note that most aspects of the classical-quantum division can be addressed without including dynamics.

Before we present QM like an algebraic system we express it by way of more commonly used axioms. We will then give a few key results, using the traditional framework, but using concepts that we will later utilize in the algebraic reformulation. This hopefully makes it more clear, already at an intuitive level, how the algebraic system can reproduce the content of the traditional formulation. There are many different axiomatizations of quantum mechanics [12][5]. We choose to use one that covers simple non-relativistic systems.

Definition 2.3.6 — Quantum system. A quantum system is a a tuple $(\mathcal{H}, S, B(\mathcal{H}), \mathcal{O})$ such that

1. $\mathcal{H}$ is a separable Hilbert space.
2. The space of states S is given by the set of one-dimensional sub-spaces of $\mathcal{H}$.
3. $B(\mathcal{H})$ is the set of bounded linear operators on $\mathcal{H}$.
4. The set of observables is the subset of self-adjoint elements $\mathcal{O} \subset B(\mathcal{H})$.

The first result will be a reformulation of the set of states, translating this into a subset of the operators on the Hilbert space. Since we will later (temporarily) discard the Hilbert space, this reformulation is quite useful.

Lemma 2.3.7 Each one-dimensional subspace of $\mathcal{H}$ corresponds to a projection operator $P \in B(\mathcal{H})$. The projection operator corresponding to the subspace containing the vector $|\psi\rangle \in \mathcal{H}$ is given in dyadic notation as $P_\psi \coloneqq |\psi\rangle\langle\psi|$.

More precisely the dyadic notation is shorthand for an element of $\mathcal{H} \otimes \mathcal{H}^*$, where $\mathcal{H}^*$ is the dual space of $\mathcal{H}$. An element T of $\mathcal{H} \otimes \mathcal{H}^*$ clearly defines a linear map $T\colon \mathcal{H} \to \mathcal{H}$.

Theorem 2.3.8 — Spectral theorem. Any bounded self-adjoint operator $\hat{O} \in \mathcal{O}$ can be expressed as

$$\hat{O} = \sum_{\lambda \in \sigma(\hat{O})} \lambda \, |\psi_\lambda\rangle\langle\psi_\lambda| = \sum_{\lambda \in \sigma(\hat{O})} \lambda P_\lambda$$

Where $\sigma(\hat{O})$ is the spectrum of $\hat{O}$ and ψ_i is the corresponding eigenvectors, and P_λ is the corresponding one-dimensional projection .

A mathematical system has many component that we can investigate and find the properties off. However, a mathematical system does not become physics until we assign the system some semantic rules. Such rules may be (or at least feel) very natural, but it is important to note that such rules

are in principle arbitrary and can never be derived from the mathematical theory itself. Indeed in early QM the semantic rules were very much in flux, and it was not at all clear to the creators of the theory what the appropriate rules were. Establishing a semantic rule is a two step process consisting of first defining some mathematical object and then assigning it a real-world interpretation. To not be overly pedantic in these matters we do both in one go. We do not actually give a concrete procedure for what these terms "look" like in the experimental situations but instead rely on using commonly used words that have a reasonably well agreed meaning. One could say that using word like states and observables also in many cases represent semantic rules, usually we will not make such implications, but the reader will sometimes have to rely on context to decipher what is the intended meaning.

Definition 2.3.9 — Interpretation of a quantum system. We create an interpretation of a quantum system $(\mathcal{H}, B(\mathcal{H}))$ by making the following identifications

1. The expectation value of an observable $\hat{O}$ with respect to the state ψ is given by
$$E_\psi(\hat{O}) = \langle\psi|\,\hat{O}\,|\psi\rangle \tag{2.3.1}$$
2. The probability distribution on the spectrum $\sigma(\hat{O})$ of $\hat{O}$ with respect to the state ψ is given by
$$\mu\,(i) \coloneqq \langle\psi|\,\hat{P}_i\,|\psi\rangle = \langle\psi|\psi_i\rangle\,\langle\psi_i|\psi\rangle = |\,\langle\psi_i|\psi\rangle\,|^2. \tag{2.3.2}$$

We can instead of these axioms just focus on the algebraic properties of the operators, and the set of linear functionals (defined by the states through the expectation values). In this direction we first note the two following theorems [12][13][14].

Theorem 2.3.10 — C*-algebra. The algebra $B(\mathcal{H})$ is a unital C*-algebra.

Theorem 2.3.11 — States. The set of Hilbert space states are normal (faithful) linear functional (i.e. algebra states) on the algebra $B(\mathcal{H})$.

We have started with traditional axiom and laid out some important consequences and definitions in a language suitable for out goal. These results are meant to make it plausible that it is actually possible to reformulate the whole system without using a Hilbert space. In analogy with what we did for classical mechanics we start with the following definition. Note that the only change from the definition of a classical system is that the algebra is now noncommutative.

Definition 2.3.12 — Algebraic quantum system. An algebraic quantum system is

1. A noncommutative C*-algebra $\mathcal{A}$
2. A full set of normal (faithful) states $\omega\colon \mathcal{A} \to \mathbb{C}$

We now validate the choices made in this definition by indicating how one derives the axioms that was used previously.

Deriving all quantum properties

We have introduced various more abstract definitions of quantum systems above. We want to explicitly relate the definitions used to standard formalism used. The essential fact that we need is stated in the following theorem. It tells us that an abstract C*-algebra can always be seen (represented) as an algebra of operators on a Hilbert space.

Theorem 2.3.13 — C*-algebra representation. Any C*-algebra is isomorphic to an algebra of operators on a Hilbert space.

Proof. See [30, p15] or [31, p33] ■

This means that any C*-algebra can be realized as an operator algebra on a Hilbert space. In our algebraic approach the real-valued observables would correspond to the set of self-adjoint elements of the algebra. Together this is basically the same as the ordinary definition of quantum observable as the set of self-adjoint operators on some Hilbert space.

We will quickly discuss the translation of states from our algebraic framework to the standard formalism. The linear functionals of the algebra represents expectation values of the observables [30]. We can identify this with a vector in a Hilbert space because we now have a representation of the algebra. For the specific case of pure states, this is equivalent to algebra states can be represented by a unit vector in the Hilbert space. Knowing the expectation value for any state allows us to reconstruct the probability distribution for this variable. The full reconstruction of the Hilbert space uses the GNS construction [30][12][13]. This establishes the equivalence of the algebraic formulation to the usual von Neumann formulation of quantum mechanics.

2.4 Mechanics as a probability theory

The whole point of the analysis in section 2.3 was to reformulate the description of physical systems in a common algebraic form, and then to see that this form corresponds closely to algebraic probability theory. The following paragraphs contains a quick and rather trivial check that the definitions that we have obtained for probability theories and physical theories are in agreement.

Classical mechanics

To establish that the kinematics of classical mechanics can be represented by a commutative probability space, extended with a full set of states, we work our way through the definitions. Using very similar language in all the definitions this task is already partially accomplished. The first step is the reproduction of the measure theoretic structure of statistical mechanics in the algebraic version of commutative probability theory. We only need to establish that the algebra of observables (or a subset of it) is a von Neumann algebra. We have in fact already established that any such space that has the structure of a measure space has an equivalent representation by a von Neumann algebra. This establishes the essential part of what we want to claim. One would usually say that a phase space of classical (statistical) mechanics has additional differentiable structure beyond its topological and measure space structure. We assume that this differentiable structure is also determined by the algebra, even though no simple reconstruction theorem exists (see chapter 5).

Quantum mechanics

We now complete the picture by establishing that we can place QM under the umbrella of a noncommutative probability space extended with a full set of states. We have actually already almost completed this step. The definitions of a quantum system and a noncommutative probability space are identical except for the detailed nature of the algebra. We specified a C*-algebra for QM and a W*-algebra for noncommutative probability. For the case of finite dimensional and infinite dimensional non-relativistic QM the relevant algebras possesses both qualities. In the relativistic case this is a significant issue, and it is not quite settled which choice is more appropriate.

2.5 Dynamics

In this section dynamics will be added. We will do this for both the classical and the quantum case. The actual implementation of dynamics in both cases will be identical. Concretely, dynamics will be defined by using one-parameter families of algebra automorphisms [32][12]. Sometimes dynamics can be derived from an underlying structure. When present, the underlying structure inducing this implementation is similar but not identical between classical and quantum systems.

Classical dynamics

In general the time evolution of a classical system can be described as a one-parameter Abelian group of algebra automorphisms. [12] In this regard we are following a path close to the Heisenberg picture of time evolution, where the observables change but the states remain fixed [33]. Of course, there is no real meaning to change of either of these parts alone, change is only meaningful when combining states and observable, and this is what leads to time dependent measurements. The Heisenberg picture is most familiar from QM, but applies equally to CM [33]. We do not a priori require the time-evolution to be derivable from any other principles, it is simply a definition of what time-evolution must mean. However, in most cases this family of automorphisms can be derived from a structure already present. Let us see how this usually comes about. In the traditional formulation, phase space is expected to carry the structure of a symplectic manifold (or even a cotangent bundle, which is always symplectic) (see chapter F and section F.4 for a brief review of symplectic manifolds). The symplectic structure gives rise to an additional product on the associative *-algebra of phase space functions. This additional product is a Lie algebra product, which is also a derivation on the associative *-algebra. An associative algebra with an added Lie algebra structure is called a Poisson algebra.

Definition 2.5.1 — Poisson algebra. A Poisson algebra over the field K is a triple $(\mathcal{P}, \cdot, \{\cdot,\cdot\})$ such that the following holds.

1. The item $\mathcal{P}$ is a K-module.
2. The pair $(\mathcal{P}, \cdot)$ is an associative K-algebra.
3. The pair $(\mathcal{P}, \{\cdot,\cdot\})$ is a Lie algebra over K.
4. The bracket $\{\cdot,\cdot\}$ is a derivation on $(\mathcal{P}, \cdot)$, i.e. $\{a, b\cdot c\} = \{a,b\}\cdot c + b\cdot\{a,c\}$.

We will only consider $K = \mathbb{R}$ or $K = \mathbb{C}$.

In the setting of a phase space with a Poisson bracket, selecting one specific phase space function, leads by using the Poisson bracket to a definition of a derivation on the set off all phase space functions (see section F.4). By these means we can postulate the traditional time evolution equations used in Hamiltonian mechanics. Let $f_t \colon [0,1] \to \mathcal{A}$ represent a time-evolution path f_t in the space of phase space function $\mathcal{A}$ that we wish to determine. Let $h(x)$ be an arbitrary phase space function that we will use to define time evolution. The Poisson bracket with one slot filled by $h(x)$ is a derivation on the algebra of functions $\mathcal{A}$. Assuming we have specified f_t at some initial point we can write the following differential equation determining the path f_t.

$$\frac{\delta\,[f_t]}{\delta t} \coloneqq \{h, _\}\,[f] = \{h, f\} \tag{2.5.1}$$

The bracket symbol with the empty underline is meant to signify the bracket as an operator taking a single function (in this case f) as an input variable. From a more geometric perspective a derivation is equivalent to a vector field. A vector field defined by a derivation can be integrated to give curves which define one-parameter flows on the phase-space. We took the detour to symplectic manifolds and phase space to use this more familiar terrain to introduce the time evolution logic. We could equally well do this directly in the algebraic framework. Just as the combination of any phase space function and the symplectic form on phase space induces a flow on the phase space, in the

purely algebraic version, the Poisson bracket of the Poisson algebra combined with one chosen algebra element induces a one-parameter flow on the algebra elements. We round of this segment by summarizing our discussion in two slightly different definitions of a classical dynamical system. The first leaves the nature of the automorphisms unspecified but the second includes the Poisson structure to derive the automorphism from.

Definition 2.5.2 — Classical algebraic dynamical system. A classical algebraic dynamical system is given by

1. A commutative *-algebra $\mathcal{A}$.
2. The complete set of normal (faithful) states $\omega : \mathcal{A} \to \mathbb{C}$.
3. A one-parameter group of automorphisms of $\mathcal{A}$.

Definition 2.5.3 — Classical algebraic dynamical system (with derived dynamics). A classical algebraic dynamical system (with derived dynamics) is given by

1. A commutative *-algebra $\mathcal{A}$
2. The complete set of normal (faithful) states $\omega : \mathcal{A} \to \mathbb{C}$
3. An extra Poisson bracket structure on the algebra making $\mathcal{A}$ a *-Poisson algebra.

Quantum dynamics

In quantum system dynamics can also be described as a one-parameter Abelian group of algebra automorphisms [12]. In the general case the family of automorphisms will have to be supplied as an extra piece of data. In this case the situation is identical to the classical case and we do not need to discuss these case further. However, as in the CM case the set of automorphism can often be derived from a choice of one specific element (observable) of the algebra. The derivation of a time-evolution in the quantum case is similar to the classical derivation based on a Poisson algebra. The type of algebra used in the quantum case is called a Jordan-Lie-Banach algebra (JLB)[34]. This is a type of algebra with two product operations, where one is commutative and the other is a Lie product. We start by giving the definition.

Definition 2.5.4 — JLB algebra. A JLB algebra is an algebra with two products ($x \circ y$ and $x \bullet y$) that satisfy

1. $x \circ y = y \circ x$ (commutative)
2. $x \bullet y = -y \bullet x$ (antisymmetric)
3. $x \bullet (y \bullet z) + y \bullet (z \bullet x) + z \bullet (x \bullet y) = 0$ (Jacobi identity)
4. $x \bullet (y \circ z) = (x \bullet y) \circ z + y \circ (x \bullet z)$ (Leibniz rule)
5. $(x \circ y) \circ z - x \circ (y \circ z) = (x \bullet z) \bullet y$ (non-associative property)
6. $\|x \circ x\| = \|x\|^2$
7. $\|x \circ x\| \leq \|x \circ x + y \circ y\|$

Definition 2.5.5 — Associated JLB algebra. Any C*-algebra defines an associated JLB algebra by the following defining the following two product operations

1. The non-associative but commutative Jordan algebra product is given by

$$a \circ b = a \cdot b + b \cdot a \tag{2.5.2}$$

2. The JLB Lie algebra product given by the commutator

$$a \bullet b = [a, b] = a \cdot b - b \cdot a. \tag{2.5.3}$$

Thus the starting point will be the C*-algebra used in the definition of a quantum system. From this algebra we derve the associated JLB algebra. It is then true by definition that the Lie algebra structure of the associated JLB is a derivation of the commutative JLB product. This derivation defines the time evolution of the system. Give any algebra element this induces a one-parameter flow on the algebra. The flows can be seen as giving solutions of the differential equations specifying the time-evolution of observables.

The situation with dynamics in the quantum case turn out to be quite similar to the classical case, however, we need to note one crucial difference. In the classical case the Lie algebra structure used to specify the derivations on the algebra had to be given as an extra structure, but in the quantum case this structure was derived directly from the existing algebra of observables. For presentational symmetry with the classical section, we end the discussion of the quantum system with the two corresponding definitions for quantum systems presentational [12][13][14].

Definition 2.5.6 — Quantum algebraic dynamical system. An quantum algebraic dynamical system is defined by

1. A noncommutative C*-algebra $\mathcal{A}$
2. A complete set of normal (faithful) states $\omega\colon \mathcal{A} \to \mathbb{C}$
3. A one-parameter group of automorphisms of $\mathcal{A}$.

Definition 2.5.7 — Quantum algebraic dynamical system (with derived dynamics). A quantum algebraic dynamical system (with derived dynamics) is defined by

1. A noncommutative C*-algebra $\mathcal{A}$.
2. A complete set of normal (faithful) states $\omega\colon \mathcal{A} \to \mathbb{C}$.
3. A derived JLB algebra structure.

2.6 Composite systems

In this brief section we consider the handling of composite systems in classical and quantum theory. Although this section is very brief, the facts herein are of the greatest importance. It may be claimed that it is in the properties of composite system that the greatest divide between classical and quantum systems arises. Remarkably, this is not due to any difference in how composite systems are defined or described, but is simply comes down to the presence or absence of noncommutativity. Let us now see how this happens.

The description of composite system by tensor products of simpler systems is usually considered a separate axiom of quantum theory [35]. The corresponding situation in classical systems is seldom explicitly discussed. However, in the algebraic framework we are adopting, the properties of classical and quantum composite systems can be treated in the same manner. This means that the exact same notions and axioms applies to both to classical and quantum systems. General composite system are simply given by tensor products of the algebras representing the subsystems.

The differences in classical and quantum case comes from the intrinsic properties of the algebras. In the classical case the tensor product structure leads to the usual Cartesian product states because the algebras are commutative [36][37]. There is therefore no entaglement in classical systems. In the case of of quantum systems and their associated noncommutative algebras, the tensor product axiom universally leads to the presence of entangled states [36][37]. When identical particles are present a subspace of the full tensor product space must be used.

The importance of entanglement can hardly be overstated. The whole concept of quantum information theory and quantum computation centers around this phenomenon [35]. It is the presence of entanglement that excludes the possibility of hidden variable classical theories. Entanglement could

also be the explanation for the emergence of a classical world and the emergence of the laws of statistical mechanics (see section VII) [38]. Perhaps more surprisingly, the entanglement structure of quantum systems will also be very important for quantum gravity consideration in chapter 9 and 10. Exploring the consequence of entanglement could easily fill many chapters on its own. We will not go further into the details of this subject at this moment.

2.7 Notes

Extra references

Useful general references using partly or fully an algebraic perspective on quantum mechanics are [30] [26] [39] [13] [14] [40, 41, 42].

Note on generality

We have claimed in this chapter that a three-part structure can describe physical theories. How general is this claim? Certainly CM, QM and QFT seems to be included. It can also be applied to classical field theories. String theory should also be amenable to such a treatment. We assume (without proof) that path-integral formulations of dynamics are equivalent to some form of algebra automorphism setup. In the case of background independent theories the algebra automorphism part has to be replaced by a relational definition of dynamics. General relativity and loop quantum gravity fit this modified schema. In noncommutative geometry models the general schema more or less applies but in addition we need to include a representation of the algebra and a Dirac type operator to define the full model.

Note on states

We have defined physical systems as algebras with states. Strictly speaking, since we are including all states, the specification of the states is redundant. The set of available states is defined by the algebra itself. We still chose to do it in this manner as it stays more similar to the probability spaces and also to the common perception of physics. Note that for the probability spaces we are talking about one specific state, so in this case the information is not redundant.

It is not absolutely clear what restrictions should be put on algebra states to correspond to a physical state (Hilbert space state, density matrix or phase space probability distribution). We require any algebra states to be positive and normalized. We require any algebra state corresponding to physical states to be normal and possibly faithful.

3. Fixed background theories - QFT

> ... our purpose in theoretical physics is not just to describe the world as we find it, but to explain – in terms of a few fundamental principles – why the world is the way it is.
>
> Weinberg

Summary

QFTs are quantum theories defined on a fixed background spacetime. Such theories derive their structure from representations of the background space isometry group. The isometry group of Minkowski space is the Poincaré group. Finding irreducible representations of this group leads almost directly to Yang-Mills theories with spin-1/2 fermions and scalar fields. Gauge redundancies are an inevitable consequence of implementing massless spin-1 Poincaré representations as vector fields. Spontaneous breaking of gauge symmetries by scalar fields is a naturally occurring possibility in such models. In the low energy EFT limit, all non-renormalizable interactions are suppressed. There is a substantial number of free parameters and one specific set of parameter values defines the standard model.

In this chapter we will analyze physical theories defined on fixed background spacetimes. Some parts of the discussion will involve general theories on arbitrary backgrounds, but for the most part we will be dealing with quantum theories on Minkowski space. We will refer to a quantum theory defined on a fixed background Minkowski spacetime as quantum field theory (QFT). A large part of the explosion of knowledge that modern physics represents stems from the use of QFT. QFT is the backbone of the standard model (SM), and this model has lead to the most accurate predictions ever seen in scientific theories [43]. It is therefore essential to seek the most substantive and satisfying way to understand QFT.

3.1 Introduction

> ...although you cannot argue that relativity plus quantum mechanics plus cluster decomposition necessarily leads only to quantum field theory, it is very likely that any quantum theory that at sufficiently low energy and large distances looks Lorentz invariant and satisfies the cluster decomposition principle will also at sufficiently low energy look like a quantum field theory.
>
> Weinberg

The viewpoint on QFT and gauge symmetries presented in this chapter is slightly different from the one found in many current QFT books. Our presentation leans on the more foundationally oriented texts of Weinberg [44], Wald [45] and McCabe [46]. We also take some clues from the work of Arkani-Hamed and collaborators [47].

While our presentation here agrees fully in its physical content with canonical presentations of QFT, the emphasis and the logical ordering is different. In most presentations, QFT is seen as a quantization of classical field theories, and gauge symmetries are seen as a new and fundamental principles of nature. Specific equations like the Klein-Gordon equation and the Dirac equation are presented through their historical route, and seen as brilliant singular insights that have been handed down to the present age by history. Renormalization is often presented as an obvious a priori requirement that any sensible theory must satisfy. Thus, while introducing basically the same rules and structures as the present chapter will do, no proper fundamental basis is given for these concepts. The epistemological status given to the various "principles" that are introduced are very misleading, and the whole setup is not conducive to a productive analysis of the basis for modern QFT.

The approach used in this chapter is to deduce QFT from a set of three fundamental principles. The three principles are, the general theory of probability, the isometry group of Minkowski space, and the cluster decomposition principle (CDP). To prepare for this quite lengthy sequence of elucidation (see figure 3.4), we first briefly present the overall structure of the argument.

3.2 Summary of deducing QFT

> ...non-Abelian gauge theories... the important feature of these theories is not gauge invariance but the existence of massless spin-1 particles that are charged ...
>
> Schwartz

Isometries makes it possible to define many different observational frameworks for the same event. The observational frameworks are related by these isometries. The relations provided by the isometries make it both arbitrary and inconsistent to prefer one of these frameworks over another. We are hence led to think that these observers must in some sense observe "the same things" when looking at the same event [45]. These considerations have profound consequences enabling us to deduce many of the central structures of physics.

Using the isometries of Minkowski space we deduce the existence of mathematical entities, which we naturally and convieniently identify with free particle states, with specific spin and mass. This closely follows Wigner's ground breaking analysis [48][49].

The next step is to analyze the time-evolution of such systems, in the restricted setting of experiments with asymptotic states corresponding to free particles states. Demanding manifest Poincaré invariance, locality and unitarity, and requiring that the principle of cluster decomposition holds, we find that the time-evolution structure essentially must be expressible as a local self-commuting

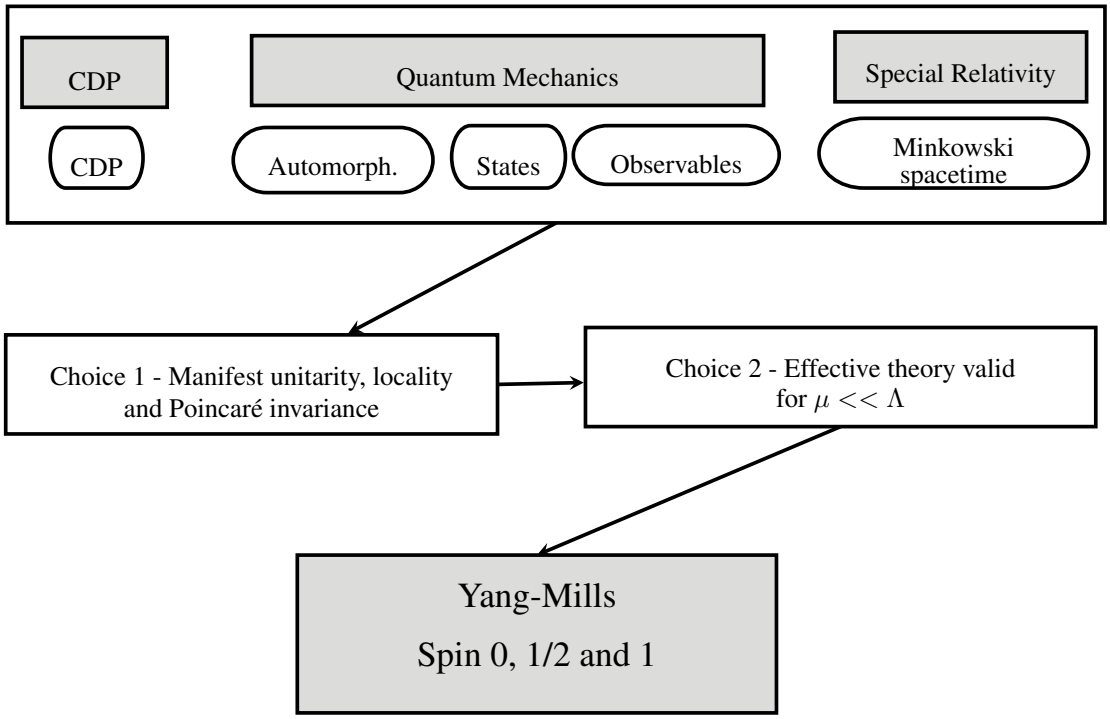


Figure 3.1: The diagram shows how the structure of QFT is build from the three basic principles (CDP + QT + SR). Note that by SR we are referring just to the geometric structure of Minkowski space. See chapter 2 for an explanation of the three-part structure indicated for QM/QT. The symbol μ represents some accessible "low" energy, which is small compared to some hypothetical "new physics energy" Λ. Abbreviations: QT=Quantum theory, CDP=Cluster decomposition principle

Hamiltonian scalar density constructed from operator fields [44]. These (free particle) operator fields must themselves be constructed from the Fourier transforms of creation and annihilation operators of the free particle states. This result already establishes most of the basic structure of modern quantum field theory. If we include massless spin-1 fields among the asymptotic states, this necessarily leads to introducing gauge redundancies [50][51].

The basic structure of gauge theory is thus in place, but there is an infinite set of possible Lorentz and gauge invariant terms in the Hamiltonian. An effective field theory (EFT) based understanding tells us that the only important terms at low energies are terms that are renormalizable [51][44]. This set of terms is quite restricted.

If Minkowski space actually has a time orientation as well as a spatial orientation, the isometry group is restricted from the full Poincaré group to the proper isochronous Poincaré group [46]. This restriction leads to the possibility of chiral spin-1/2 particles. The combination of chiral fermions with gauge invariance excludes mass terms for fermion fields [43]. Scalar fields with a potential that has a non-gauge invariant vacuum state should a priori be included in the list of possible Lagrangian terms. Such scalar fields define a new perturbative vacuum that induces (emergent) mass terms for both fermions and gauge bosons [43]. We are thus able to deduce the complete structure of renormalizable spontaneous symmetry breaking gauge theories based on spin-0, spin-1/2 and spin-1 fields. Specifying parameter values for this general model leads to a multitude of specific models, one such model is the standard model. See figures 3.1, 3.2, 3.3, and 3.4 for a graphical presentation of this argument. In the following sections we present the details of the various steps.

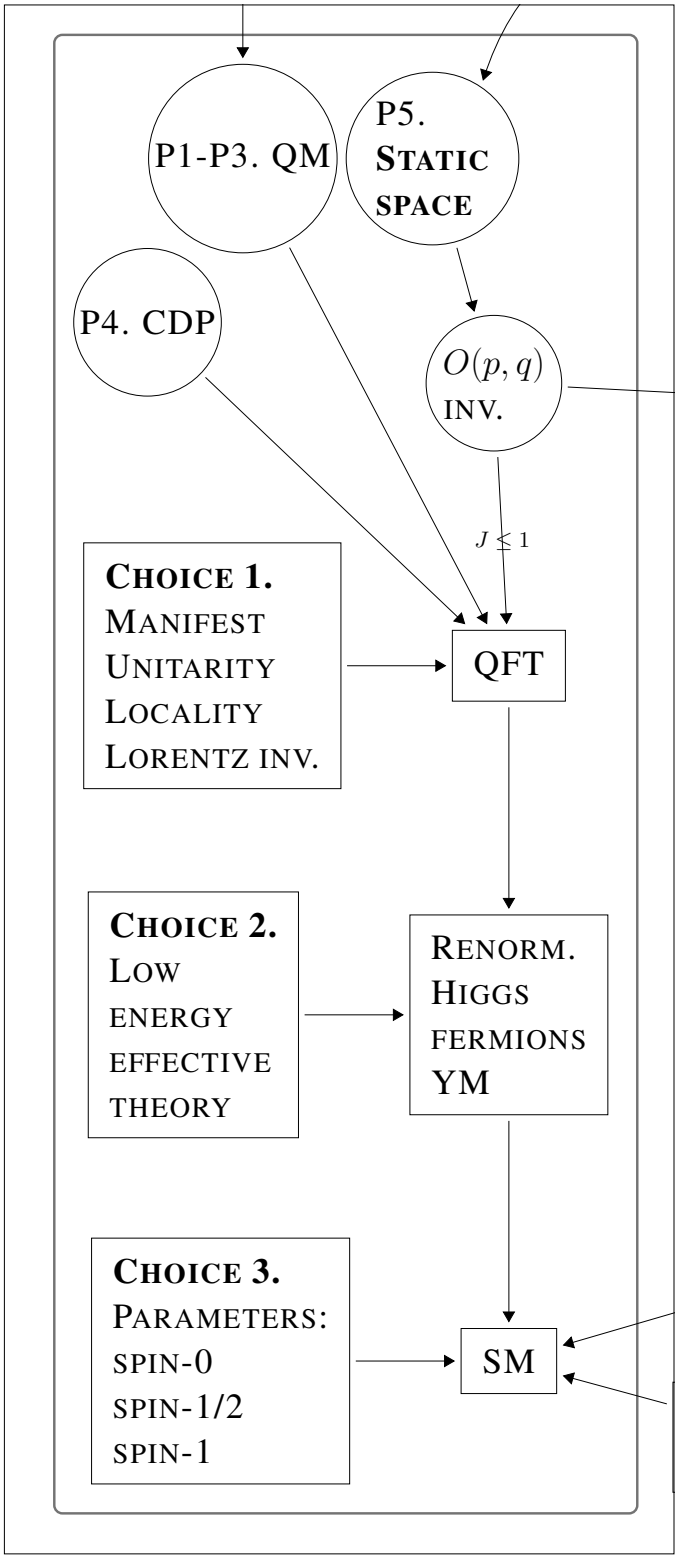


Figure 3.2: The structure of quantum theories in fixed backgrounds. The basic principles of QM, CDP and static space enables us to deduce the structure of QFT. When focusing on low energy effective theories only renormalizable actions based on spin-0, spin-1/2 and spin-1 fields are relevant. The spin-0, spin-1/2 and spin-1 sectors have various free parameters. A set of specific choice for all these parameters defines what is known as the standard model. Abbreviations: see figure 1.

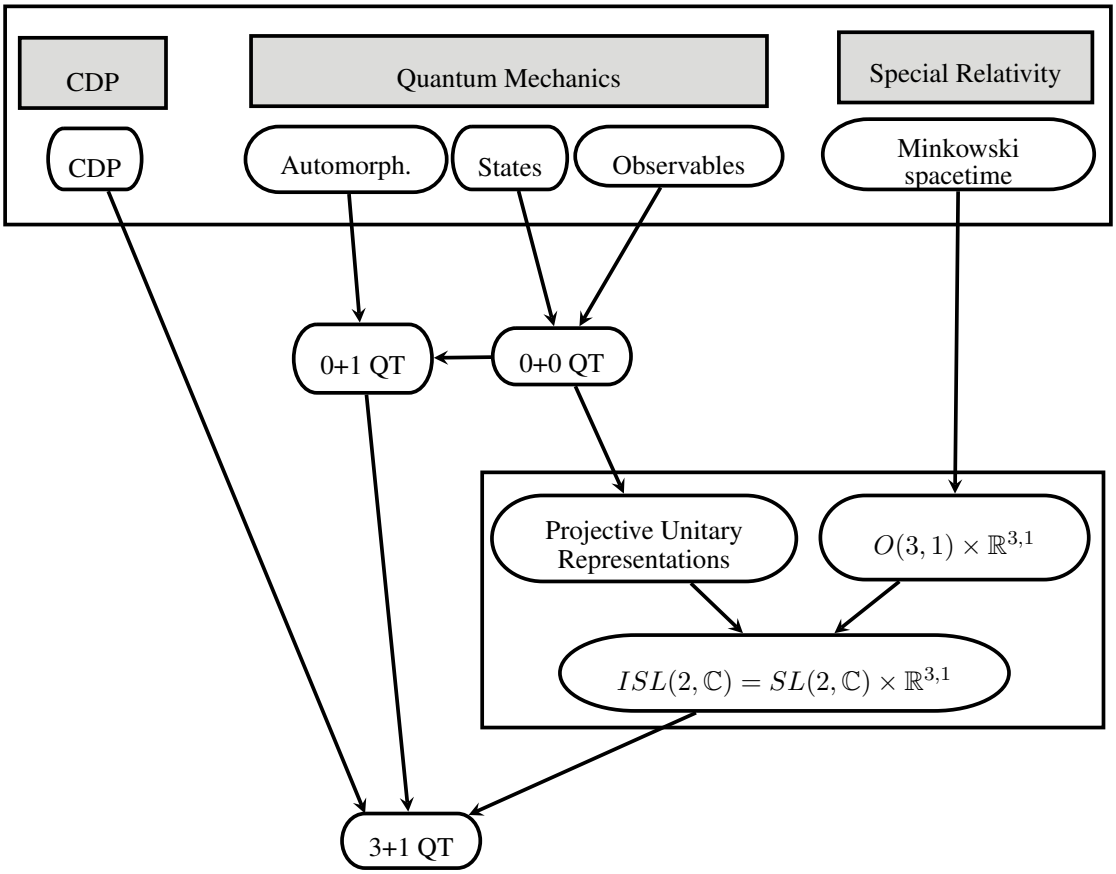


Figure 3.3: The diagram shows how the structure of QFT is build from three principles (CDP + QT + SR). The lower right-hand box shows that QT dictates the use of projective unitary representations of the isometry group, and SR dictates that the isometry group is the Poincaré group. Together these two inputs dictate the use of proper representations of the universal cover of the Poincaré group. Abbreviations: QT = quantum theory, CDP = cluster decomposition principle, SR = special relativity ISL = inhomogenous special linear group

3.3 Isometries in general physical theories

We will start with a short note on terminology and formalism. In the analysis of general physical systems in chapter 2, we found them to be expressed by states and observables (disregarding dynamics). This picture does not explicitly include neither time nor space. After adding automorphisms to introduce dynamics, there is still no concept of space. It is in this sense that quantum mechanics is sometimes referred to as (0+1) QFT, meaning a QFT with no space dimensions and a single time dimension. In the same manner the timeless version (of the early part of chapter 2) could be labeled (0+0) QFT.

In chapter 2 we used a formulation of quantum theory based on noncommutative probability, but it was shown that it was equivalent to the ordinary Hilbert space formulation. It is possible to perform the analysis of this chapter using just the probability framework, but it is more technically complicated than doing the same using the Hilbert space formulation. We will therefore freely use the Hilbert space formulation in what follows.

Isometries and states

We will now argue that the existence of spacetime isometries implies that there exists an action of those isometries on the states of the theory [45] [52] [53]. For a given spacetime M the existence

of an isometry group defines classes of equivalent spacetimes related by isometries. From a mathematical point of view these are identical (with respect to the currently considered properties). We define a reference frame $\mathcal{F}$ by a set of spatially orthogonal coordinates (corresponding to coordinate basis vectors of the tangent space at each point being orthogonal). These reference frames must now be connected with our previous definition of physical systems in terms of states $\mathcal{S}$ and observables $\mathcal{O}$. Only the invariance of *combinations* of observables and states are strictly meaningful, but we will here define an action only on states (we are not assuming that observables are not affected, they will just not enter into the picture in this argument). Let G be the abstract isometry group of M realized for each $g \in G$ as an isometry $\phi_g \colon M \to M$. An isometry also induces a map between coordinate systems as well as between tangent spaces. Let A and B be two observational frames related by the induced map of the isometry ϕ_g. We assume for simplicity that a state $s \in \mathcal{S}$ can be fully characterized by a set of n measurements. More concretely we assume s can be completely specified by an element in $\mathbb{R}^n$. Let $f_A(x, s) \in \mathbb{R}^n$ represent the measurement of the full set of independent properties of the state s in a frame A at the point x. We now demand that for each isometry

$$\phi_g \colon M \to M, \tag{3.3.1}$$

there is a corresponding map $\hat{\phi}_g$ given as

$$\hat{\phi}_g \colon \mathcal{S} \to \mathcal{S}, \tag{3.3.2}$$

such that

$$f_A(x, s) = f_B(\phi_g(x), \hat{\phi}_g(s)). \tag{3.3.3}$$

For each set of isometries $g_1, g_2 \in G$ we have the following relation between the state maps

$$\hat{\phi}_{g_1} \circ \hat{\phi}_{g_2} = \hat{\phi}_{g_1 \cdot g_2}. \tag{3.3.4}$$

Thus the maps $\hat{\phi}_g$ form a representation of the group G. This concludes the argument that for theories defined on a fixed background, the existence of isometries of the background space implies an action, in the form of a representation of the isometry group, on the states of the theory.

3.4 Isometries in quantum theory

> Our only direct experience with the physical universe is by means of the massless representations of the Poincaré group...
>
> Mirman

In the last section we analyzed how isometries affects the states of a generic theory, when isometries are considered as maps between equivalent observational frames. The analysis we presented also applies for more general symmetries than just isometries. Even though our aim is primarily to understand the consequences of isometries, in this section we consider arbitrary symmetries. So far we have not specified any details of the physical theory that these symmetries are acting on. The primary aim of this section is to see how the previous analysis is implemented when the physical theory in question is quantum theory.

From isometries to representations

The concepts introduced in equations 3.3.2,3.3.3 and 3.3.4 can be adapted almost directly for the case of symmetries in a quantum theory. The first adaption that we make is to specify that the observables (or measurements) to be preserved are the transition probabilities.

We start with specifying some notation and a key definition. A quantum symmetry is defined to be an abstract group G that has been given a representation on the set of states. (Note: The notation in the following involves the symbols g, $\hat{\phi}_g$ and $\hat{U}_g$ which are all different objects. In the previous section we used a fourth variation ϕ_g, which is also a different object. The notation of this section matches the previous section, but we are no longer assuming g to be an isometry.)

Definition 3.4.1 — Quantum symmetry group. A quantum symmetry group is a group G, that is represented by maps between states

$$\hat{\phi}_g \colon \mathcal{S} \to \mathcal{S} \tag{3.4.1}$$

such that all probabilities are unchanged. This is to be understood as follows. That the maps $\hat{\phi}_g$ form a representation means that for each $g_1, g_2 \in G$ we have

$$\hat{\phi}_{g_1} \circ \hat{\phi}_{g_2} = \hat{\phi}_{g_1 \cdot g_2}. \tag{3.4.2}$$

Let $Prob(S, R)$ be the probability of obtaining the state R when the initial state was S. That all probabilities are unchanged means that

$$Prob(S, R) = Prob(\hat{\phi}_g(S), \hat{\phi}_g(R)). \tag{3.4.3}$$

The next step is to translate these mappings of states to mappings on the Hilbert space. We recall that a quantum state is one dimensional subspace of the Hilbert space. The subspace can be represented by any unit vector belonging to the subspace. The representation of the symmetry group on the Hilbert space will be different from the representation on the set of states. Before proceeding we need two definitions.

Definition 3.4.2 — Unitary and Anti-unitary. A unitary operator on a vector space V with inner product $(,)$ is a map $U \colon V \to V$ such that

$$U(k_1 v + k_2 w) = k_1 U(v) + k_2 U(w) \tag{3.4.4}$$
$$(U(v), U(w)) = (v, w). \tag{3.4.5}$$

An anti-unitary operator is an operator that fulfills the following criteria

$$U(k_1 v + k_2 w) = \bar{k}_1 U(v) + \bar{k}_2 U(w) \tag{3.4.6}$$
$$(U(v), U(w)) = \overline{(v, w)}. \tag{3.4.7}$$

Unitary operators preserve the inner product of a vector space. Since we are concerned with the modulus squared of inner products we can also allow anti-unitary maps as these do not change the modulus squared.

The following theorem by Wigner tell us exactly how the maps between states is translated to a map between Hilbert space vectors.

Theorem 3.4.3 — Wigner's theorem. A quantum symmetry represented by a group of maps on the set of states, is represented on the corresponding Hilbert space by the action of unitary or anti-unitary operators.

Proof. See [54] for a proof. ■

We note that the properties of linearity and anti-linearity follows from unitarity and anti-unitarity. Continuous quantum symmetries represents an important special case. For continuous symmetries a representation using anti-unitary operators is not possible.

Theorem 3.4.4 A continuous quantum symmetry is represented by the action of unitary operators.

Proof. See [55, p101] for a proof. ■

Since we started by giving the representation of quantum symmetries as an action on the set of states, but we are now applying it to vectors instead of states, we need to find the relation between the group elements and the corresponding Hilbert space operators. If g is an element of the symmetry group G we denote the map between states by $\hat{\phi}_g$, and the map between vectors as $\hat{U}_g$. For each set of group elements $g_1, g_2 \in G$, we have instead of

$$\hat{\phi}_{g_1} \circ \hat{\phi}_{g_2} = \hat{\phi}_{g_1 \cdot g_2}, \tag{3.4.8}$$

a new relation given by

$$\hat{U}_{g_1} \cdot \hat{U}_{g_2} = \exp\{i\theta\}\hat{U}_{g_1 g_2}. \tag{3.4.9}$$

Thus the maps $\hat{U}_g$ forms what is called a projective representation of the group G.

Definition 3.4.5 — Projective representation. A map from group elements to automorphism group of a vector space is called a projective representation of the group if for each group element $g_1, g_2, g_3 \in G$ we have

$$\hat{U}_{g_1} \cdot \hat{U}_{g_2} = \exp\{i\theta\}\hat{U}_{g_1 g_2}. \tag{3.4.10}$$

By this analysis applies we can conclude the following for continuous symmetries in general.

Theorem 3.4.6 A group of continuous symmetries is represented in quantum mechanics by a projective unitary representation on the Hilbert space.

It turns out that whether such projective representations are equivalent to a proper representation of the group depends on whether the group is simply connected. For a group that is not simply connected we can instead use the associated simply connected covering group.

Lemma 3.4.7 A projective unitary representation on the Hilbert space is equivalent to a proper representation of the covering group.

Proof. See [45, p345] for a proof. ■

The isometries of manifolds are continuous symmetry group and in this case the following lemma applies.

Lemma 3.4.8 The group of isometries G is represented on the Hilbert space of a quantum system by a unitary representation of the covering group of G.

Proof. This is immediate from the previous lemma. ■

This concludes our analysis of the connection between a physical theory defined on a fixed background space and unitary representations of the covering group of isometry group of the background space.

3.5 Special relativity

So far we have established how the symmetries of a generic background space are implemented in quantum theory. The next task is to consider specific background spaces. Special relativity is the most common background structure used in QFT. Special relativity can be defined in many equivalent ways, but it is most commonly developed by starting from the two postulates that Einstein gave, the constancy of the speed of light, and the principle of relativity [56].

Definition 3.5.1 — Special relativity (1). Special relativity is the teory defined by the two postulates:

1. The speed of light it constant and equal in all inertial reference frames.
2. Physics is described by the same equations in all reference frames.

Einstein's two postulates gave a beautiful and simple physical basis for the Lorentz transformation. However, the postulates are not conducive to generalizations when expressed in their classical form, and they also obscure the connection to geometry. One could also point out that, while the existence of a limiting speed follows logically from the theory, the question of whether or not anything (e.g. light) actually moves at that exact speed is an independent inquiry. Of course, Einstein gave his postulates in the context of Maxwell's laws, and was motivated by Maxwell's deduction of the universal speed of electromagnetic waves. Today we know that this issue is directly related to the (perfect) masslessness of the photon, which is very difficult to establish empirically [57]. The point being, that the correctness of Einstein's special relativity is not at all dependent on whether or not photons actually move at the universal speed defined by the theory. For all of these reasons we prefer to formulate SR as a statement of geometry, specifically by stating that spacetime is four-dimensional Minkowski space. From this geometric postulate the existence of an invariant maximal speed follows.

Definition 3.5.2 — Special relativity (2). Special relativity is the postulate that physical spacetime is represented by four-dimensional Minkowski space.

Since our analysis is focused on isometries, we want to reformulate this geometric postulate in terms of isometries. The set of isometries of flat four-dimensional Minkowski space is given by the ten-dimensional Poincaré group [45]. The elements of this group represents the various translations, boost and rotations that preserve the Minkowski metric. The isometry group uniquely identifies the space, and all the relevant information about the space can be given by specifying the isometry group.

Definition 3.5.3 — Special relativity (3). Special relativity is the postulate that $O(3,1) \ltimes \mathbb{R}^{3,1}$ is the isometry group of physical spacetime.

3.6 Representations of the Poincaré group and $ISL(2,\mathbb{C})$

From the analysis in the previous sections we have found that the representation theory of the group of isometries is the key to unearthing the physical consequences of implementing a physical theory defined on a fixed background space. For a quantum theory in Minkowski space this means that we must study the representations of the convering group of the $O(3,1) \ltimes \mathbb{R}^{3,1}$ group. The universal covering group of $O(3,1) \ltimes \mathbb{R}^{3,1}$ is the group $ISL(2,\mathbb{C}) = SL(2,\mathbb{C}) \ltimes \mathbb{R}^{3,1}$ [45][46]. When combining Minkowski space with quantum theory we must therefore study the unitary representation of $ISL(2,\mathbb{C})$ [45] [46].

There are no finite dimensional unitary representations of the Lorentz group or the Poincaré

group. This follows from the from the following theorem.

Theorem 3.6.1 Non-compact groups have no finite dimensional unitary representation.

Proof. See [58, p241] for a proof. ■

Even so, the finite dimensional representations of the Lorentz and the Poincaré group are also relevant for developing QFT. For the Lorentz group these can be specified by giving two positive integers or half-integers (m, n) (see appendix 2 or any reference on the Poincaré group). However, as we have seen, the interface of representation theory with quantum theory requires the use of unitary representations, and we must therefore turn to infinite-dimensional representations. For finite-dimensional representations it is always true that we can express any representation in terms of irreducible representations (irreps). For infinite-dimensional representations this is not true in general, but for the Poincaré group (and its cover) it is true that any such representation can be expressed using only irreps [45]. Irreps are by definition non-reducible entities, and since we are seeking the proper mathematical objects to represent indivisible fundamental physical entities, it is natural to focus on irreducible representations.

Thus the goals in now to find infinite-dimensional unitary irreps of $ISL(2, \mathbb{C})$. These irreps are characterized by the values (m, s) of the two Casimir operators m^2 and S^2 of $ISL(2, \mathbb{C})$ [51]. We only consider representations for which m^2 is non-negative, and s is an integer or a half-integer. For integers these irreps are also (proper) Poincaré representations, but for half-integers they are just projective representations [45]. Irreps can be found by the method of induced representations using the stability subgroup (little group) invented by Frobenius. Applications of this technique to the Poincaré group was pioneered by Wigner, Bargman and Mackey [48] [49][59][60]. More details on this technique can be found in [61][62] [63] [44] [64] [45]. The physically relevant representations are listed in table 3.1.

Type	Repr. point	Orbit	Stab. subgrp.	Rep.	Dim.
1 - massive	(E,0)	-	SO(3)	(m,s)	2s + 1
2 - massless	(E,-E)	-	SE(2)	(0,s)	2

Table 3.1: Physically relevant irreducible unitary representations of the $ISL(2, \mathbb{C})$ group. The pair (m, s) shows the values of the two Casimir operators. They represent mass and spin.

3.7 From representations to quantum fields

From an analysis of the joint requirements of SR and QM we find that irreducible unitary representations of universal covering group of the Poincaré group, the group $ISL(2, \mathbb{C})$, is the proper starting point for analyzing the consequences of isometries. The basis vectors of unitary irreps of $ISL(2, \mathbb{C})$ will represent the basic entities of the theory we are building. We will refer to the these vectors as one-particle states or simply particles. For each given pair (m, s) of mass and spin, there is an infinite set of states corresponding to the possible values of momenta. In addition there are the discrete degrees of freedom related to z-direction spin or helicity. The details can be found in the references of the previous paragraph. We state our current starting point as a conjecture.

Conjecture 3.7.1 — Free particle states conjecture. The fundamental entities of a theory combining SR and QM are the basis vectors of unitary irreps of $ISL(2, \mathbb{C})$. The states are

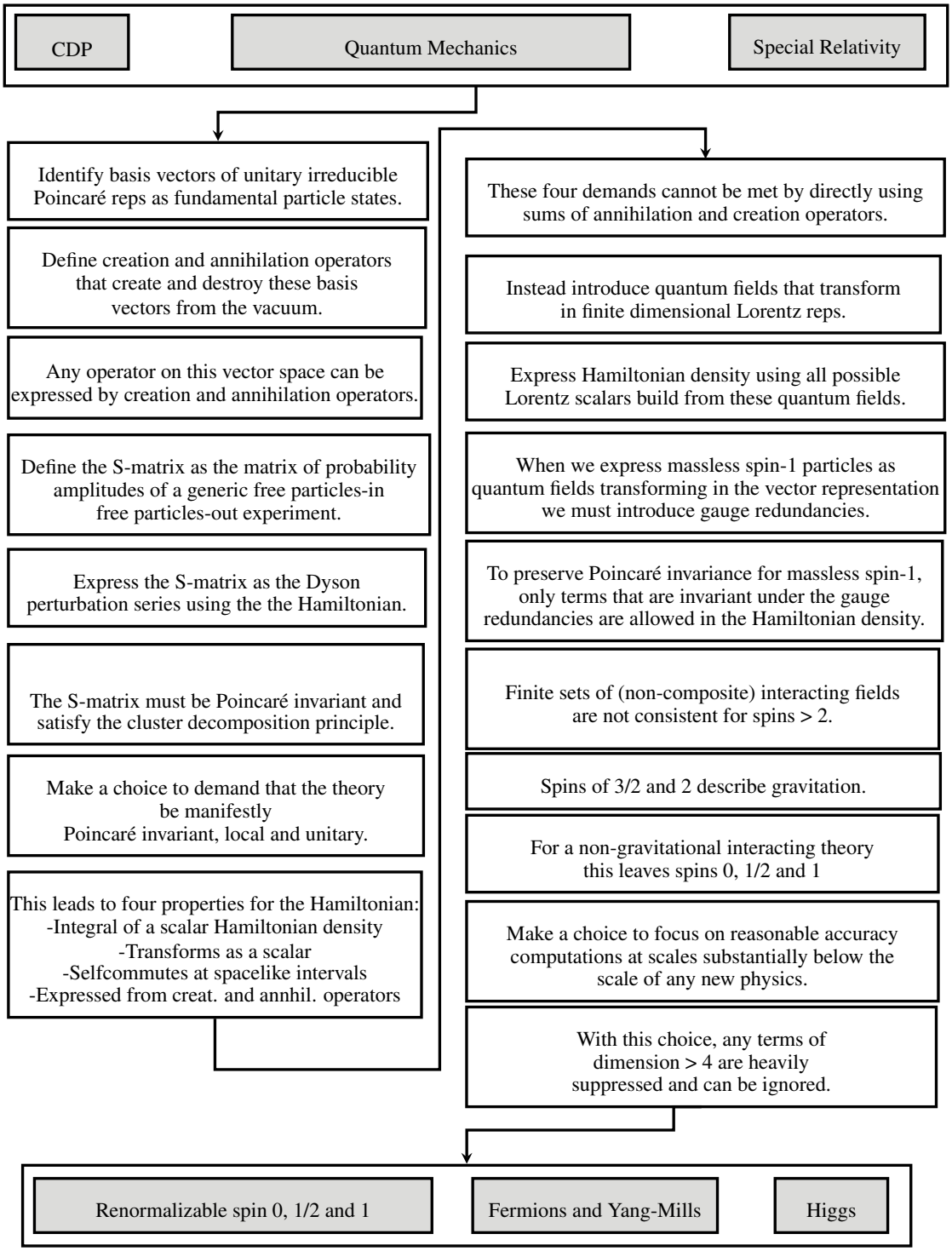


Figure 3.4: The diagram shows the long argument to arrive at QFT from three principles (CDP + Quantum Mechanics + Special Relativity). Abbreviations: CDP: Cluster decomposition principle

labeled by mass (m), spin (s), helicity (h), momentum (p) and auxiliary quantum numbers (q_i). We write them as

$$|m, s, h, p, q_i\rangle \,. \tag{3.7.1}$$

We call these states free particle states but without assuming any of the ontological implications of traditional semantics.

The next step in our deduction is to set up a framework for the dynamics of scattering processes. In scattering processes the amplitude of the transition between a set of incomming free particle states and a set of outgoing free particle states is calculated. The invariance of the theory under Poincaré transformations implies the existence of a time-translation operator $P^0 = H$. Given these assumptions, the S-matrix elements representing the quantum probability amplitudes, are given by Dyson's formula [44].

Conjecture 3.7.2 — S-matrix conjecture. Implementing QM and SR implies the existence of a time translation operator $P^0 = H$ which determines the S-matrix of quantum probability amplitudes by the formula

$$S = \int ... \int dt_1 ... dt_n T\{H(t_1)...H(t_n)\}. \tag{3.7.2}$$

The whole structure we have built so far, is based on the assumption that isometries are maps between equivalent physical settings. Therefore, for our structure to remain consistent, these S-matrix amplitudes must be invariant under Poincare transformations. In addition we assume that there are fundamental limitations restricting the possible correlations between two spatially separated physical processes. There are various possible formulations of this idea of varying strength and generality. We will use the formulation called the cluster decomposition principle (CDP).

Definition 3.7.3 — Cluster decomposition principle. The cluster decomposition principle (CDP) is the principle that all experiments that are sufficiently spatially separated are uncorrelated. We add the condition that there must not exist any prior correlations between the two experiments. For the S-matrix specifically the CDP means that widely separated scattering processes are uncorrelated, and thus can be represented by a factorized S-matrix, where each factor represents the amplitude of the individual localized scattering experiment. (Note: This definition is not intended to give a complete operational specification. To be more precise we should define far away, define the nature of the involved reference systems, and define what no prior correlations means.)

There is no unique solution to the demands of S-matrix Poincaré invariance and respecting the CPD. Indeed it is an important point that there are solutions to this requirement that look very different from QFT [47]. However, if we stipulate that the theory should be manifestly local, Poincaré invariant, and unitary, there are probably no solutions that are substantially different from what we will describe [44][64]. A rigorous proof of this is not imminent, so we will list it as two strongly supported conjectures [44][64].

Conjecture 3.7.4 — Hamiltonian density conjecture. Poincare invariance of the S-matrix coupled with the demand for manifest locality and unitarity leads to the following set of demands.

1. The time translation operator can be expressed as an integral of a scalar density over space

$$H = \int \mathcal{H}(\vec{x}, t) d^3x. \tag{3.7.3}$$

2. The density transforms under Poincaré transformations $U(\Lambda, a)$ as

$$U(\Lambda, a)\mathcal{H}(x)U(\Lambda, a) = D(\Lambda)\mathcal{H}(\Lambda x + a). \tag{3.7.4}$$

3. The density Poisson commutes with itself at spacelike separations

$$\{\mathcal{H}(x), \mathcal{H}(x')\} = 0 \quad for \quad (x - x')^2 \geq 0. \tag{3.7.5}$$

Conjecture 3.7.5 — CDP-fulfilment conjecture. The demand that the S-matrix should obey the cluster decomposition principle leads to the following requirement [44].

1. The time-translation operator (the Hamiltonian) must be expressed as a sum of creation and annihilation operators with only a single delta function.

The creation and annihilation operators transform under Poincare transformations in a manner that depends on the momentum [44]. This makes it impossible to form something transforming as a Lorentz scalar using these operators directly. To satisfy the combined demands of conjecture 3.7.4 and 3.7.5 one needs to introduce operator fields. These operator fields are formed from annihilation operators or creation operators multiplied by component functions that under Poincare transformation transform in a finite dimensional representations of the Lorentz group. We call these fields annihilation fields and creation fields repspectively. The annihilation fields and creation fields (which are made from just annihilation operators or just creation operators) does not satisfy property 3 in conjecture 3.7.4. To satisfy property 3 in conjecture 3.7.4 we need to create linear combination of annihilation fields and creation fields [44]. If some particles carry conserved charges (quantum numbers) this implies, by the requirements of Poincaré invariance, the existence of corresponding anti-particles with opposite charges (quantum numbers) [44]. The Hamiltonian now consist of any combination of such fields that produces a scalar density. The Lagrangian approach is better suited for explicit construction of such scalar densities because it is easier to ensure that specific symmetries are realized in this approach. After quite a long argument we have thus deduced the most general form the action can take based on our stated assumptions. The next sections will add a few extra details to this picture.

3.8 Gauge invariance from Lorentz invariance

> The necessity of gauge invariance in perturbative quantum field theory stems from the fact that one and the same massless particle, thought of as a representation of the spacetime isometry group, in general admits (infinitely) many implementations in terms of quantum fields sitting in different Lorentz tensors obeying respective free equations of motion.
>
> Bekaert et al.

The gauge principle has been heralded as one of the great discoveries of the 20th century [65]. Undoubtedly the success of particle physics and the standard model hinges upon the power of the gauge principle. However this does not answer where the gauge principle comes from, and whether it is a fundamental physical principle.

It is common to present the gauge principle as a distinct and fundamental principle of nature. In our view this distorts the nature of gauge symmetries. We prefer instead to approach gauge invariance from the perspective of being a direct consequence of realizing massless spin-1 representation as vector fields [66][44][51]. Since spin-1 representations have two degrees of freedom and vectors fields have four, this mapping is impossible without introducing redundancies which again leads

to gauge invariance. Thus gauge invariance is a consequence of demanding explicit locality and Lorentz invariance a theory that contains massless spin-1 fields.

Let us now see this in slightly more detail. From the representation theory of $ISL(2\mathbb{C})$ we know that massless spin-1 particles have 2 polarization states (see table 3.1). However in the field representation (in four dimensions) these particles are represented as vectors fields $A_\mu = \epsilon_\mu e^{ipx}$, where ϵ_{mu} is a set of polarization vectors. There is only one Lorentz invariant constraint that can be imposed on these polarization vectors, namely

$$\epsilon^\mu p_{mu} = 0. \tag{3.8.1}$$

This means that we must consider equivalence classes of polarization vectors related by

$$\epsilon_\mu \sim \epsilon_\mu + \alpha p_{mu}. \tag{3.8.2}$$

This again implies for the vector field themselves that we must consider equivalence classes of vectors related by

$$A_\mu \sim A_\mu + \partial\Lambda. \tag{3.8.3}$$

A similar argument exist for massless spin-1 particles with conserved charges (quantum numbers) [44]. In this case the redundancies lead to non-Abelian gauge symmetries .

3.9 Renormalization and RGE

Renormalization has been a historically important criteria in narrowing down possible terms of the Lagrangians. Later theoretical insights have modified our view on what renormalization means. We need to establish what the proper modern interpretation of this concept is.

Uncontrollable infinities in the loop diagrams of QED was a major headache for the early QFT practitioners [67]. Later Dyson, Feynman and Schwinger showed how these QED infinities could be eliminated by the process of renormalization [67]. Further analysis showed that the concept of a renormalizable theory could be extended based on on the QED example [68][69]. The renormalizability criteria can be defined at the term level. The conclusion can be put very simply, in four dimension only terms of dimension four or less are renormalizable. This was used as a fundamental and guiding principle for many years. All proposed interactions were required to be renormalizable.

Today, after the introduction of the concept of effect field theories (EFT) by Wilson, Kadanoff, Gell-Mann, Low and others, the perspective on renormalizability has changed a lot [70, 71][67]. In an EFT it is no longer natural to demand renormalizability. Instead every term compatible with the symmetries of the theory should be included. This leads to an infinite set of (non-renormalizable) terms. However, as long as the upper limit of the validity of the EFT is substantially beyond the scale we want to calculate, all the non-renormalizable will be effectively suppressed leaving only the renormalizable terms to make an actual measurable contribution to the calculations of cross-sections and life-times [44][64]. In essence the renormalizable terms are still singled out, but for a quite different reason that was originally thought to apply. This explains the success of renormalizable theories, even though the reasons for they original selection was due to an improper understanding of renormalization phenomena [44][64]. This insight is no longer new but it has still not penetrated fully in the way these things are discussed. We therefore spell out some rather straight-forward implications. In light of EFT paradigm it does not make sense to label the standard model as renormalizable. Any EFT is well approximated at low energies by a renormalizable theory. Which makes the statement "the SM is renormalizable" true, but true in the sense of a tautology. Of course it could be the case that the SM is valid at all energies (we disregard Landau

points), and that the coefficients of all non-renormalizable terms are so close to zero as to never be of any practical relevance, but this is not the EFT way of seeing things. Extra terms are not empirically excluded, they are only restricted to be sufficiently small so as to be undetectable at current experiments. Thus with regards to the SM considered as a theory in its own right, there are no empirical and theoretical foundation for claiming that it represents more than the common low energy renormalizable approximation of an infinite set of non-renormalizable theories.

The best perspective on renormalization comes from the renormalization group equation (RGE), and the flows generated by these in theory space, the RGE flows. RGE flows describe how theories (with an infinite set of terms) flow between two theory space endpoints, one in the low-energy limit (so called infrared CFT or IR-CFT) and one in the high-energy limit (so called ultraviolet CFT or UV-CFT) (see for instance [72] for more details). In this view the SM can be seen as a perturbation of a IR-CFT by some relevant operators.

There is a rephrasing of what we have already said that we find worthwhile to mention. EFT can be considered to be a method for coarse-graining. In the EFT approach one integrates out the irrelevant high-energy degrees of freedom to produces a simpler approximation that is fully adequate for the energy or resolution that is considered. Renormalizable theories can be classified as universality classes under such coarse-graining. A large class of theories will give rise to the same renormalizable theory under coarse-graining. This is often emphasized in the context of critical phenomena. Many systems, with a wide variety of dynamics, exhibit a universal behavior at critical points. However, the phenomena of universal behaviors also occurs for many other systems, including systems of elementary particles. The phenomena of universality classes can be visualized as sets of RGE flows eventually flowing to the same theory in the IR.

3.10 Symmetry breaking with scalar fields

We have already stated that gauge theory is a consequence of manifest locality and Poincaré invariance applied to massless spin-1 particles. However some bosons believed to be associated to gauge redundancies are known by experiment to be massive rather than massless. In addition, chiral fermions do not permit a Dirac mass term. Thus there is a need for a mechanism to allow combining gauge symmetry and massive gauge bosons, as well as enabling mass terms for chiral fermions. The mechanism used to achieve this in the standard model is called spontaneous symmetry breaking [43][51]. This occurs by introducing an extra spin-0 field with a ground state which is not invariant under the gauge symmetry one wants to break. We will not go into any detail about this which is well described in any recent textbook [43][51]. This point still needs to brought up because of the following. Introducing scalar fields is not introducing something new from the perspective on QFT that we have presented. Doing so does not require any new fundamental postulates, and should be regarded as one of many open possibilities in the theory that we have constructed. Quite often the Higgs mechanism is tagged as an unsightly and ad-hoc addition to the standard model. This may be true in a historical sense, but from a foundational viewpoint there is no reason to assume that scalars are not present. On the contrary, the presence of scalars is a perfectly natural occurrence from the representation theoretic point of view, and a symmetry breaking potential is one of the renormalizable possibilities for the potential terms.

3.11 The standard model

In the context of this chapter we have started out by assuming that the world is described quantum theory operating in a fixed background Minkowski space, and from this we have deduced the naturally occurring possibility of a world being described by a low energy EFT containing massless Yang-Mills gauge bosons, massless chiral fermions and mass-giving symmetry breaking scalars.

While we haven't concluded that this assumed world must by necessity contain massless particles, we have found that if they are present they must then appear in this manner.

The standard model is one specific version of a low energy EFT with several massless particles. To define the standard model starting from the general EFT model, we need to specify gauge groups, the fermion representations, the scalar representations, as well as some 20 or more parameters detailing scalar potential, interaction strengths, mass-Yukawa couplings and particle mixing. In the framework we have sketched, all of these parameters are completely arbitrary. The only partial guide is anomaly cancellation, but this principle does not bring us close to the standard model unless almost all other parameters are already given. In later chapters we look at theories that try to explain all or some of the choices necessary to get to the standard model.

3.12 Notes

History and extra references

To place some of the developments we have presented in this chapter in a more traditional perspective we present av short history of relevant developments. Early breakthroughs towards a quantum understanding were made by Planck, Einstein, de Broglie, Bohr and Sommerfield in the years 1900-1923. Quantum theory was invented by Heisenberg in 1925 [73]. Heisenberg's theory was first applied to particles. The first quantization of a field was in 1926 by Born, Heisenberg and Jordan [74].

Regarding the use of space isometries we follow the analysis of Wigner [48], Bargman [49], Mackey[59][60], Varadarajan[61], Weinberg [62], [63] [44] [64], Wald [45] and Arkani-Hamed (see almost any lecture of his).

The Abelian $U(1)$ symmetry of electromagnetism was the first example of a gauge theory. Weyl was an early proponent of the importance of the gauge symmetry in fundamental physics. But for many years this view was ignored. Gauge theory was first generalized beyond $U(1)$ by Yang and Mills [75]. They introduced non-Abelian $SU(2)$ gauge symmetries in 1954. In the initial version of the theory the corresponding bosons had to be massless so no immediate application of this theory was found, since the only known long-range microscopic interaction was the electromagnetic one. Eventually two methods (or mechanisms) were found that allowed introducing massless bosons without associated long range interactions. The first was to give the bosons a "dynamic" (or emergent) mass due to a scalar field with a non gauge invariant vacuum state. This is known as the Higg's mechanism. The other example was in the context of a confining force, like the $SU(3)$ of the strong interactions.

We have emphasized he fact that the two helicity states of the massless spin-1 bosons doesn't match up with the degrees of freedom of a vector field. Deriving both the Abelian and non-Abelian gauge theory from this idea was done by Weinberg [62] [63] [44] [64].

From Dyson's proof of the renormalizability of quantum electro dynamics in 1949, renormalizations has gone through a massive conceptual development. The new understanding of renomarlization was introduced primarity by Wilson[70, 71] in the context of statistical mechanics and critical phenomena.

4. Background free theories - GR

> The axiomatic formulation of general relativity (or gravitational theories in general) seems to resemble the myth of the Holy Grail. Serious attempts have been made to find it and everybody seems to be interested in it, but nobody actually knows where to look for it.
>
> Sotiriou et al.

Summary

The equivalence principle leads us to seek a background free description of gravity in terms of dynamic geometry. The associated symmetry of a background free description is the group of diffeomorphisms. General relativity (GR) is the unique lowest order approximation of such a theory. Classical systems like GR defined on a dynamic spacetime derives its basic structure from implementing the diffeomorphism group of spacetime in its algebra. The redundancy in this description leads to a theory defined solely by constraints. The Hamiltonian of such a system does not define a time parameter, and all Dirac observables are constant. Dynamics cannot be expressed as time-evolution and must instead be defined by using relational methods. Relational methods express dynamics as the correlated changes in two or more families of observables.

In our analysis of QFT it was the combination of the three-part picture of quantum physics plus the CDP along with the isometries of Minkowski space that dictated the content of the theory. In this chapter we discuss combining the three-part picture of classical physics with the diffeomorphism invariance of a background free theory of gravity.

4.1 Introduction

Gravitation is a force with unique properties. Gravity, unlike the other forces, affects every object equally, independent of charges and composition [76]. Einstein made this property the cornerstone of his general theory of relativity (GR). He formalized this property in the equivalence principle [56]. The equivalence principle leads to the possibility of a new and different description of gravity. Due to the universality of gravity the effects of this force can be described simply as an aspect of physical geometry. This seemingly uneventful re-coding of gravity into geometry, completely resets our understanding of gravitational interactions. Through this re-coding gravity becomes the master and the creator of the geometric, the chronometric and the inertial structure of spacetime. Understanding gravity as spacetime geometry makes spacetime without gravity loose all spatial and temporal properties. It becomes an unformed emptiness devoid of recognizable geometric and chronometric structure [77]. The dynamics of gravity dictates how this void is endowed with geometric structure. Gravity in essence becomes spacetime. Any concept of background structure becomes unobservable, and therefore physically meaningless. GR is a a background independent, or perhaps even more accurately, a background free theory. The symmetries of the theory is given by symmetries of the void. It is the group of diffeomorphisms. If we accept the picture of gravity as being given solely by (metric) geometry, and we seek to find the associated background free diffeomorphism invariant theory, we are lead to a unique sequence of terms for the gravitational Lagrangian. If we truncate the sequence at two terms, we we find the unique low-energy Einstein-Hilbert action determined by two real parameters [78] (see figure 4.2 for a diagram version of the argument from universality to the EH action).

4.2 Equivalence of gravitational and inertial forces

> Most miraculous is the connection between the gravitation of matter and its inertia. We see that any of two quantities of matter with the same inertia will exert the same gravitational effect on each other, irrespective of the substances they are made of. And in reality we do have two properties before us, two most fundamental properties of matter, which must be thought of as being completely independent of each other, but which, in our experience, and only in our experience, appear to be exactly equal.
>
> Hertz, 1884

The universality of the effects of gravity was established by Galileo Galilei's observations [79], that, contrary to Aristoteles' views, all objects free fall with the same acceleration. There are some indication that Philophonius, already in the 6th century, seem to have found universality in the rate of fall, but it was Galileo that firmly established this fact in his studies. In the end, it was Newton that could give an explanation of this, by introducing his law of gravity. Newton introduced the concepts of the inertial mass and the gravitational mass, but set them to be equal (or at least proportional). Therefore, in Newtonian theory, gravity produces the same rate of acceleration for all objects, and Galileo's universality of free fall became Newton's universal effect of gravity. It is of course not necessary that inertial mass and the gravitational mass be numerically identical for this effect, just that they have a constant ratio to each other, but we will henceforth assume that units have been re-scaled to give the identity. Note that we have not introduced the concepts of active and passive gravitational mass, as these must be equal (or proportional) by Newton's third law.

It is a direct consequence of Newton's system that a uniformly accelerated reference frame produces the same results (as concerns their classical mechanical behavior) as does a suitably uniform gravitational field. Or, to put it differently, that the gravitational mass is equal to the inertial mass for all objects.

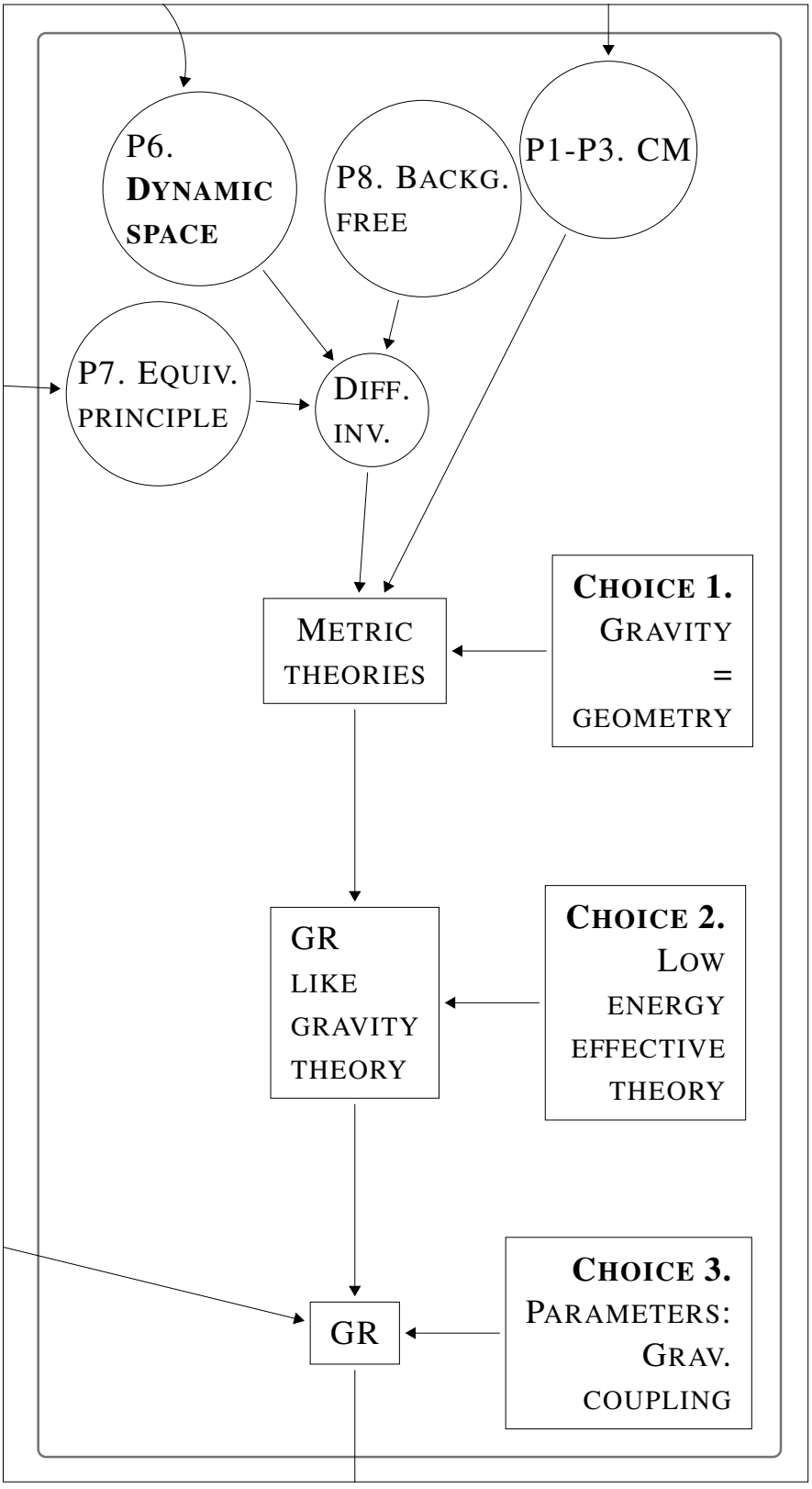


Figure 4.1: Combining dynamic space with being background free leads to metric theories of gravity (choice 1). Focusing on the low energy effective theory leads to the general form of the EH-action (choice 2). Abbreviations: see figure 1.

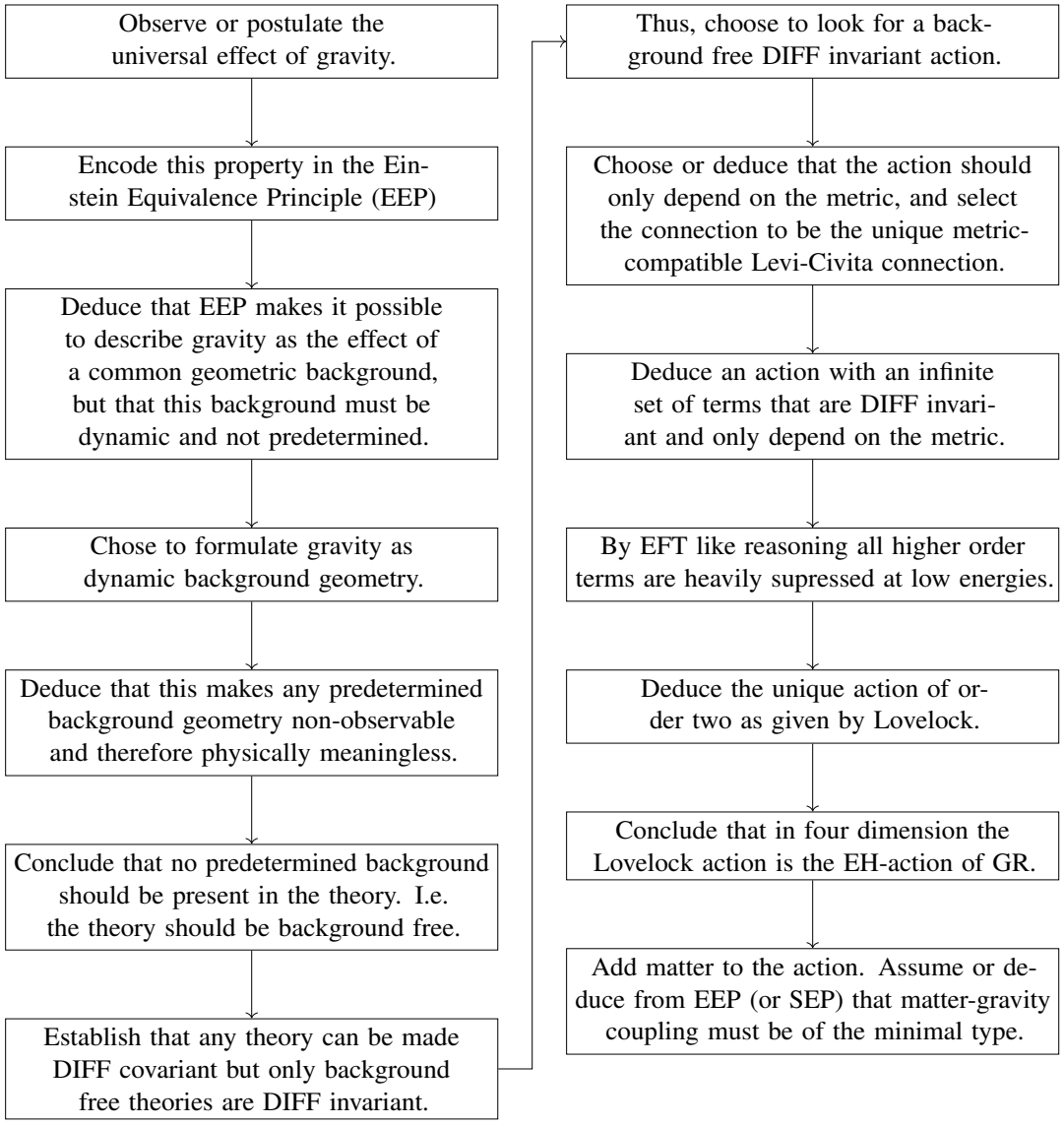


Figure 4.2: Deriving the EH action of GR from the universality of gravity. This diagram is not meant to suggest that this is a rigorous derivation. Note in particular that several choices are made in the process. Abbreviations: DIFF = diffeomorphism, GR = general relativity, EH = Einstein-Hilbert, EFT = effective field theory, SEP = strong equivalence principle.

4.3 Equivalence principles

> ...identifying gravity with spacetime geometry is not a compulsory step, but follows from adopting a view in which all universal features are ascribed to geometry.
>
> Casola and Liberati

Having established the kind of equivalence principle found in Newtonian gravity, which applies to material particles moving freely or under mechanical forces, one can go on to ask how more general physical processes are affected by gravity. This line of inquiry gives rise to a multitude of extended equivalence principles. It has become clear in the later decades that *the* equivalence principle, as the term has been used historically, is not a well defined principle [80]. Indeed, when seeking a more precise formulation, it transpires that many different versions exists. We will now

give a statement of some of the most useful versions [81] [82].

The major difference between the various formulations, is what is postulated to be equivalent (masses, mechanics, non-gravitational physics, all physics), under what conditions (locally, in the absence of self-gravitation etc.). We start with the version stipulated by Newton.

Definition 4.3.1 — Newtons's equivalence principle (NEP). In the Newtonian limit, the inertial and gravitational masses of a body are equal.

Newton's version depends quite intimately on being in the relevant domain where Newtonian concepts apply. Next in line is a formulation cementing the universality of free fall as described by Galileo. We first state an auxiliary definition.

Definition 4.3.2 — Test particle and test physics. A test particle in this context signifies a particle that has negligible back-reaction on the gravitational environment. Test physics signifies a physical system that has negligible back-reaction on the gravitational environment. (Note: Test particles/physics can have self-gravity or internal gravitational energy etc. The definition only states that the effect on the field that it is surrounded by is negligible.)

Definition 4.3.3 — Weak equivalence principle (WEP). Test particles with negligible self-gravity behave, in a gravitational field, independently of their properties.

We do not specify an operational definition of neither this nor the criteria of negligible self-gravity. The next principle removes the condition of negligible self-gravity but adds vacuum as a condition.

Definition 4.3.4 — Gravitational weak equivalence principle (GWEP). Test particles behave, in a gravitational field and in vacuum, independently of their properties.

The previous formulation have only stated equivalence with respect to the movement of mechanical particles under the sole influence of gravity. The next principle goes on to include all non-gravitational physics. This includes things like electrodynamics, thermodynamics and fluid dynamics.

Definition 4.3.5 — Einsteins's equivalence principle (EEP). Fundamental non-gravitational test physics is not affected, locally and at any point of spacetime, by the presence of a gravitational field.

The strong principle below is considerably more inclusive than the other ones listed above. It is an extension of the EEP so as to also include gravitational physics (and thereby all physics) in the list of equivalent processes. Again the test prefix limits us to systems that have negligible effect on the test environment.

Definition 4.3.6 — Strong equivalence principle (SEP). All fundamental test physics (including gravitational physics) is not affected, locally, by the presence of a gravitational field.

Note that in both EEP and SEP the conclusions are limited to what can be observed locally. Locally must interpreted as something like, for any strength of a gravitational field there is a corresponding smallness of system extension for which a system or an effect can be called local (see 4.10 for some opposing views). In the mathematical sense this is the infinitesimal limit .

4.4 The Equivalence Principle makes a geometric formulation possible

Having clarified some concepts and definitions with regards to equivalence principles, we now approach the essence of this line of reasoning. What are the roles played by the various forms of equivalence principles in suggesting or dictating gravitational theory?

The WEP specifies that the paths of movements through spacetime is independent of composition and depends only on the gravitational fields. Particles with close to the same velocity will follow close but slightly different trajectories. These trajectories will not intersect. Together the whole class of possible velocities define a complete set of possible trajectories from a point. These trajectories together cover or span the light cone at that point. This means there is a set of preferred paths through spacetime that are associated with a given configuration of the gravitational field. It seems also quite certain that these lines must become Newton's straight lines of motion in the limit of no gravitational interactions. This suggest that these lines can be seen as a generalization of straight lines whose exact path depend on the gravitational field. The mathematics behind such a concept can be found in the definition of parallel transport by some affine connection. Note that this approach is an idea for a description that might be useful, it is not forced upon us by any fact. The important thing about the equivalence principle is that it tell us that gravity couples universally. This allows for the possibility of a purely geometric description. However it does not disallow other descriptions of gravity (teleparalle gravity is one such alternative, see e.g.[83][84][85] for an introduction to teleparallel gravity).

4.5 Diffeomorphism

Diffeomorphisms are the morphisms of the category of smooth manifolds. We review some of their properties and their relation to coordinate transformation. The following section will put diffeomorphism in their proper perspective within the discussion of the foundations of GR.

A diffeomorphism is a bijection $h\colon M \to N$ between two manifolds M and N that preserves the topology as well as the differentiable structure. Since the map h preserves all aspects that we are presently interested in, and is a bijection, we can equally well consider diffeomorphism $h\colon M \to M$ on a single manifold M. From a general map $h\colon M \to M$ we get the associated maps called the pullback and the pushforward. The pushforward defines a map

$$h_*\colon T_p \to T_{h(p)} \tag{4.5.1}$$

between corresponding tangent spaces. In the same manner the pullback defines a map

$$h^*\colon T^*_{h(p)} \to T^*_p \tag{4.5.2}$$

between corresponding cotangent spaces. For a general map h the pullback and the pushforward are not invertible. When h is a diffeomorphism both of these constructions are invertible. Thus a diffeomorphism allows us to define the pushforward and pullback of any tensor. Diffeomorphisms are closely related to coordinate transformations. If we let (U_i, ϕ_i) be a chart on M and h is a diffeomorphism then we can associate to h the chart $(U'_i = h(U_i), \phi'_i = \phi_i \circ h^{-1})$. If $U'_i \cap U_i \neq \emptyset$ there is an associated coordinate transformation defined by

$$\phi'_i \circ \phi_i^{-1}. \tag{4.5.3}$$

Likewise, two overlapping charts define a diffeomorphism on the overlapping domain. The transformation of tensors under pullbacks and pushforwards are the same as those for the effective coordinate change. Note the slight complication that a chart usually has a limited domain $U_i \subset M$ while a diffeomorphism is defined for all of M. The main point to make here is that (modulo the

domain issues) there is correspondence between the set of diffeomorphisms and the corresponding set of coordinate transformations. Each of them always defines the other. On occasion words like active diffeomorphisms are used, and this usually taken to mean something else than the ordinary active/passive viewpoints that always exists for a set of transformations. We shall use a different terminology for what is usually meant by physicists by active diffeomorphisms. We get to this in the next section. In our terminology the traditional active passive viewpoints on transformations are also present for coordinate transformations and diffeomorphisms. These are just two different viewpoints on the same process with no significant distiction between them (see e.g. [86] [87]).

4.6 Background independence

Background independence is a concept that many researchers think of when trying to describe the core features of GR. We lean towards the word background free, but the intended meaning is the same. The word "background" in this setting means something like a predetermined arena for dynamics. Loosely speaking it is something predetermined and non-dynamical that plays an essential role in the equations of a physical system. This might sound quite clear cut, but in practice defining such distinctions as dynamical vs non-dynamical is quite involved.

Definition 4.6.1 — Non-dynamical field. A field g is called a non-dynamical or a background field if it is prescribed and not obtained by solving an equation.

Definition 4.6.2 — Dynamical field. A field that is found by solving an equation is called a dynamical field .

As an operational framework the definitions of dynamical and non-dynamical are not discriminating enough. It is often quite easy to set up an equation for a prescribed field, and thereby according to our definitions, change the field from non-dynamical to dynamical. More sophisticated definitions can be made but we do not pursue this issue further. In the following we do assume that a "proper" distinction can in principle be made.

Let us now set up the stage for discussing classical theories based on manifolds in some generality (see [88] or [89]). Let M be a smooth manifold. We assume that the objects of our theory are the manifold and the fields (or sections with trivializations) on the manifold (or bundle). Fields are represented as sets of maps $D_i \colon M \to V$ and $B_i \colon M \to V$ where V is some vector space. We denote the set of maps by D and B, where the letters are chosen to mean D for dynamic and B for background. (We leave out the examples using sections as this does not add anything pertinent to the present discussion.) We denote the theory by $\langle M, D, B\rangle$, and the equations of motion are given by $\mathcal{F}[D, B] = 0$. Let $\mathrm{Diff}(M)$ be the group of all diffeomorphism of M, and let G be some subset $G \subseteq \mathrm{Diff}(M)$. Let $\phi \colon M \to M$ be an arbitrary diffeomorphism $\phi \in G$. We represent the action of ϕ on the fields D and B symbolically by $\phi \cdot D$ and $\phi \cdot B$ (note that this is just notation and does not tell us what the action is).

Let us now make some definitions using this setup. First we define a theory where solutions to equations of motion are mapped to new solutions of the equations of motion when the diffeomorphism is applied both to the dynamics fields and to the background fields.

Definition 4.6.3 — Covariant theory. Let $\langle M, D, B\rangle$ be a theory with equations of motions $\mathcal{F}[D, B] = 0$, and let $\phi \in G \subseteq \mathrm{Diff}(M)$. A G covariant theory means that

$$\mathcal{F}[D, B] = 0 \iff \mathcal{F}[\phi \cdot D, \phi \cdot B] = 0 \tag{4.6.1}$$

Next we define a theory where solutions to equations of motion are mapped to new solutions of the equations of motion when the diffeomorphism is applied just to the dynamics fields and *not* to the background fields.

Definition 4.6.4 — Invariant theory. Let $\langle M, D, B\rangle$ be a theory with equations of motions $\mathcal{F}[D, B] = 0$, and let $\phi \in G \subseteq \mathrm{Diff}(M)$. A G invariant theory means that

$$\mathcal{F}[D, B] = 0 \iff \mathcal{F}[\phi \cdot D, B] = 0 \tag{4.6.2}$$

Definition 4.6.5 — Diffeomorphism invariant theory. A G invariant theory is said to be diffeomorphism invariant theory if $G =$Diff(M)

Definition 4.6.6 — Background free. A theory $\langle M, D, B\rangle$ is background free (also known as background independent) if $B = \emptyset$.

We can interpret this whole set of definitions as suggesting that a theory should be background free to be diffeomorphism invariant. The suggested definitions give a pretty good representation of what we intend to say when we say that GR is background free. We note again that this is equivalent to what some authors call "active diffeomorphisms"[77]. "Active diffeomorphism" seems to us to be an unsuitable terminology as it has too many unwanted connotations.

4.7 A diffeomorphism invariant action

In the preceeding sections we have established that a line of reasoning starting from the universal properties of gravity, via the equivalence principle to the understanding of gravity as geometry leads us to seek a diffeomorphism invariant background free action (see figure 4.2). Any properly definable scalar, when integrated with the invariant four-dimensional volume form $\sqrt{-g}\, d^4x$, will give us a diffeomorphism invariant action. The symbol g denotes the determinant of the metric tensor.(Note: A proper scalar is either an arbitrary fundamental (not derived) scalar field, or some derived scalar field defined by inner products, contractions or other invariant constructions. A scalar defined as being the n-th component of a vector would be an example of an improper scalar.) There are naturally many such potential (proper) scalar actions, and to proceed further we need to use stronger assumptions. In the thoughts leading up to this point we have identified the effects of gravity with modifications of metric geometry. Phrasing this slightly stronger we can suggest the principle that the metric and the connection are the *only* dynamical variable representing geometry, and thus by implication the *only* variables used to represent gravity. If we further assume that the connection is given as the unique torsion free metric compatible connection known as the Levi-Civita connection (see e.g. [90], we are sufficiently positioned to derive the action. We now seek a set of scalars that depend only on the metric and its (Levi-Civita) derivatives. This singles out an infinite sequence of terms. The first few terms of this sequence are given by [91][92]

$$L = \Lambda + \frac{M_p^2}{2}R + aR_{\mu\nu}R^{\mu\nu} + bR^2 + dR_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho} + \dots \tag{4.7.1}$$

Here R is the Ricci scalar, $R^{\mu\nu}$ is the Ricci tersonr, $R^{\mu\nu\lambda\rho}$ is the Riemann tensor, and Λ is the cosmological constant. The symbols a, b, d are arbitrary real constants, and M_p is the Planck mass. For each term in the sequence there is an associated real parameter. If we truncate the series after two terms, and introduce the conventional symbols for the two parameters, we get the

Einstein-Hilbert action [91][92]

$$S = \frac{1}{2\kappa} \int \sqrt{-g}(R - 2\Lambda). \tag{4.7.2}$$

Here κ denotes $\kappa = 8\pi G c^{-4}$, or just $\kappa = 8\pi G$ in natural units. This establishes a unique low-energy action (modulo two arbitrary parameters) for pure gravity. To include matter we first observe that the ordinary Lagrangian densities of spin-0, spin-1/2 and spin-1 fields is one possible proper scalar. Just adding this matter action to the Einstein-Hilbert action in the form

$$S = \frac{1}{2\kappa} \int \sqrt{-g}(R - 2\Lambda + L_{matter}), \tag{4.7.3}$$

gives us the EH-action together with what is called minimal matter-gravity coupling. It is clearly the simplest action possible that includes all the known relevant degrees of freedom. One can add terms involving the product of the matter action with any of the terms from pure gravity action. This produces direct curvature-matter couplings and is called non-minimal coupling. Non-minimal couplings are popular candidates in some theories that try to explain dark matter and dark energy by gravitational means. With the exception of a brief mention in chapter 6 we will not consider such terms further.

4.8 Dynamics of background free theories

The profound nature and importance of a background free formulation of GR is not always emphasized sufficiently. The most well known solutions of GR are derived by using symmetry principles (see e.g. [45] or [76]). These classical solutions to some extent hide the underlying changes imposed by GR on the core dynamical questions. In chapter 8 we will explore the Hamiltonian formulation of GR in more detail. For completeness we anticipate some of the results here (See chapter 8 for longer explanations and more references). The Hamiltonian formulation of GR is a fully constrained theory where the Hamiltonian is given as a linear sum of constraints. Classical Dirac observables must commute with all the constraints and must therefore also commute with the Hamiltonian. This means that the Hamiltonian flow defined by the Hamiltonian itself will not induce any changes in the Dirac observables. The whole theory appears to be "frozen" and without any dynamic evolution. This is actually what one would expect, since there is no (global) time directions available to define any form of time evolution. All "time evolutions" in a diffeomorphism invariant theory is only a coordinate transformation and does not correspond to actual physical dynamics. The dynamics of the theory must be implemented by asking the kind of questions that we alluded to in chapter 1. The coevolution of correlated values of two or more (sequences of) observables is the proper way to extract the dynamical information. This affects our general three-part setup consisting of the algebra of observables, the states of algebra and the one-parameter group of algebra automorphisms. The first two components are left unchanged, but dynamics can no longer be represented by one-parameter groups of automorphisms. There is no longer any such parameter available. Dynamics must now be extracted from the physical states by using relational methods [93]. This is a very natural generalization of the three part structure we have relied on in the previous two chapters [93][77]. It transforms and extends our view of dynamics. Dynamics is not limited to describing the change of various physical properties with time, it is more properly seens as describing the correlated changes of physical properties with respect to each other. In many cases we can recover a proper definition of time from this more evolved perspective, but the innocence of relying on time evolution as something fundamental is gone [93][77].

4.9 The spin-2 approach to gravity

> GR is the unique low energy theory for interacting massless spin 2 particles. Similarly Yang-Mills theory is the unique low energy theory for interacting massless spin 1 particles.
>
> Arkani-Hamed

It is remarkable that there exists a path to the equivalence principle and the EH-action that is completely different from the one presented in the previous sections. This spin-2 path is what we now turn our attention to (see figure 4.3 for an overview). As we have seen in chapter 3, combining Minkowski space with quantum theory leads to the considerations of irreducible unitary representations of the Poincaré group. These kinds of representations tells us a lot about possible physical theories. In chapter 3 we also analyzed the origin of so called gauge forces, and found them to be a direct consequence of the existence of interacting massless spin-1 fields. When massless spin-2 fields are analyzed in the same manner one finds similar patterns. These equations must also implement gauge redundant elements, and to be properly invariant this requires an interaction that has universal couplings to energy-momentum. That is, the interactions of a massless spin-2 field must obey a linearized version of the equivalence principle [62][63]. This can be seen as the explanation of the validity of the equivalence principle, founded on the principles of Poincaré representations 3 [62][63].

Field equations from spin-2 consistency

We argued that a linearized form of the equivalence principle can be derived from the properties of massless spin-2 fields. From another direction, we can also derive the equations for spin-2 fields as the linearized version of GR. The importance of the spin-2 approach does not stop there, a sort of reverse of the process of linearizing GR is also possible. One can apply the principle of equivalence to the spin-2 field itself, and this leads one to derive Einstein's field equations starting from the linear equations of spin-2 fields [94] [95] [96]. In this derivation there is no specific limit on the number of derivatives in the resulting action. We recover GR with a cosmological constant as the low energy effective theory. Higher order derivatives are not excluded but are irrelevant at lower energies since their contributions are heavily suppressed.

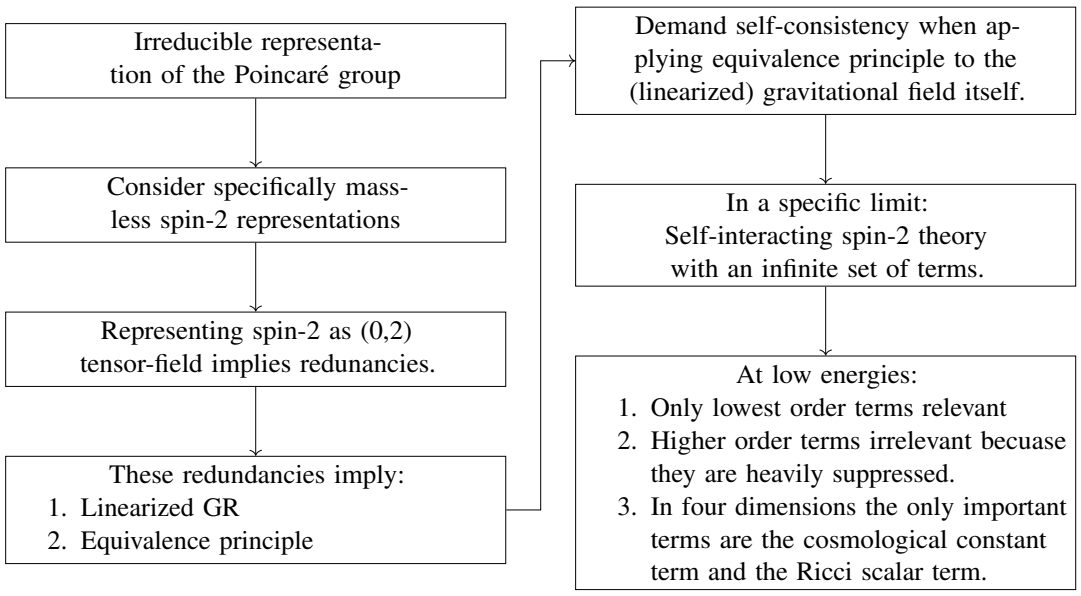


Figure 4.3: The spin-2 approach to gravity.

Note that in the process applying the equivalence principle to the spin-2 field itself, the flat background manifold is made completely invisible and non-physical. The concept of the spin-2 field is not really well defined without such a background metric, which means that the spin-2 approach to gravity sort of pulls up the ladder that was used to scale the wall. The "proper" GR approach and the spin-2 GR approach are only demonstrably equivalent in a very specific and complicated limit [95] [96]. The status of their relation or potential equivalence outside of this limit is not clear. It is therefore not certain that quantizing spin-2 fields on a flat background can capture the essential insights of GR. This kind of background dependent quantizing seems, at least superficially, to disagree with the fundamental importance that GR assigns to a background free formulation of physics. It is for instance not clear how it could be possible that GR denies time in favor of relational dynamics, while the spin-2 approach retains the time-translational invariance of flat Minkowski space. These subtle differences (and sometimes not so subtle differences), will become important when we review the spin-2 approach to gravity used in string theory and supergravity in chapters 9 and 10.

4.10 Notes

The equivalence principle

Ohanian and Ruffini (hereafter referred to as Ohanian) in their textbook "Gravitation and spacetime" seem to claim that gravity cannot be eliminated locally [97, Section 1.8]. Or in other words that the gravitational field can be determined in an infinitesimal region of space. Ohanian identifies the gravitational field with the curvature tensor and correctly states that this is an invariant quantity that cannot be eliminated in any frame. We find that this identification is not correct and that as long as one considers an infinitesimal region of *spacetime* the effects of gravity can be eliminated locally. Ohanian's viewpoints would invalidate the formulations of equivalence principles that we have used in this chapter.

III Extended theories I - Noncommutative geometry

5. Noncommutative geometry - mathematics

We need to go back to the insights behind general relativity and quantum field theory, learn to hold them together in our minds, and dare to imagine a world more strange, more beautiful, but ultimately more reasonable than our current theories of it.

Baez

Summary

Both physical systems and geometric spaces can be described by algebraic methods. Classical systems and ordinary continuum space can be described by commutative algebras. In physical systems the transition from a commutative algebra to a noncommutative algebra signifies the transition form classical to quantum. For spaces the corresponding transition leads from ordinary differential geometry to noncommutative (differential) geometry (NCG). NCG is a significant generalization of the concept of space and spacetime. Ordinary topological spaces, vector bundles, measure spaces, one-forms, and connections can all be given a commutative algebraic formulation. These spaces and structures can then be generalized to the noncommutative case. We study the constructions of such noncommutative spaces and structures in this chapter. The most important construction we present is the generalization from (commutative) Riemannian manifolds to noncommutative spin manifolds. These new spaces are defined by spectral triples. Tensor products of triples will form the basis of the physical models of chapter 6 and 7.

In this chapter we review the mathematical foundations of noncommutative geometry. The two following chapters utilizes this mathematical framework for building physical theories. We believe it is essential to have some understanding of the general mathematical field of noncommutative geometry to have any appreciation of the merits of the approach we are presenting in the subsequent chapters.

5.1 Introduction

Investigating noncommutative geometry is the start of a journey into a new understanding of what physical and mathematical space is. This journey is equal in magnitude to the travels started by Euclid, continued by Descartes, reshaped by Newton, and revolutionized by Gauss, Riemann and Einstein. Our concept of what space is, is a central part of our understanding of almost any field of knowledge. It is especially important in mathematics and in physics. A (mathematical) space can be defined as a set of points that satisfies some properties. Things like measure spaces, topological spaces, varieties and manifolds are usually defined this way (see chapters D E and F). A different approach to defining these spaces is to focus on the set of continuous (complex) functions defined on these spaces. The set of functions on a space can naturally be given the structure of a commutative algebra (see definition 5.2.1 and lemma 5.2.2). This algebra will automatically be of certain kind depending on the type of space the functions are defined on. We can then ask whether we can start with just the abstract algebra, and from this recover the space in question. In many cases this is indeed possible, and we can derive the space and its properties just from the algebra (see table 5.1). These types of correspondences will be our main concern in this chapter.

Geometry	**Algebra**
Measure space	Measurable functions
Topological space	Continuous functions
Algebraic variety	Polynomial functions
Manifold	Smooth functions

Table 5.1: A basic set of correspondences between spaces and function algebras.

Let us quickly review the generic aspects of the correspondence between spaces and algebras, and introduce some relevant terminology. We will typically have on the one side, a space defined in terms of a point set with some additional requirements on it. There may also some extra structures defined on the space. On the other side, we will have an algebra satisfying some properties, possibly with some other structures defined on it. We then wish to establish that there is a complete correspondence between theses two descriptions. This correspondence is described by saying there is a functor from the category of spaces (of the relevant type), to the category of algebras (of the relevant type), and another functor going in the opposite direction (see chapter B for category theory and functors). Any such functor between two categories A and B will also translate the morphisms between objects of A to morphisms between objects of B. A functor is called covariant or contravariant according to whether it preserves or reverses the directions of the morphisms it translates.

We can thus, for many types of spaces, define the space by specifying an algebra of a certain type. Not only can we define the space itself in this manner, but we can also define all the accessory structures associated to the space in a purely algebraic manner. Objects like vector fields and connections can be defined as associated structures on the algebra.

The mere existence of such a parallel characterization of a space and associated objects makes for an interesting fact in itself, but its significance does not end there. Algebra is richly supplied with interconnections to other mathematical fields, and it also is equipped with many powerful computational techniques. Gaining access to these techniques is another major motivation. A third reason is that an algebraic formulation opens a vista towards generalizations that simply cannot be defined outside of algebra. Here the algebra takes center stage, and we go beyond the limits of

considering just function algebras on "classical" spaces. All three aspects will be of interest to us in the following, but the main focus will be on the third reason, the possibility of generalizations. Let us proceed straight away to the specific version of this third reason that we will be the focal point of this chapter.

The algebra of complex functions on a set is always commutative (see lemma 5.2.2). From this it follows that all "classical" spaces, at least those that can be formulated algebraically, will be functorialy related to commutative algebras. This suggests a possible generalization, we can drop commutativity as a requirement, and allow for noncommutative algebras. If the required properties of the commutative algebra in question also make sense in the noncommutative case, this suggest that there might be interesting objects representing noncommutative generalization of classical spaces. Note that any associated objects will then also have to be defined algebraically in a manner that make sense without commutativity. This is the essence of noncommutative geometry.

In this chapter we take noncommutative geometry to mean noncommutative differential geometry in the manner first suggested by Alain Connes [17] [98][99]. The commutative to noncommutative transition underlying this generalization is by itself a simple idea, but the complications involved in making any kind of mathematical sense of this program are substantial. Luckily, from all these complications, a beautiful and powerful framework arises. This framework provides access to a large set of new and interesting "abstract" spaces. Not only are many new spaces provided for, but also in many cases a much improved understanding of several classical spaces that cannot be properly understood by classical methods, nor from their commutative function algebra [17]. [100]Last but not least, noncommutative geometry is strongly related to, and inspired by, the process of quantization in quantum physics [17][100]. For this reason this field has always had a close ties with physics. Several applications have been found in fundamental quantum theory, in statistical mechanics, in particle physics and in quantum gravity [17][101][102]. In this chapter we focus on the mathematical framework, while in the next chapters we focus on applying this framework to physics.

5.2 Noncommutative topology

Topological spaces are among the most fundamental spaces in modern mathematics [103][104]. The axioms of a topological space define the abstract version of “the property of being close to” or “closeness” (see chapter E a brief review on topology). There are no angles or distances in a topological space, and therefore no geometry proper. Instead there is something we could characterize as “overall fuzzy shape” or “overall closeness structure”. It is hard (or perhaps even impossible) to draw a topological space in the sense that we can only draw object having some geometry, and a topological space has no geometry. When we see a drawing of a topological space we must interpret it as a representative of an infinite equivalence class of different geometries having the same topology. This is a way for us, as geometric beings, to "de-geometrize" topological objects. Only a set of specialized properties are meaningful for a topological space. These are properties like compactness, connectedness, and of being simply connected (see chapter E). These properties tell us which topological space we are dealing with. In addition to these binary properties there are also properties that are expressed by numerical and algebraic means. These are properties like the genus and the fundamental group that are defined in algebraic topology (which is not quite the same as the algebraic topology we are currently discussing)(see chapter I for more on algebraic topology). Genus is an integer topological property that applies to compact surfaces. The genus of a doughnut and a coffee mug are both equal to one.

The aim of this section is to show that, in addition to the quite simple construction of a contravariant functor (cofunctor) from the category of compact Hausdorff topological spaces to the category of unital commutative C*-algebras, there is also a cofunctor in the opposite direction

(see figure 5.1 top row). The existence of this reverse direction functor makes these two categories equivalent (see figure 5.1 bottom row). (To be precise one category is equivalent to the opposite of the other category.) We further show that it is meaningful to extend this construction to noncommutative C*-algebra thereby arriving at noncommutative topological spaces.

Let us limit our attention to a compact Hausdorff topological space X. We study this topological space by studying the set of continuous complex valued function $C(X)$ on it. This set of functions has a natural addition, multiplication and complex scalar multiplication defined on them by pointwise addition and multiplication in the codomain of complex numbers. Complex conjugation in the codomain also gives a natural involution operation. We list the formal definition of these operations in the following definition.

Definition 5.2.1 — Algebra of functions. Let X be a topological space and let $C(X)$ be the set of continuous complex valued functions on X. We define the following operations on $C(X)$.

$$(f+g)(x) = f(x) + g(x) \tag{5.2.1}$$
$$(f \bullet g)(x) = f(x) \cdot g(x) \tag{5.2.2}$$
$$(\lambda f)(x) = \lambda [f(x)], \ \lambda \in \mathbb{C} \tag{5.2.3}$$
$$f^*(x) = [f(x)]^* \tag{5.2.4}$$

The function $f(x) = 1$ act as a unit in this algebra. The definitions for norms, spectra, representations, Banach algebras, *-algebras, C*-algebras and much more that are used in the following, can be found in the "additional mathematical topics" chapters.

Lemma 5.2.2 — Algebra of functions. The set of continuous complex valued functions $C(X)$ with the operations of addition, multiplication, scalar multiplication and star conjugation as defined above is a (complex) unital *-algebra.

Proof. Trivial. ■

To this we add a norm given by the supremum norm.

Definition 5.2.3 — Supremum norm. The supremum norm for the algebra of continuous complex valued functions $C(X)$ on a compact space X is defined by

$$\|f\| = \sup_{x \in X} |f(x)| \in \mathbb{R}^+. \tag{5.2.5}$$

Proposition 5.2.4 The normed algebra $(C(X), \|\|)$ of continuous complex valued functions on on a compact space X is a Banach algebra.

Proof. This follows from the fact that $\mathbb{C}$ is a Banach space. See [105, p165] ■

We next establish some important properties of the norm.

Lemma 5.2.5 For the algebra $C(X)$ as given above we have $\|f\| = \|f^*\|$ making it a Banach $*$-algebra

Proof. To prove this we need to prove that $\|f\|_{sup} = \|f^*\|_{sup}$. This is straight forward, we have that $\|f^*\|_{sup} = \sup_{x \in X} |[f(x)]^*| = \sup_{x \in X} |[f(x)]| = \|f\|_{sup}$. Where the second equality follows from $|[f(x)]^*| = |[f(x)]|$. (The absolute value of a complex number is equal to the absolute value of its complex conjugate.) This makes $C(X)$ a normed *-algebra. By proposition 5.2.4 $C(X)$ is a Banach algebra. Therefore $C(X)$ is a Banach *-algebra. ■

Lemma 5.2.6 For the algebra given above we have $\|f^*f\| = \|f\|^2$ making it a C^*-algebra

Proof. For any complex number $z = a + ib$ we have that $z\bar{z} = (a + ib)(a - ib) = a^2 + b^2 = |z|^2$. From this it follows that $|f^*f| = |f|^2$. From this is follows that $\|f^*f\| = \|f\|^2$. ■

Let now X and Y be two compact Hausdorff spaces and $C(X)$ and $C(Y)$ their corresponding function algebras. Given a continuous function (topological morphism) $f\colon X \to Y$ we define a map Cf by

$$Cf\colon C(Y) \to C(X) \tag{5.2.6}$$
$$Cf\colon C(Y) \ni g \mapsto h = (g \circ f) \in C(X). \tag{5.2.7}$$

The map Cf is an *-algebra homomorphism. We can then define the functor C by

$$C\colon \textbf{Top-Comp.-Hausdorff} \to \textbf{Alg-C*-Commut.} \tag{5.2.8}$$
$$C\colon X \mapsto C(X) \tag{5.2.9}$$
$$C\colon f \mapsto Cf \tag{5.2.10}$$

Our next target is to construct a functor in the opposite direction. We start by introducing the concept of the character of an algebra.

Definition 5.2.7 — Character of an algebra. A character μ of a complex Banach algebra A is a nonzero algebra homomorphism $\mu\colon A \to \mathbb{C}$.

In can be shown that such a character is always continuous, surjective and has norm one. For a C*-algebra a character is always a *-morphisms. The set of all characters (the character space) of an algebra A is denoted by $M(A)$.

Definition 5.2.8 — Gelfand spectrum. The set $M(A)$ of all characters of a commutative Banach algebra A is called the Gelfand spectrum of A. (When applied to a commutative C* algebra A, this is sometimes called just the spectrum of A or the Gelfand space of A.)

The kernel of each character is a maximal ideal, and there is thus a bijective correspondence between the set of characters and the set of maximal ideals [106][100].

Let A^* be dual space of A defined as the Banach space of all continuous linear functions $\mu\colon A \to \mathbb{C}$ (i.e. the topological dual space). The weak topology defines a topology on X by specifying the smallest topology making the elements of the (topological) dual space X^* continuous. The weak* topology defines a topology on X^* by requiring the the functions of X^{**} be continuous on X^*. We can now define a topology on the Gelfand spectrum $M(A)$ considered as a subspace of X^*.

Definition 5.2.9 — Gelfand topology. Since we have $M(A) \subset A^*$ we have a natural inclusion $\iota : M(A) \to A^*$. The relative (i.e. subspace) weak* topology on $M(A)$ is called the Gelfand topology.

For a unital algebra A the topological space $M(A)$ is compact. We can now define the functor we were seeking. Let A and B be two commutative unital C*-algebras, and let $M(A)$ and $M(B)$ be the associated character spaces. Let $\phi\colon A \to B$ be a *-algebra homomorphism. We define a map $M\phi$ by

$$M\phi\colon M(B) \to M(A) \tag{5.2.11}$$
$$M\phi\colon M(B) \ni \omega \mapsto \xi = (\omega \circ \phi) \in M(A). \tag{5.2.12}$$

The map $M\phi$ is a homeomorphism (topological morphism). We can then define the functor M by

$$M\colon \textbf{Alg-C*-Commut.} \to \textbf{Top-Comp.-Hausdorff} \tag{5.2.13}$$
$$M\colon A \mapsto M(A) \tag{5.2.14}$$
$$C\colon \phi \mapsto M\phi. \tag{5.2.15}$$

Our next task is to establish that the categories **Top-Comp.-Hausdorff** and **Alg-C*-Commut.** are equivalent. We do this by demonstrating that the composition of the functors M and C have the required properties. What we need to show is that $M \circ C$ is the identity functor on the category **Top-Comp.-Hausdorff**, and that $C \circ M$ is the identity functor on the category **Alg-C*-Commut.**

If we imagine the algebra A to be a function algebra of function into $\mathbb{C}$, then we can think of points as objects that maps each function into $\mathbb{C}$. A character is easily seen to be such a "point-like" map. In the other direction the evaluation of a function at a point provides a character for the set of functions on a set. The following definition makes this notion precise.

Definition 5.2.10 — Evaluation map (1). Let $A = C(X)$ and $x \in X$. The evaluation map ϵ_x on $C(X)$ is defined by

$$\epsilon_x\colon C(X) \to \mathbb{C} \quad \epsilon_x\colon f \mapsto f(x) \tag{5.2.16}$$

The map ϵ_x is a character of $C(X)$. That is $\epsilon_x \in M(C(X))$. We can use this to define a second map ϵ_X (note the uppercase X).

Definition 5.2.11 — Evaluation map (2). Let $A = C(X)$ and $x \in X$, and let ϵ_x be the evaluation map defined in definition 5.2.11. The evaluation map ϵ_X on X is defined by

$$\epsilon_X\colon X \to M\left(C(X)\right) \quad \epsilon_X\colon x \mapsto \epsilon_x \tag{5.2.17}$$

The map ϵ_X is a homeomorphism between topological space X and the Gelfand spectrum $M(C(X))$ with the Gelfand topology. Thus $\epsilon_X = M \circ C$ and it is the identity functor on the category **Top-Comp.-Hausdorff**.

Next we note that there are restrictions on the values that a character can take on any given algebra member.

Lemma 5.2.12 Let A be complex Banach algebra, and let $\sigma(a) \subset \mathbb{C}$ be the spectrum of $a \in A$. Let μ be character on A, then $\mu(a) \in \sigma(a)$.

Proof. Let 1_A and $1_{\mathbb{C}}$ be the multiplicative unit elements of A and $\mathbb{C}$. Since $\mu(a) \in \mathbb{C}$, the expression $a - \mu(a)1_A$ represents an element of A. We calculate the character of this element to be $\mu(a-\mu(a)1_A) = \mu(a)-\mu\left(\mu(a)1_A\right) = \mu(a)-\mu(a)\mu(1_A) = \mu(a)-\mu(a)1_{\mathbb{C}} = \mu(a)-\mu(a) = 0_{\mathbb{C}}$. Setting $b = (a - \mu(a)1_A)$, and then taking the inverse $b^{-1} = (a - \mu(a)1_A)^{-1}$ leads to $\mu(b^{-1}) = \mu(b)^{-1} = 0_{\mathbb{C}}^{-1}$ which is false. Hence $b = (a - \mu(a)1_A)$ has no inverse which (by definition) gives $\mu(a) \in \sigma(a)$. ∎

The next step is to define complex functions $\hat{a}$ on the set of characters $M(A)$, and a mapping from elements of the algebra (i.e. $a \in A$) to those functions (i.e. $\hat{a}$).

Definition 5.2.13 — Gelfand transform. The Gelfand transform are two maps $(\hat{a}, \mathcal{G})$ defined

by

$$\hat{a}\colon M(A) \to \mathbb{C} \quad \hat{a}\colon \mu \mapsto \mu(a) \tag{5.2.18}$$

$$\mathcal{G}\colon A \to C\left(M(A)\right) \quad \mathcal{G}\colon a \mapsto \hat{a}, \tag{5.2.19}$$

where a represents elements of A and μ represents characters in $M(A)$.

Lemma 5.2.14 Let A be a C^*-algebra and let $a \in A$ be self-adjoint ($a^* = a$), then

$$\forall\left(\mu \in M(A)\right))\, \mu(a) \in \mathbb{R}. \tag{5.2.20}$$

This means that self-adjoint elements of the algebra are mapped by Gelfand transform to real-valued functions on the set of characters.

Lemma 5.2.15 Any element $a \in A$ can be written as $a = a_1 + ia_2$ where a_1, a_2 are self-conjugate.

Lemma 5.2.16 Let $\hat{a}$ be the Gelfand transform of a and $\widehat{a^*}$ be the Gelfand transform of a^*, then

$$\widehat{a^*}(\mu) = \hat{a}(\mu)^* \tag{5.2.21}$$

Proof. Any element a can be written as $a = a_1 + ia_2$ where a_1, a_2 are self-conjugate. We calculate $\widehat{a^*}(\mu) = \mu(a^*) = \mu(a_1 - ia_2) = \mu(a_1) - i\mu(a_2) = \overline{\mu(a_1 + i\mu(a_2)} = \overline{\mu(a)} = \hat{a}(\mu)^*$. ■

We are now ready to state the main result for commutative C^*-algebras [98].

Theorem 5.2.17 — Commutative Gelfand-Naimark theorem. Let A be a commutative C^*-algebra and let X be the Gelfand spectrum of A. The Gelfand representation $\mathcal{G}$ is an isometric *-isomorphism from A to $C(X)$.

Proof. See [98, p7] and [107, p60]. ■

Another way of stating this result is to say that $\mathcal{G} = C \circ M$ and that this is the identity functor on the category **Alg-C*-Commut.**.

The Gelfand-Naimark theorem can be generalized to non-unital commutative C*-algebras, which corresponds to locally compact Hausdorf spaces, but we shall not do so here (see [98]). We note that when the algebra A is commutative we have three ways of constructing a corresponding topological space. The set of characters, the set of one-dimensional irreducible representations, and the set of maximal ideals. They are all equivalent (see [106][101][100]).

Generalizing the above result to the noncommutative case is the starting point of noncommutative geometry. In the noncommutative case the algebra is the fundamental definition of a space. We cannot prove an equivalence to any space defined by point-set methods for there is no such space. The algebraic definition of these spaces is the only definition we have of them. The generalization to noncommutative spaces is a leap of faith (or simply a postulate or a definition) that is supported by intuition and all the useful and interesting results it brings. In some cases we can determine that space defined by a noncommutative algebra to related to a space defined by a Morita equivalent commutative algebra (ref fix this).

Before we close this section we include a very important result that characterizes the general C^*-algebra. It shows that focusing on operator algebras is not a restriction on the set of algebras we are considering, since noncommutative C^*-algebras can all be realized as operator algebras on a Hilbert space.

Theorem 5.2.18 — Gelfand-Naimark theorem. Every C*-algebra A (commutative or noncommutative) is isometrically *-isomorphic to some C*-algebra of bounded operators on a Hilbert space.

Proof. See [98, p31] and [107, p109]. ■

Object	Classical definition	Commutative algebraic definition
Algebra	Continuous functions $C(X, \mathbb{C})$	Commutative C*-algebra $\mathcal{A}$
Points of space	$\forall x \in X$	All homomorphisms $h\colon \mathcal{A} \to \mathbb{C}$
Functions	Continuous functions $f(x) \in C(X, \mathbb{C})$	$h\,(f)$ with $f \in \mathcal{A}$
Morphisms	Continuous function $f\colon X \to Y$	C*-homomorphism $f\colon \mathcal{A} \to \mathcal{B}$
Isomorphisms	Homeomorphism $f\colon X \to Y$	C*-isomorphism $f\colon \mathcal{A} \to \mathcal{B}$

Table 5.2: Some of the central objects of topology and their algebraic counterparts.

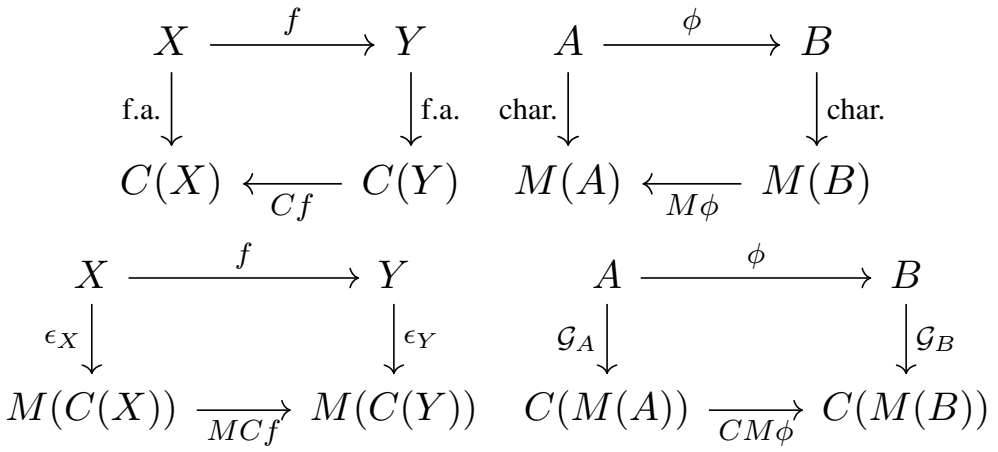


Figure 5.1: The upper two diagrams show the two Gelfand cofunctors (hereafter just called functors one and two). The left upper diagrams shows functor one that maps from a compact Hausdorff space X to the commutative unital C*-algebra $C(X)$. The right upper diagram shows functor two that maps from a commutative C*-algebra A to a compact Hausdorff space $M(A)$ defined by the characters of A with the Gelfand topology. The maps $Cf, M\phi$ are the maps induced by the other maps in the respective diagrams. E.g. f is a a continuous function and Cf is the corresponding algebra morphism (the image of f under functor one). $C(X)$ are the continuous functions on X and $M(A)$ is the set of characters of A. The bottom two diagrams shows how we get a identity map (isomorphism) on the respective categories when we compose the two functors from the upper two diagrams. E.g. (see lower left-hand diagram) if we map the compact Hausdorff space X to its corresponding commutative unital C*-algebra of functions by functor one, and the by functor two map this algebra to it character set, we get an isomorphic (=homeomorphic) topological space. Note that since we are composing two arrow reversing maps we end up with an arrow preserving map. Note also that while e.g. A and $C(M(A))$ are isomorphic algebras, $C(M(A))$ have the extra structure of being an algebra of functions. This is the essence of the categorical equivalence, the information about the "domain" and the "functions" is already implicitly present in the algebra. (Abbreviations: f.a.= map to function algebra, char. = map to set of characters)

5.3 Noncommutative measure theory

Measure spaces abstract the basic properties of integration and probability using basic set operations in combination with a map from a class of subsets to the positive reals (see chapter D). We have already analyzed measure spaces in the form of noncommutative probability spaces in chapter 2, so we will only repeat it briefly here from the perspective of illustrating the methodology of noncommutative geometry.

Let $(X, \mathcal{M}, \mu)$ be a measure space (see appendix D for a brief review on measure and probability spaces). We have seen that such a space can be reconstructed from $(\mathcal{A}, \omega)$ where

$$\mathcal{A} \coloneqq L^\infty(X, \mu, \mathbb{C}), \quad \omega \colon L^\infty \ni f \mapsto \int f d\mu \in \mathbb{C}.$$

Generally given any pair $(\mathcal{A}, \omega)$ where $\mathcal{A}$ is a commutative von Neumann algebra, and ω is faithful linear functional, there is a functor to the category of measure spaces, such that for the corresponding measure space we have

$$(\mathcal{A}, \omega) \simeq \left(L^\infty, \int d\mu \right).$$

In table 4 a list of all the central objects of measure theory as with their classical definition alongside their commutative algebraic definition.

Object	**Classical definition**	**Commutative algebraic definition**
Algebra	Bounded functions $L^\infty(X, \mu, \mathbb{C})$	Commutative W*-algebra $\mathcal{A}$
Points of space	$\forall x \in X$	All homomorphisms $h \colon \mathcal{A} \to \mathbb{C}$
Functions	Bounded functions $L^\infty(X, \mu, \mathbb{C})$	$h\left(A_f\right)$
Integral	$\int f d\mu$	$\omega\left(A_f\right)$
Measurable sets	$\forall m \in \mathcal{M}$	Projectors P in $\mathcal{A}$
Measure of a set	$\mu(m)$	$\omega(P)$
Morphisms	Measurable map $f \colon X \to Y$	W*-homomorphism $f \colon \mathcal{A} \to \mathcal{B}$

Table 5.3: Objects of measure theory

Except for the "points of space" and "functions" all these definition continue to make sense when we modify the algebra to allow for noncommutative W*-algebras [17]. Generalizing to noncommutative algebras defines noncommutative measure theory and noncommutative probability theory [38]. As we saw in chapter 2 noncommutative probability theory is one way to formulate basis of quantum mechanics (see also [38]).

5.4 Vector Bundles

Having looked at the algebraic definition of topological spaces and measure spaces in the two previous sections, we now move on to finding algebraic equivalents of vector bundles (see appendix F for a brief review on bundles). We will not develop this aspect in a detailed manner, but we want to mention the fundamental concepts. The basic idea is that the set of sections over some base space M is a module over the algebra of functions $C(M)$ [98]. Let us now make this more explicit. Consider a compact Hausdorff space M with continuous real-valued functions $C(M)$. Let E be a vector bundle $\pi \colon E \to M$ and let $\Gamma(E, M)$ be the space of continuous sections over M. Each

section is a "vector field" on M and these can be multiplied by functions from $C(M)$, and thus we see (informally) that $\Gamma(E, M)$ is a $C(M)$-module. Considering this in more detail one finds the following theorem.

Theorem 5.4.1 Let M be a compact space. The set of sections $\Gamma(E, M)$ is a finitely generated projective $C(M)$-module.

Proof. See [98, p59]. ■

The map defined by this theorem is a functor. Our major interest lies in finding a functor in the reverse direction. This will establish a new equivalence of categories. We simply state the result.

Theorem 5.4.2 — Serre-Swan theorem (1). Let M be a Hausdorff topological space. A finitely generated projective module over the algebra $C(M)$ is isomorphic to $\Gamma(E, M)$ for some vector bundle $\pi\colon E \to M$.

Proof. See [98, p59]. ■

We restate the above result using the more appropriate category theoretic language.

Theorem 5.4.3 — Serre-Swan theorem (2). Let M be a compact Hausdorff topological space, and let $A = C(X)$. Let **VecBun**(X) be the category of vector bundles over X, and let **FinProjMod**(A) be the category of finitely generated projective modules over A. The two categories **VecBun**(X) and **FinProjMod**(A) are equivalent as categories.

Proof. See [108, p7]. ■

This equivalence of categories forms the basis of a generalization to noncommutative algebras. We state it in a definition. As far as we can tell there is no well established terminology for these objects.

Definition 5.4.4 — Algebraic vector bundle. An algebraic vector bundle is a finitely generated projective A-module over some (commutative or noncommutative) algebra A.

We see the algebraic vector bundle as a vector bundle over the base space defined by the algebra A [100]. Again we rely on intuition and the interesting results that follows from this to substantiate our choice of definition.

This concludes our run through of the three major classic equivalences between algebras (or modules over algebras) on the one side, and topological spaces, measure spaces and vector bundles on the other side.

5.5 Differential calculus

Vectors and one-forms are important concepts from linear algebra that can be transported to the context of manifolds in differential geometry (see appendix F and [90]). To further enrich the idea of a space defined by an algebra, it is necessary to find algebraic replacements for the classical differential geometric definitions of vectors and one-forms.

We start by noticing that, even in the classical differential geometric setting, these structures are often introduced by purely algebraic definitions. Vector fields can be defined as the set of all derivations of the algebra $C^\infty(X)$ (see e.g. [45]). After introducing vector fields as derivations one can define one-forms (one-form fields) as linear maps on the vector fields. However, defining vector fields as derivations does not work in the general noncommutative setting [109], and beyond what is implicit in the previous section definition of vector bundles, we shall not even define the concept

of a vector. Instead, we focus on finding an appropriate generalization of differential calculus on manifolds by using differential forms [109].

The set of all forms combined with the exterior derivative operation, the exterior algebra, is an example of a differential algebra (the de Rham complex). The term differential algebra is even more appropriate in the purely algebraic setting. We therefore set our sights on finding differential algebras defined on general associative algebras.

Though we seek the full analog of the exterior algebra, it is sufficient to define this at the first-order level. The rest of the exterior algebra structure follows directly from this ansatz by linear extension [98][109]. Therefore, as a first step, we consider first-order differential algebras (FODA) which are also known as first-order differential calculi (FODC) [110][111][98].

Definition 5.5.1 — First-order differential calculus. Let $\mathcal{A}$ be an algebra and let $\Omega^1\mathcal{A}$ be an $\mathcal{A}$-bimodule. A first-order differential calculus is a triple $(A, \Omega^1\mathcal{A}, d)$ where d is a linear map $d\colon \mathcal{A} \to \Omega^1\mathcal{A}$ such that

$$d(ab) = d(a)b + ad(b) \text{ (graded Leibniz rule)}, \tag{5.5.1}$$

and every element $\omega \in \Omega^1\mathcal{A}$ can be expressed as

$$\omega = \sum_{n=1}^{N} a_n(db_n) \quad a_n, b_n \in \mathcal{A}. \tag{5.5.2}$$

To see how this relates to the definition of a full differential graded algebra (DGA) see appendix C. (Note that the graded Leibniz rule for a FODC is a special case with of general graded Leibniz rule with $deg(a) = 0$.) The classical model for differential algebras and calculi is the exterior derivative calculus (see appendix C).

■ **Example 5.5.2 — Exterior derivative.** The idea behind the FODC definition is to generalize the setup for functions and one-forms. Let X be a manifolds and $\mathcal{A} = C^\infty(X)$. The exterior derivative defines a map

$$d_{ext}\colon C^\infty(X) \to \Gamma(T^*(X)) \tag{5.5.3}$$

where $\Gamma(T^*(X))$ are the sections of the cotangent bundle. Setting $\Omega^1\mathcal{A} = \Gamma(T^*(X))$ we can declare that

$$(\mathcal{A} = C^\infty(X), \Omega^1\mathcal{A} = \Gamma(T^*(X)), d = d_{ext}) \tag{5.5.4}$$

is a FODC. ■

In the case of a classical manifold, an in example 5.5.2, the linear operator $d = d_{ext}$ is uniquely determined by the stipulated requirements. When considering noncommutative algebras there are many possible choices for a FODC [98][109]. However, even though there is no unique choice for $(\mathcal{A}, \Omega^1\mathcal{A}, d)$, there is one FODC construction that exists for every unital algebra $\mathcal{A}$. This construction has universal properties (in the categorical sense), and all of the FODC constructions we need can be made with this FODC as the starting point. It is called the universal FODC [98][109].

Definition 5.5.3 — Universal FODC. Let $\mathcal{A}$ be algebra with unit and let m be the tensor multiplication map

$$m\colon \mathcal{A} \otimes \mathcal{A} \to \mathcal{A} \tag{5.5.5}$$

$$m\colon a \otimes b \mapsto ab \tag{5.5.6}$$

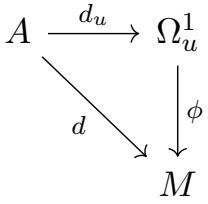


Figure 5.2

Both $\mathcal{A}$ and $\mathcal{A}\otimes\mathcal{A}$ are $\mathcal{A}$-bimodules by using the multiplication in $\mathcal{A}$. The map m is in fact a bimodule morphism. This means that the kernel of this map is a sub-bimodule $\Omega^1_u(\mathcal{A})$ of $\mathcal{A}\otimes\mathcal{A}$. We write

$$\Omega^1_u(\mathcal{A}) \coloneqq \ker(m\colon \mathcal{A}\otimes\mathcal{A}\to\mathcal{A}) \tag{5.5.7}$$

We define a map

$$d_u\colon \mathcal{A}\to\Omega^1_u(\mathcal{A}) \tag{5.5.8}$$
$$d_u\colon a\mapsto(1\otimes a - a\otimes 1). \tag{5.5.9}$$

We call the $(\mathcal{A},\Omega^1_u(\mathcal{A}),d_u)$ defined above the universal FODC associated to $\mathcal{A}$.

The differential d_u is also called the Karoubi differential [110].

Proposition 5.5.4 The triple $(\mathcal{A},\Omega^1_u(\mathcal{A}),d_u)$ defined in 5.5.3 is a FODC.

Proof. The proof consist of showing that the map is linear and Leibniz, and that every element is expressible in the required manner. Step 1: The map d_u is linear since by the properties of the tensor product we have $d_u(\alpha a+\beta b)=1\otimes(\alpha a+\beta b)-(\alpha a+\beta b)\otimes 1=1\otimes\alpha a+1\otimes\beta b-\alpha a\otimes 1-\beta b\otimes 1=\alpha(1\otimes a)+\beta(1\otimes b-b\otimes 1)=\alpha d_u(a)+\beta d_u(b)$. Step 2: The map d_u satisfies the Leibniz property since $d(ab)=1\otimes ab-ab\otimes 1=1\otimes ab+(-a\otimes b+a\otimes b)-ab\otimes 1=(1\otimes ab-a\otimes b)+(a\otimes b-ab\otimes 1)=(1\otimes a-a\otimes 1)b+a(1\otimes b-b\otimes 1)=d(a)b+ad(b)$. Step 3: We now show that any element $\omega\in\Omega^1_u(A)$ can be expressed as $\sum a_i d(b_{i)}$. First of all a generic element ω can be written as $\sum a_i\otimes b_i$ with the condition that $\sum a_i b_i=0$. We now show that in fact $\omega=\sum a_i b_i=\sum a_i d(b_i)$ for all a_i,b_i that satisfy $\sum a_i b_i=0$. This is true since $\sum a_i d(b_i)=\sum a_i(1\otimes b_i-b_i\otimes 1)=\sum a_i\otimes b_i-\sum a_i b_i\otimes 1=\sum a_i\otimes b_i-\sum a_i b_i(1\otimes 1)=\sum a_i\otimes b_i=\omega$. ■

The universal FODC gets its name from the fact that is has universal properties. We state this property in the usual category theoretic manner.

Theorem 5.5.5 Let $(\mathcal{A},\Omega^1_u(\mathcal{A}),d_u)$ be the universal FODC associated to $\mathcal{A}$, and let $(\mathcal{A},M,d)$ be some other differential calculus with M and $\mathcal{A}$-bimodule and d a derivation $d\colon\mathcal{A}\to M$. Then there exists a unique bimodule morphism $\phi\colon\Omega^1_u\to M$ such that the diagram in figure 5.2 commutes.

Proof. See [111] with a further reference to [112]. ■

The universal property means that any FODC on an algebra $\mathcal{A}$ is isomorphic with the quotient of the universal FODC with the kernel of the morphism ϕ [111].

The universal FODC is useful in its own sense, but in general it does not have all the properties we desire. A more powerful version comes from introducing algebra representations [109][113].

Lemma 5.5.6 Let π be a representation of the algebra $\mathcal{A}$ on the Hilbert space $\mathcal{H}$ by the map $\pi\colon \mathcal{A} \to L(\mathcal{H})$. For each linear operator D on H, the map defined by

$$\pi_D(a_1 da_2) = \pi(a_1)\,[D, \pi(a_2)] \tag{5.5.10}$$

defines a representation of the algebra $\Omega^1_u(\mathcal{A})$ on $\mathcal{H}$.

Proof. See [109, p41] ■

The image of this representation is not itself a FODC. To make the representation itself a proper FODC we need to divide out an ideal I of so called junk to get Ω/I. If we let D be an operator with some extra properties we can employ this definition directly for the first-order case [113].

Definition 5.5.7 — Connes differential one-forms. Let D be the Dirac operator of a spectral triple $(\mathcal{A}, \mathcal{H}, D)$ (see later sections) with a representation of $\mathcal{A}$ on $\mathcal{H}$ given by π. Connes differential one-forms are defined as $\mathcal{A}$-bimodule of operators on $\mathcal{H}$ of the form

$$\Omega^1\mathcal{A} = \left\{\omega \mid \omega = \sum \pi(a_1)\,[D, \pi(a_2)]\right\}. \tag{5.5.11}$$

We will use the Connes differential one-forms in the next chapter. When we want to extend this definition to higher-orders we need to divide out the "junk" as mentioned above [113].

5.6 Connections on modules

Bundles are a generalization of the idea of product spaces (see chapter F for brief review on bundles). More precisely, a bundle is a space that is locally a product space. We denote a bundle by giving the map

$$\pi\colon E \to M. \tag{5.6.1}$$

Another way to look at bundles is that they define a space (i.e. a vector space) over each point of a base space (often a manifold). The space over the point x is denoted E_x and is called the fiber over x. The fiber over x is given by $E_x = \pi^{-1}(x)$. Often one demands that all fibers are isomorphic. However, in this setting there is no canonical isomorphism between the fibers, which makes it impossible to (canonically) identify an element in one fiber with a specific element in another fiber. For such identifications we need a connection.

For each given path between two points x and y in the base manifold M, a connection provides an isomorphism between the fibers E_x and E_y. The definition of a connection can be expressed in many equivalent forms, but one that is suitable for generalization is the following: The connection on a given vector bundle E over the base space M is given by a map [98]

$$\nabla\colon \Gamma(E) \to \Gamma(E) \otimes \Omega^1(A). \tag{5.6.2}$$

The intuitive interpretation of this is that $\Gamma(E)$ represents a vector field $\mathcal{A}$ on M, and given another vector field B (as input to $\Omega^1(M)$) we get a new vector field C that represents the transport of vectors on $\mathcal{A}$ in the direction of B.

This can all be generalized without changes to the noncommutative case [98].

Definition 5.6.1 — Connection on module. Let E be an $\mathcal{A}$-module, and let $(\mathcal{A}, \Omega^1(\mathcal{A}), d)$ be the universal FODC on $\mathcal{A}$. A connection on the module E is a map

$$\nabla\colon \Gamma(E) \to \Gamma(E) \otimes \Omega^1(\mathcal{A}), \tag{5.6.3}$$

that satisfies the Leibniz rule

$$\nabla(sa) = (\nabla s)a + s \otimes da, \tag{5.6.4}$$

with $s \in E$ and $a \in \mathcal{A}$.

5.7 Spin manifolds

So far we have defined analogs of topological spaces, measure spaces, vector bundles, differential forms and connections. In the final step of our journey towards an algebraic understanding of space we will define (Riemannian) manifolds by algebraic means (see chapter F for a brief review on manifolds). As a first step towards categorizing a manifold algebraically we can state the following theorem.

Theorem 5.7.1 Let M and N be two smooth manifolds. Iff there is an $\mathbb{R}$-algebra isomorphism between $C^\infty(M)$ and $C^\infty(N)$ then M and N are diffeomorphic.

Proof. See e.g. [114]. ■

There is as of yet no direct algebraic characterization of the algebras involved in this result. Such a characterization is required for a reconstruction theorem. It is also believed that $C^\infty(M)$ is not quite enough to reconstruct the more complicated Riemannian manifolds. The missing piece is geometry. In differential geometry, geometry is provided by the metric, which is the defining property of a Riemannian manifold. We therefore need to find an algebraic structure that is analogous to the metric. This will be done by introducing the (generalized) Dirac operator. A generalized Dirac operator is a first order differential operator that squares to a Laplacian-like second order operator [98][115][116]. The definition of Dirac operators depends on the presence of spinors and Clifford algebras. We therefore specialize to the case of manifolds that allow such structures. These manifolds are called spin manifolds. Spin manifold are manifolds that allow the existence of a continuous orthonormal frame bundle [98][115]. From a physics perspective spin manifolds are essential to be able to treat fermionic fields. To define spin manifolds we first define what is known as a spin structure.

Definition 5.7.2 — Spin structure. Let E be an oriented n-dimensional vector bundle over a manifold X. Let $P_O(E)$ be its bundle of orthonormal frames. This is a principal O_n bundle. Choose an orientation on E which again chooses a principal $P_{SO}(E)$ bundle. Let $n \geq 3$. A spin structure on E is a principal Spin_n-bundle $P_{Spin}(E)$ together with a double covering

$$\rho\colon P_{Spin}(E) \to P_{SO}(E) \tag{5.7.1}$$

such that $\rho(pg) = \rho(p)\rho(g)$ for all $p \in P_{Spin}(E)$ and $g \in Spin_n$.

A spin manifold is a straight forward application of the idea a spin structure to a Riemannian manifold.

Definition 5.7.3 — Spin manifold. A spin manifold is an oriented Riemannian manifold with a spin structure defined on its tangent bundle.

The metric on a Riemannian manifold defines a distance function on the manifold.

Definition 5.7.4 — Distance function. Let M be Riemannian manifold with metric g, and let

$x, y \in M$. The distance function $d\colon M \times M \to \mathbb{R}$ on M is given by

$$d(x,y) = \inf\left\{\int \sqrt{g(\gamma'(t),\gamma'(t))}dt \mid \gamma\colon I \to M \wedge \gamma(0) = x \wedge \gamma(1) = y\right\} \tag{5.7.2}$$

The distance function d determines and actual metric on M. Knowledge of the distance function is enough to reconstruct the metric tensor g. It is therefore sufficient to find the algebraic equivalent of the distance function. Given the algebra $C^\infty(M)$ we wish to define a distance function on the characters space $M(C^\infty(M))$ of the algebra. The essential connection between a Riemannian metric and the Dirac operators is expressed in the following definition [17][101].

Definition 5.7.5 — Connes distance function. Let $\mathcal{A}$ be the algebra $C^\infty(M)$ and let $M(C^\infty(M))$ be the character space. Let $x, y \in M(C^\infty(M))$ and $f \in C^\infty(M)$, and let D_M be the Dirac operator on $L^2(M,S)$. We let $C^\infty(M)$ be represented on $L^2(M,S)$ as multiplication operators. For the characters x, y we define $f(x) = x(f)$ and $f(y) = y(f)$. The Connes distance function is given by

$$d(x,y) = sup\{|f(x) - f(y)| \mid f \in C^\infty(M),\, [D_M, f] \leq 1\} \tag{5.7.3}$$

Proposition 5.7.6 — Connes distance function. The Connes distance function reproduces the distance function defined by the Riemannian metric.

Proof. See [101] for a proof. ■

We can now start the process of collecting the algebraic pieces that we need to define a spin manifold M. First we temporarily hypothesize that cthe topological and differential structure to be given by the algebra $C^\infty(M)$. Next we introduce a representation of this algebra. The algebra will be represented on the Hilbert space defined by the square integrable spinor fields on M. We denote this Hilbert space as $L^2(M,S)$. The algebra acts by straight-forward multiplication. Finally we shall need a generalized Dirac operator, and in this case it is given by $D = i\gamma^\mu\nabla^S_\mu$. This three part structure are combined in the definition of a canonical triple [17][101][98].

Definition 5.7.7 — Canonical triple. A canonical triple is a three part structure (described above) associated to a compact Riemannian n-dimensional spin manifold M. It is given by $\left(\mathcal{A}_M = C^\infty(M), \mathcal{H}_M = L^2(M,S), D_M = i\gamma^\mu\nabla^S_\mu\right)$.

The interesting fact about the canonical triple is that Connes has proved that the elements of a canonical triple are sufficient to reconstruct the original spin manifold [117]. Thus this three part structure forms a natural starting point for noncommutative generalization.

Definition 5.7.8 — Spectral triple. A spectral triple is a tuple (A, H, D), where A is a unital *-algebra, H is a Hilbert space, D is a Dirac operator on H. In addition there is a faithful *-algebra representation π of A on H. We usually suppress the representation referring to both a and $\pi(a)$ as a. Theses objects must satisfy the following properties.

1. The operator D is self-adjoint.
2. The operator D has a compact resolvent $(i + D)^{-1}$.
3. The commutator $[D, a]$ is a bounded operator.

A spectral triple is sometimes referred to as an unbound Fredholm modulule. The structure (H, D) in the above definition is called a K-cycle on A. We can add further structure to this triple in the form of a grading operator [101][98].

Definition 5.7.9 — Even spectral triple. An even spectral triple is a spectral triple (A, H, D) where there is also a self-adjoint operator γ on H that gives a $\mathbb{Z}_2$-grading on H. The following equalities hold:

$$\gamma^2 = 1 \tag{5.7.4}$$
$$\gamma a = a\gamma \tag{5.7.5}$$
$$\gamma D = -D\gamma \tag{5.7.6}$$
$$\tag{5.7.7}$$

We can also add a real structure which induces an A-bimodule structure on H [101][98].

Definition 5.7.10 — Real spectral triple. A real spectral triple of KO-dimension $n \in \mathbb{Z}_8$ is a spectral triple (A, H, D) where there is also an anti-linear and anti-unitary operator J on H. With $\epsilon, \epsilon', \epsilon'' \in \{-1, 1\}$ valued according to the KO-dimension as given by the table below, the following equalities hold

$$J^2 = \epsilon \tag{5.7.8}$$
$$JD = \epsilon' DJ \tag{5.7.9}$$
$$\tag{5.7.10}$$

With $b^\circ = Jb^*J^{-1}$ the following equalities hold

$$[a, b^\circ] = 0 \tag{5.7.11}$$
$$[[D, a], b^\circ] = 0 \tag{5.7.12}$$

If the triple is also an even triple we refer to is as real even spectral triple, and require that

$$J\gamma = \epsilon''\gamma J \tag{5.7.13}$$

Note that real spectral triple is a short name for a spectral triple with a real structure and that is does not imply that the algebra of other elements are real. The definition of a real structure used here is a generalization of a the usual definition of a real structure (see appendix J).

k	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
ϵ	1	1	-1	-1	-1	-1	1	1
ϵ'	1	-1	1	1	1	-1	1	1
ϵ''	1		-1		1		-1	

Table 5.4: Definition of the KO-dimension of a spectral triple

It will turn out that there is a space of related Dirac operators for a given algebra and Hilbert space representation. In the next chapter we will see that spectral triple can be used to formulate physical theories that include general relativity and the standard model. Basically we can write down a functional over this space of Dirac operators and we can show that this functional is closely related to the structure of action functional for GR coupled to the standard model.

We include some related definitions that we will use in the next chapter.

Definition 5.7.11 — Finite spectral triple. A finite spectral triple $(\mathcal{A}_F, \mathcal{H}_F, D_F)$ is a spectral triple where the algebra A is a finite dimensional algebra.

Definition 5.7.12 — Separating vector. Let $\mathcal{A}$ be an algebra represented by π on $\mathcal{H}$. A separating vector for the representation of $\mathcal{A}$ is a vector $\xi \in \mathcal{H}$ such that

$$\pi(a)\xi = 0 \implies a = 0. \tag{5.7.14}$$

Definition 5.7.13 — Irreducible finite spectral triple. An irreducible finite spectral triple $(\mathcal{A}_F, \mathcal{H}_F, D_F)$ is a finite spectral triple where

1. The operator J_F and the representation of $\mathcal{A}_F$ acts irreducibly.
2. The representation of $\mathcal{A}_F$ has a separating vector.

Definition 5.7.14 — Product triple. By the tensor product of two triples $(\mathcal{A}_1, \mathcal{H}_1, D_1, J_1, \gamma_1)$ and $(\mathcal{A}_2, \mathcal{H}_2, D_2, J_2, \gamma_2)$ we mean the triple

$$(\mathcal{A}_1, \mathcal{H}_1, D_1, J_1, \gamma_1) \otimes (\mathcal{A}_2, \mathcal{H}_2, D_2, J_2, \gamma_2) \tag{5.7.15}$$
$$= (\mathcal{A}_1 \otimes \mathcal{A}_2, \mathcal{H}_1 \otimes \mathcal{H}_2, D_1 \otimes \mathbb{I} + \gamma_1 \otimes D_2, J_1 \otimes J_2, \gamma_1 \otimes \gamma_2) \tag{5.7.16}$$

Definition 5.7.15 — Almost commutative space. An almost commutative space is a spectral triple given as a tensor product of a canonical triple and finite triple. We write

$$(\mathcal{A}_M, \mathcal{H}_M, D_M, J_M, \gamma_M) \otimes (\mathcal{A}_F, \mathcal{H}_F, D_F, J_F, \gamma_F) \tag{5.7.17}$$

The resulting triple has the following content

$$(\mathcal{A}_M \otimes \mathcal{A}_F, \mathcal{H}_M \otimes \mathcal{H}_F, D_M \otimes \mathbb{I} + \gamma_M \otimes D_F, J_M \otimes J_F, \gamma_M \otimes \gamma_F) \tag{5.7.18}$$

We have implicitly assumed the existence of an inner product by labeling $L^2(M, S)$ as a Hilbert space. We shall need to know the explicit form of this inner product. Note the usage of angled and round parentheses and the use of a comma. This notation is conventional in NCG:

Definition 5.7.16 — Spinor inner product. Let $\pi\colon S \to M$ be a vector bundle. From the canonical fiberwise inner product w cam define Hermitian pairing

$$(\ ,\ )\colon \Gamma(S) \times \Gamma(S) \to C(M), \tag{5.7.19}$$

we define the inner product on $L^2(M, S)$ as

$$\langle f(x), g(x) \rangle = \int_M (f(x), g(x)) \sqrt{g} dx \quad x \in M, f, g \in L^2(M, S). \tag{5.7.20}$$

6. Noncommutative geometry - physics overview

Alain Connes' noncommutative geometry (...) is a systematic quantization of mathematics parallel to the quantization of physics effected in the twenties. (...) This theory widens the scope of mathematics in a manner congenial to physics

Kastler

Summary

Canonical spectral triples define classical spin manifolds, while finite spectral triples define zero-dimensional finite spaces. A tensor product of a canonical triple with a finite triple defines an almost commutative space. When used as a model for physics we will refer to such spaces as Connes-Lott-Chamseddine (CLC) models. CLC models allow us to derive GR and spontaneously broken Higgs-Dirac-Yang-Mills theory, with minimal gravity-matter coupling. The algebra of the finite triple determines the details of the resulting gauge theory, and is an essential input to the CLC models. Two set of arguments lead to the prediction of a (practically) unique finite triple for four-dimensional CLC models from first principles. The algebra for this triple is $M_2(\mathbb{H}) \oplus M_4(\mathbb{C})$. This algebra leads to the Pati-Salam model with $SU(2) \oplus SU(2) \oplus SU(4)$ gauge symmetry. It contains the $U(1) \oplus SU(2) \oplus SU(3)$ standard model. CLC models provide a compact geometric explanation of several parts of the standard model.

In this chapter we review how tensor products of spectral triples are used to make physics theories of the Connes-Lott-Chamseddine (CLC) type.

6.1 Introduction

In GR we observed that gravitational physics is modeled by the metric geometry of a four-dimensional Lorentzian manifold (see chapter 4). The metric, in its role as the specifier of geometry, is considered the key physical variable. The symmetry of the system is diffeomorphism invariance, and the physics is extracted by postulating a diffeomorphism invariant action principle which is varied with respect to the metric. A similar idea is present in the principal bundle based understanding of gauge theories and gauge symmetries (see e.g. ([118]). However, even though the diffeomorphism symmetries and the gauge symmetries can be combined in a single geometric model, these two types of symmetries are always treated as separate entities [119]. In this chapter we will see a different approach. In noncommutative geometry diffeomorphisms and gauge symmetries are seen as diffeomorphisms of a generalized form of space (see figure 6.1 for a diagram of what we are about to describe).

To start off we imagine a space X whose diffeomorphisms $\mathrm{Diff}(X)$ are equal, not just to the symmetries of gravitational physics, but is equal to the symmetries of all of physics. Armed with such a space one then tries to determine a diffeomorphism invariant action for the metric. At the Planck scale this could be some highly noncommutative space with an unknown set of symmetries given by $\mathrm{Diff}(X)$. Finding such a space X would be the ultimate goal of the NCG program. A stepping stone to uncovering such a space could be to consider a simpler space Y which would serve as a valid approximation at lower scales. The symmetries $\mathrm{Diff}(Y)$ of Y could be seen as remnants of the full symmetries $\mathrm{Diff}(X)$ of X.

We know from experiments that the symmetry of physics below the Planck scale seems to be well approximated by $\mathrm{Diff}(M) \rtimes G$ for some gauge group G. If the symmetries are to be given by $\mathrm{Diff}(Y)$ the space Y cannot be an ordinary manifold, because ordinary manifolds cannot have a non-simple diffeomorphisms group. However, noncommutative spaces in the form of spectral triples can have a non-simple diffeomorphisms group. Specifically, product triples of the almost commutative kind (AC) are spaces with just such diffeomorphisms groups [119]. These AC spaces are a kind of product spaces that we can write symbolically as $M \times F$. More accurately they are the product of a canonical triple, representing the continuous spacetime part M, and a finite triple, representing a finite noncommutative aspect F. This product space ansatz is superficially similar to the setup found in Kaluza-Klein theories and in Calabi-Yau compactifications of superstrings (see chapter 9 for more on Kaluza-Klein and Calabi-Yau compactifications).

When using spectral triples the space $Y = M \times F$ is represented by the algebra $\mathcal{A} = \mathcal{A}_M \otimes \mathcal{A}_F$, and the symmetry principle $\mathrm{Diff}(Y)$ is replaced by $\mathrm{Aut}(\mathcal{A})$, the group of automorphisms of the algebra $\mathcal{A}$. The finite dimensional part of the algebra $\mathcal{A}_F$ is noncommutative and will give rise to the group of inner automorphism $\mathrm{Inn}(\mathcal{A})$. The group of outer automorphism $\mathrm{Out}(\mathcal{A})$ is isomorphic to the group of diffeomorphism $\mathrm{Diff}(M)$. For an AC spectral triple we will symbolically have $\mathrm{Diff}(M \times F) \simeq \mathrm{Aut}(\mathcal{A}) \simeq \mathrm{Out}(\mathcal{A}) \rtimes \mathrm{Inn}(\mathcal{A}) \simeq \mathrm{Diff}(M) \rtimes G$.

Just as in GR the metric will be the primary dynamical variable. In spectral triples the metric is determined by the Dirac operator. We therefore seek an action principle invariant under $\mathrm{Aut}(\mathcal{A})$ that can determine the metric (in the form of a Dirac operator). In the process of analyzing spectral triples for use in physics, one realizes that the individual spectral triples should be considered as representatives of a (pseudo-)equivalence class of spectral triples. Most of the structure of the spectral triple is unchanged among the different members of the equivalence class, but the Dirac operator experiences non-trivial changes [119][101]. One therefore seeks to find a general expression for the Dirac operators of the spectral triples of a given equivalence class.

The general form of the Dirac operator is called the fluctuated Dirac operator. It is the noncommutative aspect of the finite-dimensional algebra that induces fluctuations in the Dirac operator. The fluctuated form of the Dirac operator can therefore be parameterized by the group of

inner automorphism of the algebra (modulo a minor modification). These fluctuation are called inner fluctuations, and enter into the fluctuated Dirac operator in a manner similar to how gauge fields appear in covariant derivatives. The fluctuated Dirac operator is more general than the gauge covariant derivative in that the inner fluctuations will take on the role of both the Yang-Mills fields *and* the Higgs field.

The inner fluctuation express possible variation of the Dirac operator for a given (pseudo-) equivalence class. The again expresses the possible variations of the metric. From this ansatz one seeks variational principles that, based on the spectral information of the Dirac operator, determine the form of the metric. Such action principles have been devised for both the bosonic and the fermionic parts of the action [120]. The fermionic part can be straightforwardly calculated, but the bosonic part requires the use of sophisticated approximation methods.

Given an AC spectral triple based on $\mathcal{A} = \mathcal{A}_M \otimes \mathcal{A}_F$, one can deduce a physical action which includes GR and spontaneously broken Higgs-Dirac-Yang-Mills theory with minimal coupling to gravity [120]. The details of the Higgs-Dirac-Yang-Mills theory depend on the finite triple F. Therefore, if one accepts the conceptual framework we presented, the decisive input to the model is the form of the finite triple. Various argument have been proposed to deduce ab initio the form of this triple [120][121][122][123].

We start our presentation by finding the general form of the Dirac operator as part of a set of Morita self-equivalent spectral triples (section 6.3). This motivates some practical definitions relating to gauge groups and gauge Lie algebras (section 6.4). We examine the postulated action principles, and check that they have the proper invariance properties (section 6.6). We then look at how to approximate the bosonic action principle, and derive the general form of the action based on a generic finite triple (section 6.7). Finally we look at how we can deduce the algebra of the finite triple from first principles (section 6.8).

6.2 Notation

For a brief introduction to spectral triples and spectral triple terminology see section 5.7. See section H.5 for information on *-algebras and their associated morphisms. Throughout this chapter we assume that we are given a real even spectral triple $(\mathcal{A}, \mathcal{H}, D, J, \gamma)$. Usually it will be in the form $(\mathcal{A}_M, \mathcal{H}_M, D_M, J_M, \gamma_M) \otimes (\mathcal{A}_F, \mathcal{H}_F, D_F, J_F, \gamma_F)$, where $(\mathcal{A}_M, \mathcal{H}_M, D_M, J_M, \gamma_M)$ is given as $\big(C^\infty(M), L^2(M,S), i\gamma^\mu\nabla^S_\mu, J_M, \gamma_M\big)$ or $\big(C^\infty(M), L^2(M,S), D_M, J_M, \gamma_M\big)$. Note that the representation π of $\mathcal{A}$ on $\mathcal{H}$ is implicitly assumed whenever we have expression involving some $a \in \mathcal{A}$ with D, J, γ or when an algebra element is applied to some $\xi \in \mathcal{H}$. E.g. aJ means $\pi(a)J$ and $a\,,\psi\rangle$ means $\pi(a)\,,\psi\rangle$. Several definitions are written like e.g. $X(\mathcal{A})$ just referencing $\mathcal{A}$, but do in fact depends on the full $(\mathcal{A}, \mathcal{H}, D, J, \gamma)$, this should be understood as defining $X(\mathcal{A})$ for the $\mathcal{A}$ being part of the triple $(\mathcal{A}, \mathcal{H}, D, J, \gamma)$. Note that this chapter uses an Euclidean metric for spacetime.

It is perhaps helpful to suggest some approximate(!) and informal physics interpretations of the basic spectral triple before going into details. The Hilbert space $\mathcal{H}_M$ contains spinor fields $,\psi\rangle$ and J_M is the charge conjugation operator for such fields, and γ_m is the chirality operator. The Hilbert space $\mathcal{H}_F$ is a finite dimensional vector space where each basis vector (e.g. $,e\rangle$) represent the various possible chiral fermions and anti-fermions. The elements of the finite dimensional vector space only become proper spinors when tensored with the elements of continuous Hilbert space, and therefore can not be classified as Dirac or Weyl by themselves. The elements of the continuous Hilbert space are Dirac spinors. Sometimes we use the chirality operator to separate this continuous Hilbert space into a direct sum of Weyl spinor spaces. The operator J_F maps particles to anti-particles and vice versa, and the operator γ_F provides the Z_2-grading that gives $+1$ for left-handed fermions and -1 for right-handed fermions. The elements of $f \in \mathcal{A}_M$ are basically

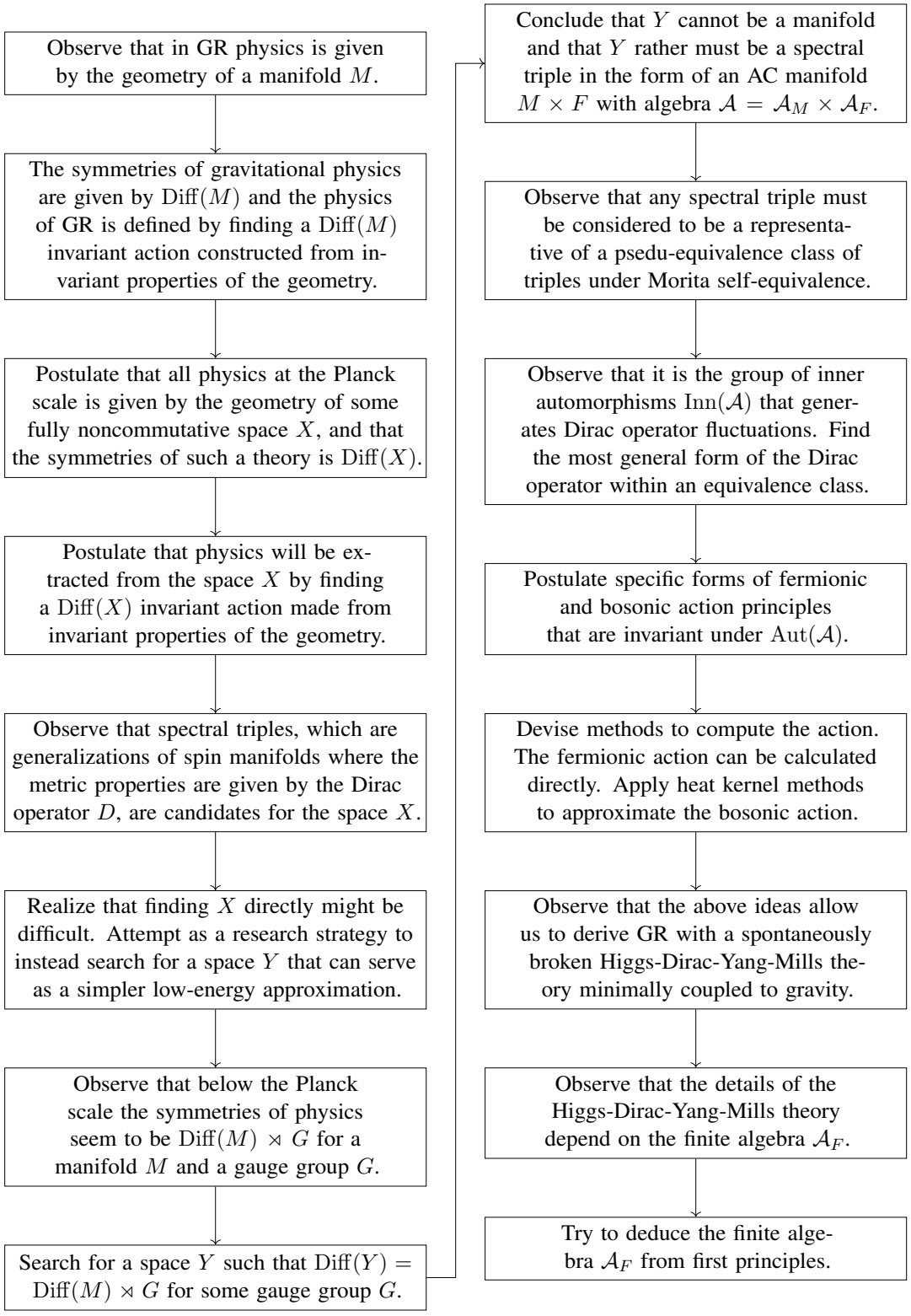


Figure 6.1: The conceptual build up of CLC type NCG models for physics. Note that the diagram includes a hypothetical space X which is more general than the one used in CLC models.

scalar fields, and the unitary elements of $\theta \in A_F$ are the group elements that gauge fields take their values in. Thus, very roughly, fermion fields are represented as $, \psi\rangle \otimes , e\rangle \in \mathcal{H}_M \otimes \mathcal{H}_F$, and gauge

fields as $f \otimes \theta \in \mathcal{A}_M \otimes \mathcal{A}_F$.

The notation (ψ, ξ) denotes the fiberwise inner product on $\mathcal{H}_M = L^2(S, M)$. The inner product on $\mathcal{H}_M = L^2(S, M)$ is denoted $\langle \psi, \xi \rangle$. The inner product on $\mathcal{H} = \mathcal{H}_M \otimes \mathcal{H}_F$ is denoted $\langle \psi \otimes e, \xi \otimes \nu \rangle$.

6.3 Inner fluctuations

We start by discussing how a spectral triple must be considered to be a representative of a (pseduo-) equivalence class of spectral triples under Morita equivalence [124][119]. The only part of the triple that varies within an equivalence class is the Dirac operator. It is therefore convenient that we extract a parameterized general representative of the Dirac operators of an equivalence class. This general form of the Dirac operator will be referred to as the fluctuated Dirac operator.

For commutative algebras the natural equivalence relation is given by ordinary isomorphisms. For noncommutative algebras an equivalence relation based on having equal representation theory is more natural. This stems, among other things, from the fact that the set of inequivalent representations define a set of points for the algebra [101]. The algebra elements can be considered to be functions on this set of points. This type of generalized algebra equivalence goes by the name of Morita equivalence [101][119].

Definition 6.3.1 — Morita equivalence. Two algebras $\mathcal{A}$ and $\mathcal{B}$ are Morita equivalent if the associated categories of $\mathcal{A}$-modules and $\mathcal{B}$-modules are equivalent as Abelian categories.

The following theorem expresses Morita equivalence in a form that will be useful.

Theorem 6.3.2 Two unital algebras $\mathcal{A}$ and $\mathcal{B}$ are Morita equivalent iff for some finitely generated projective $\mathcal{A}$-module $\mathcal{E}$ we have $\mathcal{B} \simeq End_{\mathcal{A}}(\mathcal{E})$.

Proof. See [101, p110] for a proof. ■

Note that $End_{\mathcal{A}}(\mathcal{E})$ is automatically a Hilbert $\mathcal{A}$-bimodule with a Hermitian structure $\langle\ ,\ \rangle : \mathcal{E} \times \mathcal{E} \to \mathcal{A}$. The next theorem records the form of the fluctuated Dirac operator in a very general case where we have two Morita equivalent algebras $\mathcal{A}$ and $\mathcal{B}$. Later we will specialize to the case of Morita self-equivalence where $\mathcal{B} = \mathcal{A}$.

Theorem 6.3.3 Let $(\mathcal{A}, \mathcal{H}_{\mathcal{A}}, D_{\mathcal{A}})$ be a spectral triple and let $\mathcal{B}$ be an algebra that is Morita equivalent to $\mathcal{A}$. Then $(\mathcal{B}, \mathcal{H}_{\mathcal{B}} = \mathcal{E} \otimes \mathcal{H}_{\mathcal{A}}, D_{\mathcal{B}} = \mathbb{I} \otimes D_{\mathcal{A}} + \nabla \otimes \mathbb{I})$ is a spectral triple, where $\mathcal{E}$ is a finitely generated module such that $\mathcal{B} = \mathrm{End}_{\mathcal{A}}(\mathcal{E})$, and ∇ is a Hermitian connection $\nabla \colon \mathcal{E} \to \mathcal{E} \otimes \Omega^1_{D_{\mathcal{A}}}(\mathcal{A})$.

Proof. Let us consider a spectral triple $(\mathcal{A}, \mathcal{H}_{\mathcal{A}}, D_{\mathcal{A}})$. We want to find another triple $(\mathcal{B}, \mathcal{H}_{\mathcal{B}}, D_{\mathcal{B}})$, which is based on the structure of $(\mathcal{A}, \mathcal{H}_{\mathcal{A}}, D_{\mathcal{A}})$, and where $\mathcal{B}$ is Morita equivalent to $\mathcal{A}$. Therefore by theorem 6.3.2 let us set $\mathcal{B} = \mathrm{End}_{\mathcal{A}}(\mathcal{E})$, where $\mathrm{End}_{\mathcal{A}}(\mathcal{E})$ is taken to be a $\mathcal{A}$-bimodule (with a Hermitian structure). We start the construction by defining $\mathcal{H}_{\mathcal{B}} = \mathcal{E} \otimes {}_{\mathcal{A}}\mathcal{H}_{\mathcal{A}}$. Let $\xi \in \mathcal{E}$ and $\eta \in \mathcal{H}_{\mathcal{A}}$ such that we have $\xi \otimes \eta \in \mathcal{H}_{\mathcal{B}}$. Note that this tensor product space is still a Hilbert space in a natural way. We define the inner product $\langle \xi \otimes \eta, \xi' \otimes \eta' \rangle = \langle \eta, \langle \xi, \xi' \rangle \eta' \rangle$. Since each element of $b \in \mathcal{B}$ is an endomorphism $b \colon \mathcal{E} \to \mathcal{E}$, the element b has a natural action on $\mathcal{H}_{\mathcal{B}} = \mathcal{E} \otimes {}_{\mathcal{A}}\mathcal{H}_{\mathcal{A}}$ defined by

$$b(\xi \otimes \eta) = b(\xi) \otimes \eta \tag{6.3.1}$$

If we could define an action of $D_{\mathcal{B}}$ to be given by

$$D_{\mathcal{B}}(\xi \otimes \eta) = \xi \otimes (D_{\mathcal{A}}\eta), \tag{6.3.2}$$

we would be done. But for a tensor product and with $a \in \mathcal{A}$ we must have

$$\xi \otimes (a\eta) = (\xi a) \otimes \eta. \tag{6.3.3}$$

This implies that we should have

$$D_{\mathcal{B}}(\xi \otimes a\eta) = D_{\mathcal{B}}((\xi a) \otimes \eta), \tag{6.3.4}$$

which means that we should have

$$\xi \otimes (D_{\mathcal{A}}a\eta) = (\xi a) \otimes (D_{\mathcal{A}}\eta) = \xi \otimes a(D_{\mathcal{A}}\eta) \tag{6.3.5}$$

or

$$(Da\eta) = a(D\eta). \tag{6.3.6}$$

But this is not the case since in general $Da \neq aD$. Therefore this choice of $D_{\mathcal{B}}$ is not possible. Instead we choose

$$D_{\mathcal{B}}(\xi \otimes \eta) = \xi \otimes (D_{\mathcal{A}}\eta) + (\nabla\xi)\eta, \tag{6.3.7}$$

where ∇ is a connection $\nabla\colon \mathcal{E} \to \mathcal{E} \otimes \Omega^1_{D_{\mathcal{A}}}(\mathcal{A})$. Here $(\nabla\xi) \in \mathcal{E} \otimes \Omega^1_{D_{\mathcal{A}}}(\mathcal{A})$. We know that $\Omega^1_{D_{\mathcal{A}}}(\mathcal{A}) \subset \mathcal{A}$ so that we can represent elements of $\mathcal{E} \otimes \Omega^1_{D_{\mathcal{A}}}(\mathcal{A})$ as $\xi' \otimes a$. We then define $(\nabla\xi)\eta$ as $(\xi' \otimes a)\eta = \xi \otimes a\eta$. Finally we check that $D_{\mathcal{B}}(\xi a \otimes \eta) = D_{\mathcal{B}}(\xi \otimes a\eta)$. We verify this by calculating that $D_{\mathcal{B}}(\xi a \otimes \eta) = \xi a \otimes (D_{\mathcal{A}}\eta) + (\nabla\xi a)\eta = \xi \otimes aD_{\mathcal{A}}\eta + (\nabla\xi)a\eta + (\xi \otimes [D, a])\eta = \xi \otimes D_{\mathcal{A}}a\eta + (\nabla\xi)a\eta = D_{\mathcal{B}}(\xi \otimes a\eta)$. ■

To analyze the same case for a real even spectral triple we first need to define the J-associated right action of the algebra $\mathcal{A}$ on $\mathcal{H}$.

Definition 6.3.4 — Right action on H. A right action on H is defined by $H(b) \equiv (Jb^*J^*)H$. The simultaneous left and right actions can then be written $aHb = aJb^*J^*H$. The definition of a real spectral triple stipulates that $(aH)b = a(Hb)$ which is the same as

$$[a, b^\circ] = 0 \tag{6.3.8}$$

With

$$b^\circ = Jb^*J^* \tag{6.3.9}$$

This is called the zeroth-order condition.

The next theorem extends theorem 6.3.3 to the case of a real even spectral triple..

Theorem 6.3.5 Let $(\mathcal{A}, \mathcal{H}_{\mathcal{A}}, D_{\mathcal{A}}, J_{\mathcal{A}}, \gamma_{\mathcal{A}})$ be a real even spectral triple and let $\mathcal{B}$ be an algebra that is Morita equivalent to $\mathcal{A}$. Then

$$\begin{aligned}(\mathcal{B}, \mathcal{H}_{\mathcal{B}} = \mathcal{E} \otimes \mathcal{H}_{\mathcal{A}} \otimes \bar{\mathcal{E}}, D_{\mathcal{B}} = \mathbb{I} \otimes D_{\mathcal{A}} \otimes \mathbb{I} + \nabla \otimes \mathbb{I} \otimes \mathbb{I} + \mathbb{I} \otimes \mathbb{I} \otimes \overline{\nabla},\\ J_{\mathcal{B}} = \mathbb{I} \otimes J_{\mathcal{A}} \otimes \mathbb{I}, \gamma_{\mathcal{B}} = \mathbb{I} \otimes \gamma_{\mathcal{A}} \otimes \mathbb{I})\end{aligned} \tag{6.3.10}$$

is a real even spectral triple, where $\mathcal{E}$ is a finitely generated module such that $\mathcal{B} = \mathrm{End}_{\mathcal{A}}(\mathcal{E})$, and ∇ is a Hermitian connection $\nabla\colon \mathcal{E} \to \mathcal{E} \otimes \Omega^1_{D_{\mathcal{A}}}(\mathcal{A})$, and $\bar{\mathcal{E}}$ is the conjugate module of $\mathcal{E}$. The connection $\bar{\nabla}$ is defined by anti-linear factor reversal from ∇. That is if $\nabla\xi = \xi' \otimes \omega$ then

$\overline{\nabla}\bar{\xi} = \omega^* \otimes \overline{\xi'}$.

Proof. The proof is essentially the same as for theorem 6.3.3. ■

We now want to specialize to the case of Morita self-equivalence where $\mathcal{E} = \mathcal{A}$ which also means that $\mathcal{B} = \mathrm{End}_{\mathcal{A}}(\mathcal{E}) = \mathcal{A}$. We first deal with the case where there is no real or even structure.

Theorem 6.3.6 Let $(\mathcal{A}, \mathcal{H}, D)$ be a spectral triple then

$$(\mathcal{A}, \mathcal{H}, D + A) \tag{6.3.11}$$

is also a spectral triple, where $A \in \Omega^1_D(\mathcal{A}) = \{\sum a_j\,[D, b_j] \mid a_j, b_j \in \mathcal{A}\}$ and $A = A^*$.

Proof. From theorem 6.3.3 we have that $(\mathcal{B}, \mathcal{H}_{\mathcal{B}} = \mathcal{E} \otimes \mathcal{H}_{\mathcal{A}}, D_{\mathcal{B}} = \mathbb{I} \otimes D_{\mathcal{A}} + \nabla \otimes \mathbb{I})$ is a spectral triple. But now we have $\mathcal{E} = \mathcal{A}$ and $\mathcal{B} = \mathrm{End}_{\mathcal{A}}(\mathcal{E}) = \mathcal{A}$ which implies that $\mathcal{E} \otimes \mathcal{H}_{\mathcal{A}} \simeq \mathcal{H}_{\mathcal{A}}$. This means that $D_{\mathcal{B}} = \mathbb{I} \otimes D_{\mathcal{A}} + \nabla \otimes \mathbb{I} = D_{\mathcal{A}} + \nabla$. And it also means that the connection $\nabla \colon \mathcal{E} \to \mathcal{E} \otimes \Omega^1_{D_{\mathcal{A}}}(\mathcal{A})$ is now equivalent to $\nabla \colon \mathcal{E} \to \Omega^1_{D_{\mathcal{A}}}(\mathcal{A})$. We therefor just need to specify a single value $\nabla(\mathbb{I})$ and extend by linearity. This single value is given by some $A \in \Omega^1_{D_{\mathcal{A}}}(\mathcal{A})$. Since we require the Dirac operator to be self-adjoint we must require that $A^* = A$. ■

The next theorem deals with Morita self-equivalence for a real even spectral triple.

Theorem 6.3.7 Let $(\mathcal{A}, \mathcal{H}, D, J, \gamma)$ be a real even spectral then

$$\left(\mathcal{A}, \mathcal{H}, D + A + \epsilon' J A J^{-1}, J, \gamma\right) \tag{6.3.12}$$

is also a real even spectral triple, where $A \in \Omega^1_D(\mathcal{A}) = \{\sum a_j\,[D, b_j] \mid a_j, b_j \in \mathcal{A}\}$ and $A = A^*$.

Proof. The proof is essentially the same as for the theorem above. ■

The conclusions we have reached are compactly summarized in the following definitions.

Definition 6.3.8 — Inner fluctuations. The inner fluctuations of a spectral triple $(\mathcal{A}, \mathcal{H}, D)$ are defined as the set of self-adjoint one-forms

$$A = \sum a_j\,[D, b_j] \quad A = A^*. \tag{6.3.13}$$

We can also write this definition as $A \in \Omega^1_D(\mathcal{A}) \wedge A = A^*$.

Definition 6.3.9 — Fluctuated Dirac operator. The fluctuated Dirac operator of a real even spectral triple $(\mathcal{A}, \mathcal{H}, D, J, \gamma)$ is defined in terms of the inner fluctuations as

$$D_A = D + A + \epsilon' J A J^{-1}. \tag{6.3.14}$$

Where ϵ' refers to the ϵ' of the KO-dimension of the triple (see table).

Note that this definition defines a class of Dirac operators. We want to make sure that this definition captures the general form of the Dirac operator. Thus we must make sure that further inner fluctuations does not produce a different expression. We deal with the case without real and even structure first.

Theorem 6.3.10 Let $(\mathcal{A}, \mathcal{H}, D)$ be a spectral triple. If the Dirac operator D is fluctuated by the

inner fluctuation A to be $D' = D + A$ and we fluctuate the Dirac operator D' with another inner fluctuation B to become $D'' = D' + B$ then $D'' = D' + B = D + C$ with $C = A + B$.

Proof. See [119, p195] ■

We have a similar result for the case with a real and even structure.

Theorem 6.3.11 Let $(\mathcal{A}, \mathcal{H}, D, J, \gamma)$ be a spectral triple. If the Dirac operator D is fluctuated by the inner fluctuation A to be $D' = D + A + JAJ^*$ and we fluctuate the Dirac operator D' with another inner fluctuation B to become $D'' = D' + B + JBJ^*$ then $D'' = D' + B + JBJ^* = D + C + JCJ^*$ with $C = A + B$.

Proof. See [119, p195] ■

As we saw, the expression for the fluctuated Dirac operator should be understood as the separate inner perturbations coming from the left and right action of A on H, represented by the second last and last term of $D_A = D + A + \epsilon' JAJ^{-1}$ respectively. Note that although Morita equivalence is an equivalence relation on algebras it has not been lifted to an equivalence relation on spectral triples. The inner fluctuations lead to a relation between triples that is reflexive and transitive but not in general symmetric. For this reason we have used the word pseudo-equivalence classes to refer to the resulting "partitions" (which are not proper partitions).

We now look at the fluctuated Dirac operator for an AC triple. To derive the fluctuated Dirac operator we first calculate separately the inner fluctuations A_μ and ϕ coming from the continuous and discrete parts of the original Dirac operator. The next step is to use the expression $D_A = D + A + \epsilon' JAJ^{-1}$. We state the result in a theorem.

Theorem 6.3.12 Let $\left(C^\infty(M), L^2(M,S), i\gamma^\mu \nabla^S_\mu, J_M, \gamma_M\right) \otimes (\mathcal{A}_F, \mathcal{H}_F, D_F, J_F, \gamma_F)$ be an AC spectral triple. We refer to the Dirac operator of the first part as D_M. Let the the continuous and discrete inner fluctuations be given by

$$\begin{aligned} A_\mu &= \sum a_j \left[D_M, b_j\right] \quad A_\mu = A_\mu^* \\ \phi &= \sum a_j \left[D_F, b_j\right] \quad \phi \ \ = \phi^*. \end{aligned}$$

The fluctuated Dirac operator for the AC spectral triple is given by

$$D_A = D_M \otimes \mathbb{I} + \gamma^\mu \otimes B_\mu + \gamma_M \otimes \Phi, \tag{6.3.15}$$

where

$$B_\mu = A_\mu - J_F A_\mu J_F^{-1}, \quad \Phi = D_F + \phi + J_F \phi J_F^{-1}. \tag{6.3.16}$$

The symbols γ^μ denote the basis of the Clifford algebra associated to M.

Proof. See [101, p141] for a proof. ■

These results are the essence of what we wanted to find. We could stop here and just use these results as they are. However, we would like to relate the inner perturbation more directly to the structure of the algebra. To do this we now embark on a long series of definitions. This new set of terminology will allow us to say that to each triple there is an associated gauge group, and that the the inner fluctuations that we have just discussed are determined by the gauge group. Specifically we will find that the inner fluctuations take values in the Lie algebra of the associated gauge group.

6.4 Gauge structure

In this section we introduce terminology to allow a more familiar description of the inner fluctuations. We start by noting that when a space is described by an algebra, the role of diffeomorphism is taken over by automorphism. We denote by $\mathrm{Aut}(\mathcal{A})$ the set of all automorphisms of the algebra $\mathcal{A}$. It is useful to distinguish between inner and outer automorphisms.

Definition 6.4.1 — Inner automorphism. An inner automorphism on a *-algebra A is an automorphisms $\alpha_u \colon A \to A$ given by

$$\alpha_u \colon a \mapsto uau^*,\ u \in \mathcal{U}(A). \tag{6.4.1}$$

The set of all inner automorphisms on $\mathcal{A}$ forms a group that we denote as $\mathrm{Inn}(\mathcal{A})$. The group of inner automorphisms is a normal subgroup of $\mathrm{Aut}(\mathcal{A})$ [119].

Definition 6.4.2 — Outer automorphism. An outer automorphism is an automorphism that is not inner. I.e. an automorphism $\alpha \colon A \to A$ is an outer automorphism if $\alpha \notin \mathrm{Inn}(\mathcal{A})$.

The set of all outer automorphisms on $\mathcal{A}$ does not form a group. The group of outer automorphisms, which is denote as $\mathrm{Out}(\mathcal{A})$, is instead defined as the quotient

$$\mathrm{Out}(\mathcal{A}) = \mathrm{Aut}(\mathcal{A})/\mathrm{Inn}(\mathcal{A}). \tag{6.4.2}$$

We summarize the relations between the various automorphism groups by the diagram

$$1 \longrightarrow \mathrm{Inn}(\mathcal{A}) \longrightarrow \mathrm{Aut}(\mathcal{A}) \longrightarrow \mathrm{Out}(\mathcal{A}) \longrightarrow 1\ . \tag{6.4.3}$$

Commutative algebras have no non-trivial inner automorphisms, and noncommutative algebras have only inner automorphisms. For a canonical triple representing the spin manifold M, the group of automorphisms of the algebra $C^\infty(M)$ is isomorphic to the group of diffeomorphisms of M. For a finite-dimensional noncommutative matrix algebra the group of automorphisms is isomorphic to the group of (projective) unitary elements of the algebra.

We just saw that the concepts of outer and inner automorphisms, seem to fit nicely into the idea that the total symmetry of the space is being given as a product of spacetime diffeomorphisms and local gauge transformations. The idea of a gauge group that we will present, almost corresponds to the inner automorphisms, and the inner automorphism themselves almost corresponds to the group of unitaries. The presence of the operator J, which provides a right action of the algebra $\mathcal{A}$ on the Hilbert space $\mathcal{H}$, makes things a bit more involved.

Definition 6.4.3 — Group of unitary elements. The group of unitary elements of $\mathcal{A}$ is given by

$$\mathcal{U}(\mathcal{A}) = \{u \in \mathcal{A} \mid uu^* = u^*u = 1\}\,. \tag{6.4.4}$$

Definition 6.4.4 — Lie algebra of unitaries. The Lie algebra of unitaries of $\mathcal{A}$ is defined by

$$\mathfrak{u}(\mathcal{A}) = \{a \in \mathcal{A} \mid a^* = -a\}\,. \tag{6.4.5}$$

This is the set of skew Hermitian elements of $\mathcal{A}$.

Based on these two definitions we are now ready to define the gauge group of the algebra and gauge Lie algebra of the algebra. Our fundamental motivation for these definitions is the formula for

the fluctuated Dirac operator. To motivate these definitions even further we introduce some more concepts.

In light of the right action on $\mathcal{H}$ induced by J we define the adjoint action of unitaries on $\mathcal{H}$.

Definition 6.4.5 — Adjoint action. The adjoint action of $u \in \mathcal{U}$ on $\xi \in H$ is given by

$$adj(u)\xi = u\xi u^* = uJuJ^*\xi. \tag{6.4.6}$$

Based on these preparatory definitions we can now define the gauge group of the algebra. Notice that is has the structure of the adjoint action of unitaries.

Definition 6.4.6 — Gauge group. The gauge group of a triple is given by

$$\mathfrak{G}(\mathcal{A}, \mathcal{H}, J) = \{U = uJuJ^{-1} | u \in \mathcal{U}(\mathcal{A})\} \tag{6.4.7}$$

The gauge Lie algebra is now defined to be the Lie algebra corresponding to the Lie gauge group.

Definition 6.4.7 — Gauge Lie algebra. The gauge Lie algebra of a triple is given by

$$\mathfrak{g}(\mathcal{A}, \mathcal{H}, J) = \{X + JXJ^{-1} | X \in \mathfrak{u}(\mathcal{A})\}, \tag{6.4.8}$$

where $\mathfrak{u}(\mathcal{A})$ is the Lie algebra of skew Hermitian algebra elements.

To clarify the content of these definitions we define a useful subalgebra of $\mathcal{A}$.

Definition 6.4.8 — $\mathcal{A}_J$ subalgebra. The subalgebra $\mathcal{A}_J$ induced by J is given by

$$\mathcal{A}_J := \left\{a \in \mathcal{A} \mid a = Ja^*J^{-1}\right\} \tag{6.4.9}$$

(Note that $a = Ja^*J^{-1}$ by right multiplication by J is equivalent to $aJ = Ja^*$.)

We have the following important relations

Proposition 6.4.9 The relations between the inner automorphisms and the gauge group and the group of unitaries are given by the following diagram.

$$\begin{array}{ccccccccc} 1 & \longrightarrow & \mathcal{U}(\mathcal{A}_J) & \longrightarrow & \mathcal{U}(\mathcal{A}) & \longrightarrow & \mathfrak{G}(\mathcal{A}, \mathcal{H}, J) & \longrightarrow & 1 \\ & & \downarrow & & \| & & \downarrow & & \\ 1 & \longrightarrow & \mathcal{U}(Z(\mathcal{A})) & \longrightarrow & \mathcal{U}(\mathcal{A}) & \longrightarrow & \mathrm{Inn}(\mathcal{A}) & \longrightarrow & 1 \end{array}$$

Proof. We present a partial proof. For the top line: There is is a natural map from the group of unitaries onto to the gauge group given by $f\colon u \mapsto uJuJ^{-1}$. To make this a bijection we divide by the kernel of this map, $u \mapsto 1$, or $uJuJ^{-1} = 1$, which by right multiplication by Ju^* means $uJ = Ju^*$, which is the set of all unitaries in $\mathcal{A}_J$, $U(\mathcal{A}_J)$. For the bottom line: From the definition on inner automorphism there is a natural map from the group of unitaries onto to the inner automorphisms given by $f\colon u \mapsto \alpha_u$. To make this a bijection we divide by the kernel of this map, $u \mapsto 1$, or $uau^* = a$, which by right multiplication by u means $ua = au$, which is the set of all unitaries that commute with all elements of $\mathcal{A}$, or the set of unitaries in the center of $\mathcal{A}$, $U(Z(\mathcal{A}))$. ∎

From the diagram we can conclude that

$$\mathfrak{G}(\mathcal{A}) \simeq \mathcal{U}(\mathcal{A})/\mathcal{U}((\mathcal{A})_J) \quad \text{and} \quad \mathrm{Inn}(\mathcal{A}) \simeq \mathcal{U}(\mathcal{A})/\mathcal{U}(Z(\mathcal{A})). \tag{6.4.10}$$

Only when $\mathcal{U}((\mathcal{A})_J) \simeq \mathcal{U}(Z(\mathcal{A}))$ do we have $\mathfrak{G}(\mathcal{A}) \simeq \mathrm{Inn}(\mathcal{A})$. For the Lie algebras there is a corresponding diagram.

Proposition 6.4.10 $0 \longrightarrow \mathfrak{u}(\mathcal{A}_J) \longrightarrow \mathfrak{u}(\mathcal{A}) \longrightarrow \mathfrak{g}(\mathcal{A}, \mathcal{H}, J) \longrightarrow 0$

Proof. See [125] for a proof. ■

This means that

$$\mathfrak{g}(\mathcal{A}, \mathcal{H}, J) \simeq \mathfrak{u}(\mathcal{A})/\mathfrak{u}((\mathcal{A})_J) \tag{6.4.11}$$

We now argue that the inner fluctuation A_μ is the gauge field of the gauge group $\mathfrak{G}(\mathcal{A}, \mathcal{H}, J)$ taking values in the Lie algebra $\mathfrak{g}(\mathcal{A}, \mathcal{H}, J)$.

Theorem 6.4.11 A general inner fluctuations can be expressed as $A_\mu \in i\mathfrak{g}(\mathcal{A}, \mathcal{H}, J)$ The inner fluctuations of an AC triple can be expressed as $A_\mu(x) \in C^\infty(M, ig(\mathcal{A}_F))$.

Proof. We determined that A_μ must be self-adjoint. This means that iA_μ must be skew-adjoint. Furthermore, since $A_\mu = -ia\partial_\mu b$ we can deduce that $iA_\mu \in \mathcal{A}$. This means that iA_μ is a skew-adjoint member of $\mathcal{A}$, but this mean by definition that $iA_\mu \in \mathfrak{u}(\mathcal{A})$. This is the same as saying that $(\exists b \in \mathfrak{u}(\mathcal{A}))b = iA_\mu$ or equivalently $(\exists b \in \mathfrak{u}(\mathcal{A})) - ib = A_\mu$ or $(\exists b \in -i\mathfrak{u}(\mathcal{A}))b = A_\mu$. But since $-i\mathfrak{u} = i\mathfrak{u}$ we write $A_\mu \in i\mathfrak{u}$. We can refine this further. Since any element $a_\mu \in \mathcal{U}(\mathcal{A}_F)_{J_F}$ commutes with J_F this means that $A_\mu + a_\mu - J_F(A_\mu + a_\mu)J_F^* = A_\mu + a_\mu - a_\mu J_F J_F^* - J_F(A_\mu J_F^* = A_\mu + a_\mu - a_\mu + J_F(A_\mu J_F^* = A_\mu - J_F(A_\mu J_F^*$. Therefore we can take the equivalence classes with respect to $\mathcal{U}(\mathcal{A}_F)_{J_F}$ which becomes $\mathfrak{u}(\mathcal{A})/\mathcal{U}(\mathcal{A}_F)_{J_F}$ which is just $\mathfrak{g}(\mathcal{A}, \mathcal{H}, J)$. Thus we write $A_\mu \in i\mathfrak{g}(\mathcal{A}, \mathcal{H}, J)$. For A_μ, defined by $\gamma^\mu \otimes A_\mu = \sum a_j \left[D_M \otimes 1, b_j\right]$ we get a general expression $A_\mu(x) \in C^\infty(M, ig(\mathcal{A}_F))$. ■

The transformation properties of the inner fluctuations are what we would expect for gauge fields.

Theorem 6.4.12 — Transformation of the inner fluctuations. Under a gauge transformation $U = uJuJ^*$ the inner fluctuations A transforms as

$$A \to uAu^* + u\left[D, u^*\right] = uAu^* + u(du^*) \tag{6.4.12}$$

Proof. See [119, p195] for a proof. ■

6.5 Unimodularity

When defining the derived gauge structure on a spectral triple we have not made any provisions with regard to the representation of the algebra. It turns out there is an important aspect pertaining to the representation that needs to be dealt with by an extra step. If the representation is complex it automatically satisfies the unimodularity condition given by

$$(\forall g \in \mathfrak{G}(\mathcal{A}))(det(g) = 1). \tag{6.5.1}$$

If the representation is real this condition is not automatically satisfied. For the case of real representations we define a modified gauge group called the unimodular gauge group.

Definition 6.5.1 — Unimodular gauge group. The unimodular gauge group $\mathfrak{SG}(\mathcal{A})$ is the group defined by

$$\mathfrak{SG}(\mathcal{A})\,\{g \in \mathfrak{G}(\mathcal{A}) \mid det(g) = 1\} \tag{6.5.2}$$

We naturally also define the associated unimodular gauge Lie algebra.

Definition 6.5.2 — Unimodular gauge Lie algebra. The unimodular gauge Lie algebra is the Lie algebra defined by

$$\mathfrak{sg}(\mathcal{A},\mathcal{H},J) = \{a \in \mathfrak{g}(\mathcal{A},\mathcal{H},J) \mid \mathrm{tr}(a) = 0\}\,. \tag{6.5.3}$$

In the context of chapter 7 the concrete consequences of imposing unimodularity will be to change the gauge group of the noncommutative standard model from $\mathrm{U}(1) \oplus \mathrm{SU}(2) \oplus \mathrm{U}(3)$ to $\mathrm{U}(1) \oplus \mathrm{SU}(2) \oplus \mathrm{SU}(3)$. Note that when treating the electroweak theory with gauge group $\mathrm{U}(1) \oplus \mathrm{SU}(2)$ the representation used is also a real representation which does not satisfy unimodularity. In the isolated electroweak case unimodularity is not applied as this would give $\mathrm{SU}(1) \oplus \mathrm{SU}(2)$ which would render the $\mathrm{U}(1)$ symmetry trivial.

6.6 Action principles

As we previously mention the basic idea of creating physics in NCG is that the total set of symmetries of the physical theory is given as the diffeomorphism (i.e. automorphisms) of an AC space. This is closely related to the structure of GR, and like in GR the metric is the specifier of the geometry. It is therefore natural that the physics of the theory can be expressed through an equations for the metric. In chapter 5 we saw that in NCG the metric is expressed through the Dirac operator. In the previous section we saw how inner fluctuations led to a more general expression representing a whole class of Dirac operator. The actual physics of this will be found by developing an action principle that determines the fields that represent inner perturbations of the metric (or Dirac operator).

The basic idea of the action principles is to construct geometric invariants of the space [120].

Definition 6.6.1 — Fermionic action principle. The fermionic action principle is given by

$$S = \frac{1}{2}\langle J\psi^+, D_A\psi^+\rangle \tag{6.6.1}$$

with $\psi^+ \in H^+$ where H^+ is the $+1$ component of H projected out by γ.

Definition 6.6.2 — Bosonic action principle. The bosonic action principle is given by

$$S = \mathrm{tr}\left[F\left(\frac{D_A^2}{\Lambda^2}\right)\right] \tag{6.6.2}$$

The total action is then

$$S = \mathrm{tr}\left[F\left(\frac{D_A^2}{\Lambda^2}\right)\right] + \frac{1}{2}\langle J\psi^+, D_A\psi^+\rangle. \tag{6.6.3}$$

We consider it a basic postulate of the theory that this equation determines the complete action for both fermionic and bosonic fields [120]. The spin-1 and spin-0 bosonic fields will come from the inner perturbations of the Dirac operator. The fermionic spin-1/2 fields will come from the

finite dimensional Hilbert space of the finite triple combined with the generic spinor fields from $H_M = L^2(M, S)$.

It is important that the action principles are invariant under the gauge transformations. We state the invariance properties in two theorems.

Theorem 6.6.3 — Invariance of the fermionic action. The fermionic action principle defines a functional that is invariant under the action of the gauge group.

Proof. See [101, p122] for a proof. ■

Theorem 6.6.4 — Invariance of the bosonic action. The bosonic action principle defines a functional that is invariant under the action of the gauge group.

Proof. See [101, p122] for a proof. ■

6.7 Deriving the action

In this section we will derive the expression for the action from the two action principles. Calculating the fermionic action can be done simply by substituting the relevant expressions directly into the equation

$$S = \frac{1}{2}\langle J\psi^+, D_A\psi^+\rangle. \tag{6.7.1}$$

From the general expression for the fluctuated Dirac operator of an almost commutative space

$$D_A = D \otimes \mathbb{I} + \gamma^\mu \otimes B_\mu + \gamma_M \otimes \Phi \tag{6.7.2}$$

we see that we will get terms of the form

$$\frac{1}{2}\langle J\xi\,, D_M \otimes \mathbb{I}\xi\rangle \tag{6.7.3}$$

$$\frac{1}{2}\langle J\xi\,, \gamma^\mu \otimes B_\mu\xi\rangle \tag{6.7.4}$$

$$\frac{1}{2}\langle J\xi\,, \gamma_M \otimes \Theta\xi\rangle\,. \tag{6.7.5}$$

Reading the list from top to bottom we see that these terms correspond to kinetic fermion terms, gauge-fermion interaction terms and scalar-fermion Yukawa terms. The top two lines comes from the A_μ part of the inner fluctuations, while the third line comes from the ϕ part of the inner fluctuations. This takes care of the fermionic part of the action.

To calculate the bosonic part of the action we need to rely on approximation techniques. First we note that the square of the Dirac operator for an almost commutative spectral triple is a generalized Laplacian operator P. The operator P is of the form

$$P = -\left(g^{\mu\nu} I\partial_\mu\partial_\nu + A^\mu\partial_\mu + B\right). \tag{6.7.6}$$

The operator can be expressed as the proper Laplacian Δ with an extra factor F. We write this as

$$P = \Delta - F. \tag{6.7.7}$$

The trace of the exponential of the operator P can be expanded in an asymptotic series as [101]

$$Tr\left(e^{-tP}\right) = \sum t^{\frac{k-n}{2}} a_k(P). \tag{6.7.8}$$

The coefficients $a_k(P)$ are defined as [101]

$$a_k(P) = \int a_k(P,x)\sqrt{g}d^4x. \tag{6.7.9}$$

The coefficients $a_k(P,x)$ are called Seeley-DeWitt coefficients and the first six coefficients are given (for our setup) as [101]

$$a_0(P,x) = (4\pi)^{-\frac{n}{2}}\,\mathrm{Tr}(id) \tag{6.7.10}$$
$$a_1(P,x) = 0 \tag{6.7.11}$$
$$a_2(P,x) = (4\pi)^{-\frac{n}{2}}\,\mathrm{Tr}\left(\frac{R}{6}+F\right) \tag{6.7.12}$$
$$a_3(P,x) = 0 \tag{6.7.13}$$
$$a_4(P,x) = (4\pi)^{-\frac{n}{2}}\frac{1}{360}Tr(-12\Delta R + 5R^2 - 2R_{\mu\nu}R^{\mu\nu} + 2R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} \tag{6.7.14}$$
$$+ 60RF + 180F^2 - 60\Delta F + 30\Omega_{\mu\nu}\Omega^{\mu\nu}) \tag{6.7.15}$$
$$a_5(P,x) = 0. \tag{6.7.16}$$

The symbol F is defined above. The various R's denote as usual the Riemann curvature tensor, the Ricci tensor and the Ricci scalar. The Δ is the Laplacian (differential operator). The Ω is the curvature form of the connection. Since by a generalization of Lichernowitcz formula we can write $D_A^2 = P$, this expansion applies directly to expanding the trace of the Dirac operator. By using heat kernel methods along with the Laplace-Stieltjes transform we can derive the asymptotic expansion [101]. We get the following expression

$$S[A] = \mathrm{Tr}\left(f\left(\frac{D_A}{\Lambda}\right)\right) = \mathrm{Tr}\left(g\left(\frac{D_A^2}{\Lambda^2}\right)\right) \tag{6.7.17}$$
$$= f(0)\Lambda^0 a_4(D_A^2) + 2f_2\Lambda^2 a_2(D_A^2) + 2f_4\Lambda^4 a_0(D_A^2) + \mathcal{O}(\Lambda^{-2}). \tag{6.7.18}$$

Here Λ is a cutoff constant and $f(x) = g(x^2)$ are enveloping functions. The a_k's are the (integrated) Seeley-DeWitt coefficients as given above. Upon inserting the general form of the Dirac operator and using inner product defined on the Hilbert space the action can be expressed as[101]

$$S[g_{\mu\nu}, B_\mu, \Phi] = \int \mathrm{d}^4x\,\sqrt{g}\,\left\{N\mathcal{L}_M(g_{\mu\nu}) + \mathcal{L}_B(B_\mu) + \mathcal{L}_\phi(g_{\mu\nu}, B_\mu, \phi)\right\}. \tag{6.7.19}$$

The letter N denotes the dimension of $\mathcal{H}_F$. The parts of the action are given by

$$\mathcal{L}_M(g_{\mu\nu}) = \frac{f_4\Lambda^4}{2\pi^2} - \frac{f_2\Lambda^2}{24\pi^2}R + \frac{f(0)}{16\pi^2}\left(\frac{1}{30}\Delta R - \frac{1}{20}C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma} + \frac{11}{360}R^*R^*\right) \tag{6.7.20}$$

$$\mathcal{L}_B(B_\mu) = \frac{f(0)}{24\pi^2}Tr\left(F_{\mu\nu}F^{\mu\nu}\right) \tag{6.7.21}$$

$$\mathcal{L}_\phi(g_{\mu\nu}, B_\mu, \phi) = -\frac{2f_2\Lambda^2}{4\pi^2}Tr\left(\Phi^2\right) + \frac{f(0)}{8\pi^2}Tr\left(\Phi^4\right) + \frac{f(0)}{24\pi^2}\Delta Tr\left(\Phi^2\right)$$
$$+ \frac{f(0)}{48\pi^2}RTr\left(\Phi^2\right) + \frac{f(0)}{8\pi^2}Tr\left(\left(D_\mu\Phi^2\right)\left(D_\mu\Phi^2\right)\right). \tag{6.7.22}$$

Here R is the scalar curvature, Δ the Laplacian (differential operator), and

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + i\left[B_\mu, B_\nu\right] \tag{6.7.23}$$

is the curvature of the gauge field B_μ. We also define $D_\mu = \partial_\mu + [B_\mu, \Theta]$, and $f_j = \int_0^\infty f(v)v^{j-1}$ are the moments of the function f. The term $\frac{11}{360}R^*R^*$ is called the Pontryagin class. It is a topological term which integrates to a number proportional to the Euler characteristic of the manifold M [101]. Such a topological term does not affect dynamics. The two terms involving the Laplacian Δ are boundary terms that vanish when integrating over all of M (which is a manifold without boundary). The term $-\frac{1}{20}C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma}$ involves the Weyl tensor $C^{\mu\nu\rho\sigma}$. This is a higher order gravitational term beyond the (lowest order) Einstein-Hilbert action. Relative to the R term of the Einstein-Hilbert action it is suppressed by a factor $1/\Lambda^2$, which renders it unobservable at low energies. The term $\frac{f(0)}{48\pi^2}RTr\left(\Phi^2\right)$ couples the Higgs field to the scalar curvature. This is an example of non-minimal coupling between gravity and matter. A discussion of this term in the context of inflation can be found in [126]. When we disregard topological term and the surface term we are only left with EH-action plus the Weyl tensor term. Thus the gravity Lagrangian can be written as

$$\mathcal{L}_M(g_{\mu\nu}) = \frac{f_4\Lambda^4}{2\pi^2} - \frac{f_2\Lambda^2}{24\pi^2}R - \frac{f(0)}{320\pi^2}C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma} \tag{6.7.24}$$

$$= \frac{f_4\Lambda^4}{2\pi^2} - \frac{f_2\Lambda^2}{24\pi^2}R + \frac{f(0)}{480\pi^2}\left(R^2 - 3R_{\mu\nu}R^{\mu\nu}\right). \tag{6.7.25}$$

With these caveats stipulated we can conclude that the bosonic action principle reproduces the terms found in the Einstein-Hilbert action, plus the Higgs-Yang-Mills terms minimally coupled to gravity. We have accounted for why it is reasonable to ignore the extra terms of the action when comparing it to the actions of GR and the SM. For clarity we rewrite the Lagrangian of the bosonic action as

$$\mathcal{L}_M(g_{\mu\nu}) = \frac{f_4\Lambda^4}{2\pi^2} - \frac{f_2\Lambda^2}{24\pi^2}R \tag{6.7.26}$$

$$\mathcal{L}_B(B_\mu) = \frac{f(0)}{24\pi^2}Tr\left(F_{\mu\nu}F^{\mu\nu}\right) \tag{6.7.27}$$

$$\mathcal{L}_\phi(g_{\mu\nu}, B_\mu, \phi) = -\frac{2f_2\Lambda^2}{4\pi^2}Tr\left(\Phi^2\right) + \frac{f(0)}{8\pi^2}Tr\left(\Phi^4\right) \tag{6.7.28}$$

$$+ \frac{f(0)}{8\pi^2}Tr\left(\left(D_\mu\Phi^2\right)\left(D_\mu\Phi^2\right)\right), \tag{6.7.29}$$

where the terms we assume to be ignorable are removed. Note that this is not the final version of the action since the various terms listed are constructed from components that we have yet to analyze.

6.8 Deriving the finite algebra

We have seen that any almost-commutative space will produce a gauge theory coupled to gravity [101]. The actual physics we derive depends on the choice of the finite triple, and primarily on the choice of the finite algebra. Not all selections of non-commutative algebras and Hilbert spaces fulfill the demands of being a spectral triple. There are several limitations on what a finite spectral triple can look like. We will now look at arguments that derive a more or less unique algebra that lead to the standard model or one of its close relatives.

There exists two very different derivations of the proper finite algebra for the standard model. The first derivation has gone through two slightly different version, one with and one without, the first-order condition, and we will consider them concurrently. The second derivation proceeds by a quantum geometric argument to the Pati-Salam model and will be presented in section 6.9.

KO-dimension assumption

At the outset there are no constraints on the KO-dimension of the theory. It was long believed that is should be zero. However this choice of KO-dimension does not support neutrino oscillations and leads to doubling of the fermionic degrees of freedom (the fermion doubling problem). A selection of KO-dim equal to six (mod eight) solves both of these problems and this selection is therefore made at the outset. This choide leads to the equations (see table 6.3.9 in chapter 5)

$$J_F^2 = 1 \quad J_F D_F = D_F J_F \quad J_F \gamma_F = -\gamma_F J_F. \tag{6.8.1}$$

The irreducible finite spectral triples of KO-dimension equal to six are classified in the following theorem.

Theorem 6.8.1 For an irreducible finite real spectral triple of KO-dimension 6, then for some positive integer n, the algebra is given by

$$\mathcal{A}_F = M_n(\mathbb{C}) \oplus M_n(\mathbb{C}) \tag{6.8.2}$$

Proof. See [127] or [101, p45] for a proof. ■

Symplectic structure assumption

To proceed further we add the assumptions that there is a symplectic structure $I^2 = -1$ on $\mathcal{H}_F$. Using this symplectic structure one can show that the algebra of a spectral triple must be of the form [127]

$$\mathcal{A}_F = M_n(\mathbb{H}) \oplus M_{2n}(\mathbb{C}), \tag{6.8.3}$$

while the corresponding Hilbert space must be

$$\mathcal{H} = \mathbb{C}^{8n^2}. \tag{6.8.4}$$

Grading assumption

In this step we assume that the presence of the operator γ_F also induces a non-trivial grading on $\mathcal{A}$, and defines an even subalgebra of elements that commute with γ_F [127]. We find that the smallest non-trivial selection of n to allow such a grading is $n = 2$ [101]. With $n = 2$ we get a 32-dimensional Hilbert space (or an integer multiple of 32), and we obtain the following algebra [101]

$$\mathcal{A}_F = M_2(\mathbb{H}) \oplus M_4(\mathbb{C}). \tag{6.8.5}$$

Using the grading, the first part or the algebra can be diagonalized to give [101]

$$\mathcal{A}_F = \mathbb{H}_L \oplus \mathbb{H}_R \oplus M_4(\mathbb{C}). \tag{6.8.6}$$

If we stop at this level of the derivation the prediction for the gauge group would be

$$SU(2) \oplus SU(2) \oplus SU(4), \tag{6.8.7}$$

which is the gauge group found in the Pati-Salam model. This algebra does not satisfy the first-order condition [101].

First-order condition assumption

A generic (self-adjoint) Dirac operator on the finite space is given by [101]

$$D_F = \begin{pmatrix} S & T^* \\ T & \bar{S} \end{pmatrix}. \tag{6.8.8}$$

If we impose the condition that T is non-zero, as well as the first-order condition, this leads to the conclusion that there is a unique sub-algebra (of the even algebra) of maximal dimensions satisfying these two conditions. This algebra is given by [101]

$$\mathcal{A}_F = \mathbb{C} \oplus \mathbb{H} \oplus M_3(\mathbb{C}). \tag{6.8.9}$$

Pati-Salam model

The model above is in accordance with the standard model except for the fact that it predict a Higgs mass of 170GeV [101]. We therefore consider a modified deduction that leads to a Higgs mass compatible with experiment. This modification keeps the same algebra as a starting point, as well as all but the last step of the deduction. Instead of the last step we drop the first order condition [128]. This bring us back to the situation before we used the first-order condition assumption, namely to the algebra

$$\mathcal{A}_F = \mathbb{H}_L \oplus \mathbb{H}_R \oplus M_4(\mathbb{C}) \tag{6.8.10}$$

When we take the unitary elements of this algebra we get the gauge group

$$G_F = SU(2) \oplus SU(2) \oplus SU(4) \tag{6.8.11}$$

which is known as the Pati-Salam model [129]. The Pati-Salam model can be broken down to the standard model in several different ways [130][123].

Unimodularity assumption

The unimodularity assumption says that the determinant of the representation matrices of the algebra must be equal to 1. This assumption is automatically true for complex representations but it must be imposed as an extra condition for real representations [101][119][131, 132]. There are so far no arguments that do not use this assumption to derive the gauge group of the standard model.

Algebras beyond $n = 2$

The arguments above support the claim that the standard model gauge group is close to being uniquely singled out by the almost-commutative models. There are however bigger algebras that one can try. The choice $n = 3$ does not give a viable model, but a grand unified theory is available in the form of the algebra with $n = 4$,

$$A_F = M_4(\mathbb{H}) \oplus M_8(\mathbb{C}). \tag{6.8.12}$$

Associative algebras

The gauge group of the theory comes from the unitary elements of the algebra. Many interesting gauge groups can not be obtained as the unitary elements of an associative algebra. The associative requirement can be relaxed, thereby allowing many more algebras [133, 134, 135, 136].

6.9 A quantum geometric derivation of the algebra

Recent results have used a quantized version of the orientability condition, resulting in a generalized Heisenberg commutation relation, to get a geometrical quantization of the total volume of a space defined by a spectral triple [121][122][123]. An understanding of these new ideas can be gained by first establishing a conceptual equivalence of the Dirac operator with the momentum operator in quantum mechanics. In the canonical triple the Dirac operator is a Feynman slashed spatial derivative so this identification is natural. Once this is done one realizes that one needs to find a proper expression for the corresponding position operator. The idea is now to take this to be the Feynman slashed version Y of real scalar functions. This produces the set of functions $Y = Y_+ \oplus Y_- \in C^\infty(M, C_+ \oplus C_-)$. The C_+ and C_- represents the Clifford algebra in $D+1$ dimensions of signature all plus and all minus respectively. For $D = 4$ C_+ is isomorphic to $M_2(\mathbb{H})$ and C_- is isomorphic to $M_4(\mathbb{C})$. The direct sum $C_+ \oplus C_- = M_2(\mathbb{H}) \oplus M_4(\mathbb{C})$ is exactly the algebra we previously deduced from other considerations. This is a very interesting set of arguments which also has many other consequences that are yet to be explored. However, the whole construction still remains somewhat disjoint from general ideas on quantizing spacetime, and it is therefore too early to form a clear opinion on their potential physical relevance. The mathematical formulation is certainly beautiful and provides a novel idea towards a theory of quantum gravity.

6.10 Notes

Terminology

Models based on almost commutative space are often called spectral models or are not given a specific name. We have designated them as Connes-Lott-Chamseddine (CLC) models after the main inventors of the theory. This terminology is not standard.

The terminology of pseudo-equivalence classes is used by other authors but one should be wary of misunderstandings.

We sometimes almost equate the finite triple with the finite algebra. This is not completely accurate. The finite algebra does not uniquely determine the triple.

The term diffeomorphism does not really apply to spaces that are defined purely algebraically. We follow standard usage in NCG physics when we allow ourselves to refer to the group of automorphisms as the group of diffeomorphism. In the commutative case these two groups are isomorphic.

We sometimes write $M \times F$ to represent the AC space define by spectral triple based on the algebra $\mathcal{A}_M \otimes \mathcal{A}_F$. This is just a convenient symbolic notation. There is no product space that fully captures the algebraic definition. The F space can be related to a finite set of points (corresponding to the distinct irreducible representations of $\mathcal{A}_F$).

7. Noncommutative geometry - physics details

Summary

In this chapter the details of the derivations of various Connes-Lott-Chamseddine (CLC) models are described. We construct the CLC versions of QED, the electroweak theory and the standard model. We also briefly review the Pati-Salam model.

7.1 Introduction

In this chapter we will systematically build various noncommutative particle physics models based on almost commutative spaces. We start with the simplest possible case to illustrate the methods. The first example will be an Abelian theory with U(1) symmetry (section 7.5). We will then modify this example slightly to make it conform to the known Lagrangian of QED (section 7.6). We then study the electroweak theory with its $\mathrm{SU}(2) \oplus \mathrm{U}(1)$ gauge symmetry (section 7.7). For the electroweak case we simplify matter by considering only one generation of leptons (no quarks). Building on the study of the electroweak theory, we deduce the structure of the standard model from the algebra $\mathbb{C} \oplus \mathbb{H} \oplus M_3(\mathbb{C})$ (section 7.8). For the standard model we include both quarks and leptons, and work with three generations of particles. Finally, we look more briefly at the Pati-Salam model which is a semi grand unified theory with gauge symmetry $\mathrm{SU}(2)_L \oplus \mathrm{SU}(2)_R \oplus \mathrm{SU}(4)$ (section 7.9). Note that in this chapter we will mostly ignore the pure gravity part of the theory as well as the coupling to gravity. These gravitational aspects were dealt with in the previous chapter (see note). A Yang-Mills action will therefore typically be written simply as $S = \int d^4x F_{\mu\nu}F^{\mu\nu}$. In this chapter we will mostly follow the conventions of Walt von Suijlekom and collaborators as found e.g. in [102] and [101]. All results are previously presented by others. As there are so many itermediate results, which are all well established, we will not give individual references for each statement. All the results except those related to the Pati-Salam model can be found in [101][101] [137][138] as well as in [139][119].

7.2 Useful formulas

k	0	1	2	3	4	5	6	7	Equation
ϵ	1	1	-1	-1	-1	-1	1	1	$J^2 = \epsilon$
ϵ'	1	-1	1	1	1	-1	1	1	$JD = \epsilon' DJ$
ϵ''	1		-1		1		-1		$J\gamma = \epsilon'' \gamma J$

Table 7.1: Definition of the KO-dimension of a spectral triple.

We repeat a few formulas that will be useful. With $\epsilon, \epsilon', \epsilon'' \in \{-1, 1\}$ valued according to table 7.1 and with $b^\circ = Jb^*J^{-1}$ the following equations hold

$$\gamma^2 = 1 \quad \gamma a = a\gamma \quad \gamma D = -D\gamma \quad J^2 = \epsilon \tag{7.2.1}$$

$$JD = \epsilon' DJ \quad J\gamma = \epsilon'' \gamma J \quad [a, b^\circ] = 0 \quad [[D, a], b^\circ] = 0 \tag{7.2.2}$$

$$J^* = J^{-1} \quad JJ^* = J^*J = 1 \tag{7.2.3}$$

Any anti-linear and anti-unitary operator J can be written as $J = UC$ where C is complex conjugation operator and U is a linear unitary operator. The complex conjugation operator is basis dependent and we assume that the action on a vector $, e\rangle$ in the chosen basis is $C(\alpha, e\rangle) = \overline{\alpha}, e\rangle$. We also have

$$J^* = (UC)^* = CU^* = U^T C \quad C^* = C^{-1} = C \quad CC = 1. \tag{7.2.4}$$

Note that

$$CU = \overline{U}C \neq \overline{U}. \tag{7.2.5}$$

In informal terms C passes through another matrix (both to the left and to the right) with complex conjugation. Block matrices (with same size blocks) can be multiplied as if the blocks were matrix elements. The blocks do not have to be square. The gauge group and the gauge algebra is given by

$$\mathfrak{G}(\mathcal{A}) = \mathcal{U}(\mathcal{A})/\mathcal{U}((\mathcal{A})_J) \quad \mathfrak{g}(\mathcal{A}) = \mathfrak{u}(\mathcal{A})/\mathfrak{u}((\mathcal{A})_J) \tag{7.2.6}$$

The fluctuated Dirac operator is given by

$$D_A = D_M \otimes \mathbb{I} + \gamma^\mu \otimes B_\mu + \gamma_M \otimes \Phi \quad B_\mu = A_\mu - J_F A_\mu J_F^{-1}, \quad \Phi = D_F + \phi + J_F \phi J_F^{-1}. \tag{7.2.7}$$

$$A_\mu = \sum a_j [D_M, b_j] \quad A_\mu = A_\mu^* \quad \phi = \sum a_j [D_F, b_j] \quad \phi = \phi^* \tag{7.2.8}$$

The non-gravitational Lagrangian is expressed by

$$\mathcal{L}_B(B_\mu) = \frac{f(0)}{24\pi^2} Tr\left(B_{\mu\nu} B^{\mu\nu}\right) \quad F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + i\left[B_\mu, B_\nu\right] \tag{7.2.9}$$

$$\begin{aligned} \mathcal{L}_\phi(g_{\mu\nu}, B_\mu, \phi) = &-\frac{2f_2\Lambda^2}{4\pi^2} Tr\left(\Phi^2\right) + \frac{f(0)}{8\pi^2} Tr\left(\Phi^4\right) + \frac{f(0)}{24\pi^2} \Delta Tr\left(\Phi^2\right) \\ &+ \frac{f(0)}{48\pi^2} RTr\left(\Phi^2\right) + \frac{f(0)}{8\pi^2} Tr\left(\left(D_\mu \Phi^2\right)\left(D_\mu \Phi^2\right)\right). \end{aligned} \tag{7.2.10}$$

7.3 Notation

A short note on the matrix notation used in this chapter is in order. Matrices will often be displayed in collapsed form where the entries that are shown represent sub-matrices. One needs to be careful to distinguish whether entries are represent numbers or matrices.

The form of the matrices depend on the ordering of the chosen basis. For this reason we sometimes present a column matrix showing how the basis elements are ordered to facilitate the interpretation. We surround these column matrices by straight vertical lines. They are not intended to be read as part of the matrix equation but are just annotations to the actual equation. On a few occasions we see the need to add annotations in both the vertical and the horizontal direction. In this case the relevant information about the basis is put in the outer right column and upper row. Again these are to be considered annotations and are not part of the actual matrix.

The symbol $\mathbb{I}_n$ is used to represent an $n \times n$ diagonal matrix where all diagonal entries are equal to one. The symbol $\mathbb{I}_{+n}^{-m}$ is used to represent an $(n+m) \times (n+m)$ diagonal matrix where the first m diagonal entries are equal to minus one and the last n diagonal entries are equal to one. The meaning of similar variations should be clear.

We write the transpose of a matrix X as X^T. The overbar notation $\overline{X}$ means that all entries are complex conjugated. The star notation means the combined operation of complex conjugation and transposing.

The (non-standard) notation

$$\mathcal{H} = \begin{pmatrix} \nu_R \\ e_R \\ \nu_L \\ e_L \end{pmatrix} \tag{7.3.1}$$

will be used as a shorthand notation to mean that $\mathcal{H}$ has a basis consisting of $\{\nu_R, e_R, \nu_L, e_L\}$ and that $\mathcal{H}$ equals the (complex linear) span of those basis vectors. It will also indicated the given ordering is the ordering of the basis with respect to which we are writing any matrix operating on $\mathcal{H}$.

The notation (ψ, ξ) denotes the fiberwise inner product on $\mathcal{H}_M = L^2(S, M)$. The inner product on $\mathcal{H}_M = L^2(S, M)$ is denoted $\langle \psi, \xi \rangle$. The inner product on $\mathcal{H} = \mathcal{H}_M \otimes \mathcal{H}_F$ is denoted $\langle \psi \otimes e, \xi \otimes \nu \rangle$.

7.4 Some general preparations

Before we introduce the specifics of the $\mathrm{U}(1)$ theory we introduce some preliminary observations. These observations are of a general nature and will in principle be true for any finite triple. However, they do depend on a specific way of organizing the basis of the Hilbert space (with respect to the grading operator γ_F). This way of organizing the Hilbert space will only be used for the $\mathrm{U}(1)$ theory. In later sections we organize the basis differently, and some of the findings of this section will therefore not apply.

A generic finite spectral triple is written

$$(\mathcal{A}_F, \mathcal{H}_F, D_F, J_F, \gamma_F)\,. \tag{7.4.1}$$

The matrix γ_F defines a $\mathbb{Z}_2$ grading of the n-dimensional vector space $\mathcal{H}_F$. This means that we can decompose $\mathcal{H}_F$ as $\mathcal{H}_F = \mathcal{H}_F^+ \oplus \mathcal{H}_F^-$. We cannot in general determine the dimensions k and l of $\mathcal{H}_F^+$ and $\mathcal{H}_F^-$, but we know that $k + l = n$. If we use an organization of the Hilbert space basis

based on $\mathcal{H}_F^+$ and $\mathcal{H}_F^-$, we can write γ_F as

$$\gamma_F = \begin{pmatrix} I_k & 0 \\ 0 & -I_l \end{pmatrix}, \tag{7.4.2}$$

where k and l are the dimensions of $\mathcal{H}_F^+$ and $\mathcal{H}_F^-$. The matrix D_F must be a $n \times n$ self-adjoint matrix. Any self-adjoint $n \times n$ matrix can be written as

$$D_F = \begin{pmatrix} S & T \\ T^* & S' \end{pmatrix}, \tag{7.4.3}$$

for S and S' self-adjoint $k \times k$ and $l \times l$ matrices respectively. In the basis given above the requirement $\gamma_F D_F = -D_F \gamma_F$, which is independent of KO-dimension, restricts D_F to be of the form

$$D_F = \begin{pmatrix} 0 & T \\ T^* & 0 \end{pmatrix}. \tag{7.4.4}$$

We will generally assume that we have both a real structure J_F and a $\mathbb{Z}_2$ grading γ_F, which restrict us to KO-dimensions 0, 2, 4 and 6. The form of J_F depends on the KO-dimension. With the assumptions given above, and for KO-dimension equal to zero, J_F has the form

$$J_F = \begin{pmatrix} S & 0 \\ 0 & S' \end{pmatrix} C, \tag{7.4.5}$$

with S and S' symmetric ($S^T = S, S'^T = S'$) and unitary. For KO-dimension equal to two, J_F has the form

$$J_F = \begin{pmatrix} 0 & T \\ -T^T & 0 \end{pmatrix} C. \tag{7.4.6}$$

with C unitary. For KO-dimension equal to four, J_F has the form

$$J_F = \begin{pmatrix} S & 0 \\ 0 & S' \end{pmatrix} C, \tag{7.4.7}$$

with S and S' anti-symmetric ($S^T = -S, S'^T = -S'$) and unitary. For KO-dimension equal to six, J_F has the form

$$J_F = \begin{pmatrix} 0 & T \\ T^T & 0 \end{pmatrix} C. \tag{7.4.8}$$

with C unitary. To complete the analysis one must check that $J_F D_F = D_F J_F$ as well as checking the zeroth-order and the first-order conditions given by

$$\pi(a) J_F \pi(b)^* J_F^* = J_F \pi(b)^* J_F^* \pi(a) \tag{7.4.9}$$

$$(D_F \pi(a) - \pi(a) D_F)\, J_F \pi(b)^* J_F^* = J_F \pi(b)^* J_F^*\, (D_F \pi(a) - \pi(a) D_F). \tag{7.4.10}$$

We will return to these conditions later.

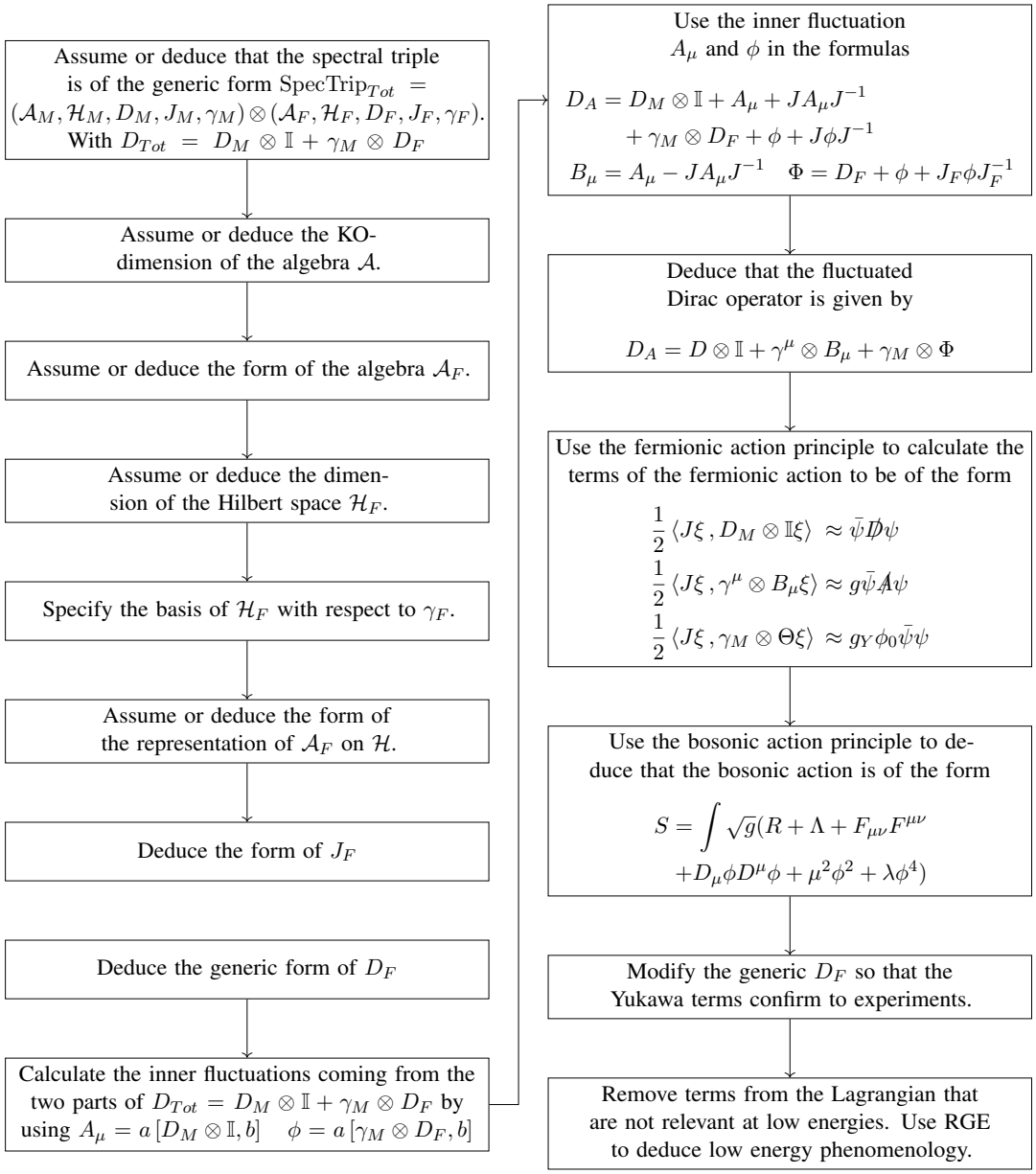


Figure 7.1: Deriving the NCG action.

7.5 U(1) gauge theory

Hilbert space and algebra representation

The simplest non-trivial finite space that can be set up is a space containing two points. The spectral triple in this case is given by

$$\left(\mathcal{A}_F = \mathbb{C}^2, \mathcal{H}_F, D_F, J_F, \gamma_F\right). \tag{7.5.1}$$

The two points correspond to the two distinct irreducible representations of the commutative algebra $\mathbb{C} \oplus \mathbb{C}$. To progress we need to establish the simplest possibilities for the other components of the triple. A faithful representation of A_F requires $\dim(H_F) \geq 2$ and hence we take $\mathcal{H}_F = \mathbb{C}^2$. The operator γ_F has eigenvalues in $\{-1, 1\}$, thus just as our preparatory section we can choose a basis such that γ_F is diagonal with only such values along the diagonal. Choosing $\mathrm{diag}\,(1, 1)$

or $\mathrm{diag}\,(-1,-1)$ leads to D_F being a matrix of zeros. This comes directly from the fact that we require $D_F\gamma_F = -\gamma_F D_F$. To have the option of a non-zero Dirac operator we therefore choose

$$\gamma_F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{7.5.2}$$

An element a of the algebra $\mathcal{A}_F$ is written $a = (a_1, a_2) \in \mathbb{C}^2$. We represent a as an operator on $\mathcal{H}_F = \mathcal{H}_F^+ \oplus \mathcal{H}_F^-$ by

$$\pi(a) \begin{pmatrix} \mathcal{H}_F^+ \\ \mathcal{H}_F^- \end{pmatrix} = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} \begin{pmatrix} \mathcal{H}_F^+ \\ \mathcal{H}_F^- \end{pmatrix}. \tag{7.5.3}$$

The finite Dirac operator

The $D_F\gamma_F = -\gamma_F D_F$ criteria restricts D_F to be

$$D_F = \begin{pmatrix} 0 & d \\ \overline{d} & 0 \end{pmatrix}. \tag{7.5.4}$$

The spectral triple so far is

$$\left(\mathcal{A}_F = \mathbb{C}^2, \mathcal{H}_F = \mathbb{C}^2, D_F = \begin{pmatrix} 0 & d \\ \overline{d} & 0 \end{pmatrix}, J_F, \gamma_F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right). \tag{7.5.5}$$

The charge conjugation operator

The only remaining undetermined element is the J_F matrix. Since we already require there to be a $\mathbb{Z}_2$-grading γ_F we are restricted (by definition) to KO-dimensions 0, 2, 4 and 6. We now check J_F case-by-case for the possible KO-dimensions. Let a, b and c be complex unitary 1×1 matrices, which means they are numbers of the form $e^{i\theta}$. For KO-dimension 0 we find J_F to be given by the matrix

$$J_F = \begin{pmatrix} j_1 & 0 \\ 0 & j_2 \end{pmatrix} C, \tag{7.5.6}$$

with j_1 and j_2 symmetric. The transpose of a number (i.e. a 1×1 matrix) is obviously equal to the number itself, therefore any 1×1 matrix is symmetric. The conjugate transpose is then found to be

$$J_F^* = \begin{pmatrix} \overline{j}_1 & 0 \\ 0 & \overline{j}_2 \end{pmatrix} C. \tag{7.5.7}$$

We can now check the zeroth-order condition

$$\pi(a) J_F \pi(b)^* J_F^* = J_F \pi(b)^* J_F^* \pi(a). \tag{7.5.8}$$

These are all diagonal complex valued matrices and they therefore all commute and the zeroth-order condition is true. Since these matrices all commute we get

$$\pi(b)^\circ = J_F \pi(b)^* J_F^* = J_F J_F^* \pi(b)^* = \pi(b)^* = \begin{pmatrix} \overline{b}_1 & 0 \\ 0 & \overline{b}_2 \end{pmatrix}. \tag{7.5.9}$$

Using this we can check the first-order condition

$$[[D_F, \pi(a)], \pi(b)^\circ] = 0. \tag{7.5.10}$$

The first part is calculated to be

$$[D_F, \pi(a)] = \begin{pmatrix} 0 & d \\ \overline{d} & 0 \end{pmatrix} \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} - \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} \begin{pmatrix} 0 & d \\ \overline{d} & 0 \end{pmatrix} \tag{7.5.11}$$

$$= \begin{pmatrix} 0 & (a_2 - a_1)d \\ (a_1 - a_2)\overline{d} & 0 \end{pmatrix}. \tag{7.5.12}$$

The whole expression now becomes

$$[[D_F, \pi(a)], \pi(b)^\circ] = \begin{pmatrix} 0 & (a_2 - a_1)d \\ (a_1 - a_2)\overline{d} & 0 \end{pmatrix} \begin{pmatrix} \overline{b}_1 & 0 \\ 0 & \overline{b}_2 \end{pmatrix} \tag{7.5.13}$$

$$- \begin{pmatrix} \overline{b_1} & 0 \\ 0 & \overline{b_2} \end{pmatrix} \begin{pmatrix} 0 & (a_2 - a_1)d \\ (a_1 - a_2)\overline{d} & 0 \end{pmatrix} \tag{7.5.14}$$

$$= \begin{pmatrix} 0 & (a_1 - a_2)(\overline{b}_1 - \overline{b}_2)d \\ (a_1 - a_2)(\overline{b}_1 - \overline{b}_2)\overline{d} & 0 \end{pmatrix} \tag{7.5.15}$$

$$= 0. \tag{7.5.16}$$

Since the a's and b's are arbitrary complex numbers we conclude that $d = \overline{d} = 0$, which means that $D_F = 0$. For KO-dimensions 2 and 6 basically the same conclusion follows. For KO-dimension = 4 we get

$$J_F = \begin{pmatrix} j_1 & 0 \\ 0 & j_2 \end{pmatrix} C, \tag{7.5.17}$$

with j_1 and j_2 anti-symmetric. Minus the transpose of a number (i.e. a 1×1 matrix) is obviously equal to the minus the number itself, therefore no non-zero 1×1 matrix is anti-symmetric. We must conclude that $j_1 = j_2 = 0$ and then $J_F = 0$, but this is not allowed by the definition of J_F, and we must conclude that KO-dimension 4 does not allow for any J_F. Since we require a J_F, KO-dimension 4 is not possible for this spectral triple. We are hence left with KO-dimensions 0,2 and 6, and they all imply that we must have $D_F = 0$.

The gauge group of a finite triple is given by

$$G = \mathcal{U}(\mathcal{A}_\mathcal{F}) / \mathcal{U}((\mathcal{A}_\mathcal{F})_J) \tag{7.5.18}$$

To establish what the group is for our finite triple, it is necessary to calculate the elements of the algebra $(\mathcal{A}_F)_J$ defined by $(\mathcal{A}_F)_J := \{a \mid a = J_F a^* J_F^*\}$ for the various KO-dimensions. For KO-dimension 0 the defining criteria for an element of $(\mathcal{A}_F)_J$ becomes

$$\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} = \begin{pmatrix} j_1 & 0 \\ 0 & j_2 \end{pmatrix} C \begin{pmatrix} \overline{a}_1 & 0 \\ 0 & \overline{a}_2 \end{pmatrix} C \begin{pmatrix} \overline{j}_1 & 0 \\ 0 & \overline{j}_2 \end{pmatrix}. \tag{7.5.19}$$

The right-hand side is calculated to be

$$\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}, \tag{7.5.20}$$

which is the same as the expression for an arbitrary element of $\mathcal{A}_F$ and thus this makes $(\mathcal{A}_F)_J = \mathcal{A}_F$. In this case the gauge group is equal to the trivial group. For KO-dimension 2 the defining criteria becomes

$$\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} = \begin{pmatrix} 0 & j \\ -j & 0 \end{pmatrix} C \begin{pmatrix} \overline{a}_1 & 0 \\ 0 & \overline{a}_2 \end{pmatrix} C \begin{pmatrix} 0 & -\overline{j} \\ \overline{j} & 0 \end{pmatrix} \tag{7.5.21}$$

$$= \begin{pmatrix} a_2 & 0 \\ 0 & a_1 \end{pmatrix}, \tag{7.5.22}$$

which is true if $a_1 = a_2$. This implies that $(\mathcal{A}_F)_J = \mathbb{C}$. In this case the gauge group is

$$G = \mathcal{U}(\mathcal{A}_\mathcal{F}) / \mathcal{U}((\mathcal{A}_\mathcal{F})_J) \tag{7.5.23}$$
$$= U(1) \times U(1)/U(1) = U(1). \tag{7.5.24}$$

For KO-dimension 6 the defining criteria becomes

$$\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} = \begin{pmatrix} 0 & j \\ j & 0 \end{pmatrix} C \begin{pmatrix} \overline{a}_1 & 0 \\ 0 & \overline{a}_2 \end{pmatrix} C \begin{pmatrix} 0 & \overline{j} \\ \overline{j} & 0 \end{pmatrix} \tag{7.5.25}$$
$$= \begin{pmatrix} a_2 & 0 \\ 0 & a_1 \end{pmatrix}, \tag{7.5.26}$$

which is true if $a_1 = a_2$. This implies that $(\mathcal{A}_F)_{J_F} = \mathbb{C}$. In this case the gauge group is again given by

$$G = \mathcal{U}(\mathcal{A}_\mathcal{F}) / \mathcal{U}((\mathcal{A}_\mathcal{F})_J) \tag{7.5.27}$$
$$= U(1) \times U(1)/U(1) = U(1). \tag{7.5.28}$$

This means that KO-dimension 2 and 6 both define $U(1)$ gauge theories and KO dimension 0 gives us just the trivial group. We now select KO-dimension 6 for further analysis. (At this point this selection is an arbitrary choice, but it is motivated by later developments which make it clear that only KO-dimension 6 is physically acceptable.) With the proviso of KO-dimension equal to 6, we have found the following spectral triple

$$\left(\mathcal{A}_F = \mathbb{C}^2, \mathcal{H}_F = \mathbb{C}^2, D_F = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, J_F = \begin{pmatrix} 0 & j \\ j & 0 \end{pmatrix} C, \gamma_F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right). \tag{7.5.29}$$

We choose to always rescale the matrix J_F such that its unitary part is represented by zeros and plus or minus one. We also write the algebra and the Hilbert space in their explicit representations. The final version of the triple is

$$\left(\mathcal{A}_F = \mathbb{C}^2, \mathcal{H}_F = \mathbb{C}^2, D_F = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, J_F = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} C, \gamma_F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right) \tag{7.5.30}$$

with representation of $\mathcal{A}$ given by

$$\mathbb{C}^2 \ni (a_1, a_2) \mapsto \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} \text{ acting on } \begin{pmatrix} e \\ \overline{e} \end{pmatrix} \in H_F. \tag{7.5.31}$$

Gauge fields

To determine the gauge algebra of the finite algebra we use the expression $\mathfrak{g} = \mathfrak{u}(\mathcal{A}_F) / \mathfrak{u}\left((\mathcal{A}_\mathcal{F})_{J_F}\right)$. From now on we use a for both a itself and the matrix representation $\pi(a)$ to make the notation more compact. One must rely on context for the interpretation but a will usually mean $\pi(a)$ except when discussing the representations. The Lie algebra of the gauge group is defined to

$$\mathfrak{u}(\mathcal{A}_F) = \{a \mid a^* = -a\}, \tag{7.5.32}$$

which means that

$$\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} = \begin{pmatrix} -\overline{a}_1 & 0 \\ 0 & -\overline{a}_2 \end{pmatrix}. \tag{7.5.33}$$

This is true if a_1 and a_2 are pure complex numbers. This can be expressed as $u\left(\mathcal{A}_F\right) = i\mathbb{R} \oplus i\mathbb{R}$. From earlier we have that

$$\left(\mathcal{A}_F\right)_{J_F} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \tag{7.5.34}$$

which means that for $\mathfrak{u}\left(\left(\mathcal{A}_{\mathcal{F}}\right)_{J_F}\right)$ we get

$$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} = \begin{pmatrix} -\overline{a} & 0 \\ 0 & -\overline{a} \end{pmatrix}. \tag{7.5.35}$$

This is only true if a is a pure complex number. This can be expressed as $\mathfrak{u}\left(\mathcal{A}_F\right)_{J_F} = i\mathbb{R}$. The Lie algebra of the finite triple thus becomes

$$\mathfrak{g}_F \coloneqq \mathfrak{g}\left(\mathcal{A}_F\right) = \mathfrak{u}\left(\mathcal{A}_F\right) / \mathfrak{u}\left(\left(\mathcal{A}_{\mathcal{F}}\right)_{J_F}\right) = (i\mathbb{R} \oplus i\mathbb{R})/i\mathbb{R} = i\mathbb{R}. \tag{7.5.36}$$

We infer from this that the gauge field is represented by $A_\mu \in (C^\infty(M, i\mathfrak{g}_F)$, where $i\mathfrak{g}_F = \mathbb{R}$. Alternatively we could go back to the foundations and reason by the use of the expression

$$\Omega^1_D(\mathcal{A}) = \left\{ \sum_i a_i \left[D, b_i\right] \mid a_i, b_i \in \mathcal{A} \right\}. \tag{7.5.37}$$

We know that gauge fields are self-adjoint members of $\Omega^1_D(\mathcal{A})$ which means that we must have $A_\mu \in \Omega^1_D(\mathcal{A})$ and $A^*_\mu = A_\mu$. Since $D_F = 0$ we have $D = D_M \otimes \mathbb{I} + \gamma_M \otimes D_F = D_M$, and from this we get

$$\Omega^1_D(\mathcal{A}) = \Omega^1_{D_M}(\mathcal{A}) = \left\{ \sum_i a_i \left[D_M, b_i\right] \mid a_i, b_i \in \mathcal{A} \right\}. \tag{7.5.38}$$

To proceed we calculate this expression and get

$$a_i \left[D_M, b_i\right] = \left(\alpha_i \otimes A_i\right) \left[D_M \otimes \mathbb{I}, \left(\beta_i \otimes B_i\right)\right] = i\alpha_i \gamma^\mu \partial_\mu \beta_i \otimes A_i B_i. \tag{7.5.39}$$

This expression is often written just

$$\Omega^1_{D_M}(\mathcal{A}) = \left\{ \gamma^\mu \otimes A_\mu \mid A_\mu \coloneqq i\alpha_i \partial_\mu \beta_i \otimes A_i B_i = i a_i \partial_\mu b \right\}. \tag{7.5.40}$$

This is possible since $f_\mu \coloneqq i\alpha_i \partial_\mu \beta_i$ is a set of (four) functions $-if_\mu \in \mathcal{A}_M$, and $A_i B_i$ is an element of $\mathcal{A}_F$. We write $A_\mu \in C^\infty(M) \otimes \mathbb{C}^2 = C^\infty(M, \mathbb{C}^2) = C^\infty(M, \mathbb{C}) \oplus C^\infty(M, \mathbb{C})$. This means that $A_\mu = (X^1_\mu, X^2_\mu)$ for to complex functions X^1_μ, X^2_μ on M. To have $A_\mu = A^*_\mu$ the two functions must be real-valued, and we must have $A_\mu \in C^\infty(M, \mathbb{R}) \oplus C^\infty(M, \mathbb{R})$. This means that A_μ can be written as

$$A_\mu = \begin{pmatrix} X^1_\mu & 0 \\ 0 & X^2_\mu \end{pmatrix} \tag{7.5.41}$$

We are interested in the full Dirac operator which is given in this case by

$$\begin{aligned} D_A &= D_M + \gamma^\mu \otimes A_\mu + J\left(\gamma^\mu \otimes A_\mu\right) J^* && (7.5.42)\\ &= D_M + \gamma^\mu \otimes A_\mu - \gamma^\mu \otimes J_F A_\mu J_F^* && (7.5.43)\\ &= D_M + \gamma^\mu \otimes A_\mu - \gamma^\mu \otimes A_\mu^T && (7.5.44)\\ &= D_M + \gamma^\mu \otimes \left(\begin{pmatrix} X_\mu^1 & 0 \\ 0 & X_\mu^2 \end{pmatrix} - \begin{pmatrix} X_\mu^2 & 0 \\ 0 & X_\mu^1 \end{pmatrix}\right) && (7.5.45)\\ &= D_M + \gamma^\mu \otimes \begin{pmatrix} X_\mu^2 - X_\mu^1 & 0 \\ 0 & X_\mu^1 - X_\mu^2 \end{pmatrix} && (7.5.46)\\ &= D_M + \gamma^\mu \otimes \begin{pmatrix} Y_\mu & 0 \\ 0 & -Y_\mu \end{pmatrix} && (7.5.47)\\ &= D_M + \gamma^\mu Y_\mu \otimes \gamma_F && (7.5.48) \end{aligned}$$

We write this as $D_M \otimes \mathbb{I} + \gamma^\mu \otimes B_\mu$ with $B_\mu = Y_\mu \otimes \gamma_F$. By this direct computation we arrive at the same definition of the gauge field as we did when working from the definitions of the gauge group and gauge algebra. This is of course by design since we defined the gauge group and gauge group Lie algebra exactly to represent the inner fluctuations of the Dirac operator.

Field strength and traces

The fields strength of the gauge field, and the associated traces, are easy to define. We have the standard definition of $B_{\mu\nu}$ as

$$B_{\mu\nu} \coloneqq \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{7.5.49}$$

We define $Y_{\mu\nu}$ as

$$Y_{\mu\nu} \coloneqq \partial_\mu Y_\nu - \partial_\nu Y_\mu. \tag{7.5.50}$$

The field $B_{\mu\nu}$ can be expressed in terms of $Y_{\mu\nu}$.

We note that the commutator part is absent since this is a commuting scalar field. The field $Y_{\mu\nu}$ is just a number and the trace is equal to the number itself. The trace of $B_{\mu\nu}B^{\mu\nu}$ is just $2Y_{\mu\nu}Y^{\mu\nu}$ since $B_{\mu\nu}$ is square and the minus signs cancel. The finite Dirac operator D_F is zero. Hence the fields ϕ and Φ are both zero. Any traces involving Φ are also zero.

Bosonic action

We have all the ingredients necessary to compute the bosonic action. Let us rename $Y_{\mu\nu}$ to the more conventional $F_{\mu\nu}$. The bosonic action becomes

$$\begin{aligned} S_b &= \int \mathrm{d}^4x\,\sqrt{g}\,\left\{2L_M + \frac{f(0)}{24\pi^2} Tr\left[B_{\mu\nu}B^{\mu\nu}\right]\right\} && (7.5.51)\\ &= \int \mathrm{d}^4x\,\sqrt{g}\,\left\{2L_M + \frac{f(0)}{24\pi^2}\left(2F_{\mu\nu}F^{\mu\nu}\right)\right\}. && (7.5.52) \end{aligned}$$

We observe that there are no scalar fields in the action since all terms related to Φ are zero.

Fermionic action

To write out the fermionic action we first need to get an expression for the +1 chiral subspace of the Hilbert space. The operator γ_M provides a Z_2 grading of $\mathcal{H}_M = L^2(M,S)$, so that we can write $L^2(M,S) = L^2(M,S)^+ \oplus L^2(M,S)^-$. Note that the elements of $L^2(M,S)$ are four-component

Dirac spinors and the elements of $L^2(M,S)^+$ and $L^2(M,S)^-$ are two-component Weyl spinors. Since we have $\mathcal{H}_F = \mathcal{H}_F^+ \oplus \mathcal{H}_F^-$ and $\mathcal{H} = \mathcal{H}_M \otimes \mathcal{H}_F$, we determine the $+1$ chiral subspace with respect to the operator $\gamma_M \otimes \gamma_F$ to be

$$\mathcal{H}^+ = \left(L^2(M,S)^+ \otimes \mathcal{H}_F^+\right) \oplus \left(L^2(M,S)^- \otimes \mathcal{H}_F^-\right). \tag{7.5.53}$$

We can express a generic vector in this space as

$$\xi = \psi_L \otimes e + \psi_R \otimes \overline{e} \tag{7.5.54}$$

with $\psi_L \in L^2(M,S)^+, \psi_R \in L^2(M,S)^-$ and $e \in \mathcal{H}_F^+, \overline{e} \in \mathcal{H}_F^-$. The fermionic action is then expressed as

$$S_f = \frac{1}{2}\left\langle J\xi\,, D_A\xi\right\rangle = \frac{1}{2}\left\langle J\xi\,, \left(D_M \otimes \mathbb{I} + \gamma^\mu \otimes B_\mu\right)\xi\right\rangle. \tag{7.5.55}$$

We calculate this in several steps. We do the left-hand side first.

$$\begin{aligned} J\xi &= \left(J_M \otimes J_F\right)\left(\psi_L \otimes e + \psi_R \otimes \overline{e}\right) && (7.5.56)\\ &= J_M\psi_L \otimes J_F e + J_M\psi_R \otimes J_F\overline{e} && (7.5.57)\\ &= J_M\psi_L \otimes \overline{e} + J_M\psi_R \otimes e. && (7.5.58)\end{aligned}$$

Then we calculate the first part of the right-hand side.

$$\begin{aligned} \left(D_M \otimes \mathbb{I}\right)\xi &= \left(D_M \otimes \mathbb{I}\right)\left(\psi_L \otimes e + \psi_R \otimes \overline{e}\right) && (7.5.59)\\ &= D_M\psi_L \otimes e + D_M\psi_R \otimes \overline{e}. && (7.5.60)\end{aligned}$$

Finally we calculate the second part of the right-hand side.

$$\begin{aligned} \left(\gamma^\mu \otimes B_\mu\right)\xi &= \left(\gamma^\mu \otimes B_\mu\right)\left(\psi_L \otimes e + \psi_R \otimes \overline{e}\right) && (7.5.61)\\ &= \gamma^\mu\psi_L \otimes Y_\mu e - \gamma^\mu\psi_R \otimes Y_\mu\overline{e}. && (7.5.62)\end{aligned}$$

The presence of a γ_F causes a sign switch in the last term. We now have the complete expression for the fermionic action.

$$\begin{aligned} S_f &= \frac{1}{2}\left\langle J\xi\,, D_A\xi\right\rangle && (7.5.63)\\ &= \frac{1}{2}\left\langle J_M\psi_L \otimes \overline{e} + J_M\psi_R \otimes e\,, D_M\psi_L \otimes e + D_M\psi_R \otimes \overline{e} + \gamma^\mu\psi_L \otimes Y_\mu e - \gamma^\mu\psi_R \otimes Y_\mu\overline{e}\right\rangle && (7.5.64)\\ &= \frac{1}{2}\int d^4x\sqrt{g}\left[\overline{\psi}_R\left(D_M - \gamma^\mu Y_\mu\right)\psi_R + \overline{\psi}_L\left(D_M - \gamma^\mu Y_\mu\right)\psi_L\right]. && (7.5.65)\end{aligned}$$

Where we have used definition 5.7.16 .

In summary we have constructed a spectral triple with $\mathrm{U}(1)$ gauge symmetry, which gives an action representing a quantum field theory of massless Weyl spinors interacting with a spin-1 $\mathrm{U}(1)$-gauge field. This is not quite equivalent to QED, and we now proceed to correct the construction to make a spectral triple that reproduces QED.

7.6 QED

In the previous section we produced a theory that was similar to QED. It fell slightly short of reproducing QED since it had only massless fermions, and too few fermionic degrees of freedom.

We will keep the algebra from the previous section, but we now represent it on a four dimensional Hilbert space (instead of the previous two dimensional space). Unlike in the previous section, this will allow a non-trivial Dirac operator. With a non-zero Dirac operator we can get mass terms for fermions. This will also double the number of degrees of freedom in the fermionic sector. Together these changes will allow us to reproduce the well-known QED theory of interacting electrons and positrons. We only analyze the KO-dimension = 6 case and we start with the spectral triple

$$\left(\mathcal{A}_F = \mathbb{C}^2, \mathcal{H}_F = \mathbb{C}^4, D_F, J_F, \gamma_F\right) \tag{7.6.1}$$

with representation on $\mathcal{H}_F$ given by

$$\mathbb{C}^2 \ni (a_1, a_2) \mapsto \begin{pmatrix} a_1 & 0 & 0 & 0 \\ 0 & a_1 & 0 & 0 \\ 0 & 0 & a_2 & 0 \\ 0 & 0 & 0 & a_2 \end{pmatrix} \begin{vmatrix} e_L \\ e_R \\ \overline{e}_L \\ \overline{e}_R \end{vmatrix}. \tag{7.6.2}$$

Note that the structure of Hilbert space basis is now different from before. This will lead to new forms of Dirac operator and more. By similar deductions as in the previous section we get the following components

$$D_F = \begin{pmatrix} 0 & d & 0 & 0 \\ \overline{d} & 0 & 0 & 0 \\ 0 & 0 & 0 & \overline{d} \\ 0 & 0 & d & 0 \end{pmatrix} \quad J_F = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} C \quad \gamma_F = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{7.6.3}$$

We can write some om the matrices in a more compact form as

$$\mathcal{A}_F = \begin{pmatrix} a_1 I_2 & 0 \\ 0 & a_2 I_2 \end{pmatrix} \quad J_F = \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix} C \quad J_F^* = \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix} C. \tag{7.6.4}$$

Gauge fields

We will now work out the form of the gauge fields. Using the simplified form of the matrices from equation 7.6.4 one finds by inspection that

$$-J_F a^* J_F^* = \begin{pmatrix} -a_2 I_2 & 0 \\ 0 & -a_1 I_2 \end{pmatrix}. \tag{7.6.5}$$

The gauge field matrix B_μ can be expressed by a single field Y_μ as

$$B_\mu = \begin{pmatrix} Y_\mu & 0 & 0 & 0 \\ 0 & Y_\mu & 0 & 0 \\ 0 & 0 & -Y_\mu & 0 \\ 0 & 0 & 0 & -Y_\mu \end{pmatrix}. \tag{7.6.6}$$

The fluctuated Dirac operator for any AC triple is given as

$$D_A = D_M \otimes \mathbb{I} + \gamma^\mu \otimes B_\mu + \gamma_M \otimes \Phi. \tag{7.6.7}$$

In this case D_F given in equation 7.6.3 commutes with the matrix representing (a_1, a_2). This means that $\phi = [D_F, a] = 0$. Thus we have that Φ is simply

$$\Phi = D_F + \phi + J_F \phi J_F^* = D_F, \tag{7.6.8}$$

and from this we get that D_A can be written as

$$D_A = D_M \otimes \mathbb{I} + \gamma^\mu \otimes B_\mu + \gamma_M \otimes D_F. \tag{7.6.9}$$

Field strength and traces

There are no complications with regards to the fields strength and the associated traces. We have the standard definition of $B_{\mu\nu}$ as

$$B_{\mu\nu} \coloneqq \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{7.6.10}$$

We define $Y_{\mu\nu}$ as

$$Y_{\mu\nu} \coloneqq \partial_\mu Y_\nu - \partial_\nu Y_\mu. \tag{7.6.11}$$

The field $B_{\mu\nu}$ can be expressed in terms of $Y_{\mu\nu}$. We note that the commutator part is absent since this is just a commuting scalar field. The field $Y_{\mu\nu}$ is just a number and the trace is equal to the number itself. The trace of $B_{\mu\nu}B^{\mu\nu}$ is just $4Y_{\mu\nu}Y^{\mu\nu}$ (since the representation is four-dimensional and the minuses cancel because the equation is quadratic).

The finite Dirac operator D_F is non-zero but commutes with the algebra and hence $\Phi = D_F$. This means that

$$Tr(\Phi^2) = Tr(D_F^2) = 4|d|^2 \tag{7.6.12}$$

$$Tr(\Phi^4) = Tr(D_F^4) = 4|d|^4 \tag{7.6.13}$$

$$\mathrm{Tr}\left(D_\mu\Phi\, D^\mu\Phi\right) = 0. \tag{7.6.14}$$

Bosonic action

The bosonic action becomes the usual GR action together with the Yang-Mills kinetic action, and some extra constant terms which are absorbed into the cosmological constant. We rename $Y_{\mu\nu}$ to the more conventional $F_{\mu\nu}$. Explicitly the bosonic action is

$$S_b = \int \mathrm{d}^4x\,\sqrt{g}\,\left\{4L_M + \frac{f(0)}{24\pi^2}Tr\left[B_{\mu\nu}B^{\mu\nu}\right] - \frac{2f_2\Lambda^2}{4\pi^2}4d^2 + \frac{f(0)}{8\pi^2}4d^4 + \frac{f(0)}{48\pi^2}R4d^2\right\} \tag{7.6.15}$$

$$= \int \mathrm{d}^4x\,\sqrt{g}\,\left\{4L_M + \frac{f(0)}{24\pi^2}4F_{\mu\nu}F^{\mu\nu} - \frac{2f_2\Lambda^2}{\pi^2}d^2 + \frac{f(0)}{2\pi^2}d^4 + \frac{f(0)}{12\pi^2}Rd^2\right\}. \tag{7.6.16}$$

Fermionic action

A general vector in the Hilbert space is now a combination $\mathcal{H} = \mathcal{H}_M \otimes \mathcal{H}_F$ of $L^2(M,S) = L^2(M,S)^+ \oplus L^2(M,S)^-$ and $\mathcal{H}_F = \mathcal{H}_F^+ \oplus \mathcal{H}_F^-$. For the $+1$ chiral subspace of we get

$$\mathcal{H}^+ = \left(L^2(M,S)^+ \otimes \mathcal{H}_F^+\right) \oplus \left(L^2(M,S)^- \otimes \mathcal{H}_F^-\right) \tag{7.6.17}$$

We can express a generic vector in this space as $+$ with $+$, and $-$ with $-$ combinations.

$$\xi = \chi_L \otimes e_L + \chi_R \otimes e_R + \psi_R \otimes \overline{e}_L + \psi_L \otimes \overline{e}_R \tag{7.6.18}$$

The fermionic action becomes

$$S = \langle J_M\chi\,, D\psi\rangle + \langle J_M\chi\,, \gamma^\mu Y_\mu\psi\rangle + \langle J_M\chi_L\,, \overline{d}\psi_L\rangle - \langle J_M\chi_R\,, d\psi_R\rangle\,. \tag{7.6.19}$$

Where $\chi = \chi_R + \chi_L$ and $\psi = \psi_r + \psi_L$. We give physical meaning to the matrix elements d of D_F by defining a real mass parameter $m = -i\,d$. This matrix element will result in a mass term for the electron and the positron by simple substitution into the action just presented. By comparing the given action with the conventional QED action we conclude that the spectral triple reproduces the action of QED of electrons and positrons.

7.7 Electroweak theory

The electroweak theory is based on the gauge group $\mathrm{U}(1) \oplus \mathrm{SU}(2)$. We present this model by itself as a stepping stone to the standard model. Most of the structure we present here will be unchanged when we treat the standard model in the next section.

Hilbert space and algebra representation

To build the electroweak theory in NCG we use the algebra $\mathcal{A}_F = \mathbb{C} \oplus \mathbb{H}$ represented on an 8-dimensional Hilbert space (representing one generation of leptons). Elements of $\mathbb{H}$ are written as pairs of complex numbers $(\alpha, \beta) \in \mathbb{C}^2$, while $\lambda \in \mathbb{C}$ represents complex numbers. The triple we will use is given by

$$\left(\mathcal{A}_F = \mathbb{C} \oplus \mathbb{H}, \mathcal{H}_F = \mathbb{C}^8, D_F, J_F, \gamma_F\right). \tag{7.7.1}$$

The representation of an element $(\lambda, (\alpha, \beta)) \in \mathcal{A}_F$ on $\mathcal{H}_F$ is given by

$$\mathbb{C} \oplus \mathbb{H} \ni (\lambda, (\alpha, \beta)) \mapsto \begin{pmatrix} \lambda & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \overline{\lambda} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha & \beta & 0 & 0 & 0 & 0 \\ 0 & 0 & -\overline{\beta} & \alpha & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{pmatrix} \begin{matrix} \nu_R \\ e_R \\ \nu_L \\ e_L \\ \overline{\nu}_R \\ \overline{e}_R \\ \overline{\nu}_L \\ \overline{e}_L \end{matrix}. \tag{7.7.2}$$

Chirality and charge conjuagtion operators

The KO-dimension is set equal to six. With our chosen basis we can now easily write down J_F and γ_F. This is simply a matter of writing down in mathematical terms the well known meaning of these operators. The chirality operator γ_F should give the value $+1$ for all left-handed particles (as well as all right-handed anti-particles). For all others particle it should give the value -1. The charge conjugation operator should map each particle to its corresponding antiparticle, while preserving chirality and with complex coefficients conjugated. Anti-particles should be mapped to their corresponding particle. The matrix form of the chirality operator is given by

$$\gamma_F = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}. \tag{7.7.3}$$

The charge conjugation operator J_F is written as

$$J_F = \begin{pmatrix} 0 & \mathbb{I}_4 \\ \mathbb{I}_4 & 0 \end{pmatrix} C. \tag{7.7.4}$$

The Hilbert space $\mathcal{H}_F$ can be decomposed as we did for QED into subspace of particles or anti-particles. We can write this as $\mathcal{H}_F = \mathcal{H}_F^l \oplus \mathcal{H}_F^{\bar{l}}$ with

$$\mathcal{H}_F^l = \begin{pmatrix} \nu_R \\ e_R \\ \nu_L \\ e_L \end{pmatrix} \quad \text{and} \quad \mathcal{H}_F^{\bar{l}} = \begin{pmatrix} \overline{\nu}_R \\ \overline{e}_R \\ \overline{\nu}_L \\ \overline{e}_L \end{pmatrix}. \tag{7.7.5}$$

The finite Dirac operator

To find the structure of the Dirac operator we start by writing it in its most general possible form as

$$D_F = \begin{pmatrix} S & T^* \\ T & S' \end{pmatrix}, \tag{7.7.6}$$

where S and S' must be self-adjoint. This is nothing more than writing down a general expression for what it means for D_F to be self-adjoint when decomposed on $\mathcal{H}_F = \mathcal{H}_F^l \oplus \mathcal{H}_F^{\bar{l}}$. From the condition $D_F J_F = J_F D_F$ we deduce that $S' = \overline{S}$ and $T = T^T$. The last equality implies that $T^* = \overline{T}$. In the next step we use the condition $\gamma_F D_F = -D_F \gamma_F$. First, we get that

$$S = \begin{pmatrix} 0 & Y^* \\ Y & 0 \end{pmatrix}, \tag{7.7.7}$$

where Y is an arbitrary 2×2 matrix. Second, we find that

$$T = \begin{pmatrix} T_R & 0 \\ 0 & T_L \end{pmatrix}, \tag{7.7.8}$$

where T_R and T_L are arbitrary symmetric 2×2 matrices. Instead of keeping T and Y in their most general possible form we now restrict their form based on what we desire for the final Lagrangian. This can of course be postponed until after deriving the Lagrangian. One would then restrict the extra parameters to be zero or unmeasurably small on the basis of input from experiment. Concretely, we restrict Y to be of the form

$$Y = \begin{pmatrix} Y_\nu & 0 \\ 0 & Y_e \end{pmatrix}. \tag{7.7.9}$$

We also set $T_L = 0$ while we restrict T_R to be of the form

$$T_R = \begin{pmatrix} Y_R & 0 \\ 0 & 0 \end{pmatrix}. \tag{7.7.10}$$

The final form for D_F then becomes

$$D_F = \left(\begin{array}{cc|cc|cc|cc} 0 & 0 & \overline{Y}_\nu & 0 & \overline{Y}_R & 0 & 0 & 0 \\ 0 & 0 & 0 & \overline{Y}_e & 0 & 0 & 0 & 0 \\ \hline Y_\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & Y_e & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline Y_R & 0 & 0 & 0 & 0 & 0 & Y_\nu & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & Y_e \\ \hline 0 & 0 & 0 & 0 & \overline{Y}_\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \overline{Y}_e & 0 & 0 \end{array}\right). \tag{7.7.11}$$

Gauge fields

We can write the inner fluctuation of the continuous Dirac operator D_M as $A_\mu = -ia\partial_\mu b$, with the requirement that $A_\mu = A_\mu^*$. Let us define two elements of $\mathcal{A} = \mathbb{C} \oplus \mathbb{H}$ as $a = (\lambda, q)$ and $b = (\lambda', q')$. Since each matrix element of the representation is linear with respect to algebra components, and the representation is block diagonal we can define

$$\Lambda_\mu = -i\lambda\partial_\mu\lambda' \tag{7.7.12}$$
$$\Lambda'_\mu = -i\bar{\lambda}\partial_\mu\bar{\lambda}' \tag{7.7.13}$$
$$Q_\mu = -iq\partial_\mu q'. \tag{7.7.14}$$

The matrix A_μ can be represented in terms of these fields. Since A_μ must be self-adjoint, this implies that its block diagonal components $\Lambda_\mu, \Lambda'_\mu$ and Q_μ must be self-adjoint. We impose self-adjointness of Λ_μ by requiring that

$$(\Lambda_\mu)^* = (\Lambda_\mu). \tag{7.7.15}$$

However, (Λ_μ) is just a complex number which means that it must be in the subset of real number. Hence we have $(\Lambda_\mu) \in \mathbb{R}$. Furthermore we see that

$$(\Lambda_\mu)^* = (-i\lambda\partial_\mu\lambda')^* = i\bar{\lambda}\partial_\mu\bar{\lambda}' = -\Lambda'_\mu. \tag{7.7.16}$$

We require Q_μ to be self-adjoint which means that

$$(Q_\mu)^* = (Q_\mu). \tag{7.7.17}$$

This last condition restrict Q_μ to be given by

$$Q_\mu = \begin{pmatrix} Q_{1\mu} & -iQ_{2\mu} + Q_{3\mu} \\ iQ_{2\mu} + Q_{3\mu} & -Q_{1\mu} \end{pmatrix}, \tag{7.7.18}$$

with $Q_{j\mu} \in \mathbb{R}$. Thus Q_μ is an element of the real algebra $i\mathfrak{su}(n)$.

The reasoning we have just presented leads to the expression

$$A_\mu = \begin{pmatrix} \Lambda_\mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\Lambda_\mu & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & Q_{1\mu} & -iQ_{2\mu} + Q_{3\mu} & 0 & 0 & 0 & 0 \\ 0 & 0 & iQ_{2\mu} + Q_{3\mu} & -Q_{1\mu} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Lambda_\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Lambda_\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Lambda_\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Lambda_\mu \end{pmatrix}. \tag{7.7.19}$$

To complete the fluctuated Dirac operator we need an expression for $J_F A_\mu J_F^*$. This sandwiching between J_F and J_F^* results in exchanging the upper left block for the lower left block and taking the complex conjugate at the same time. We get the matrix

$$J_F A_\mu J_F^* = \begin{pmatrix} \Lambda_\mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \Lambda_\mu & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \Lambda_\mu & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Lambda_\mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Lambda_\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\Lambda_\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & Q_{1\mu} & iQ_{2\mu} + Q_{3\mu} \\ 0 & 0 & 0 & 0 & 0 & 0 & -iQ_{2\mu} + Q_{3\mu} & -Q_{1\mu} \end{pmatrix}. \tag{7.7.20}$$

Combining the above results by matrix subtraction we find an expression for B_μ given as

$$B_\mu = A_\mu - J_F A_\mu J_F^* \tag{7.7.21}$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2\Lambda_\mu & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\Lambda_\mu + Q_{1\mu} & -iQ_{2\mu} + Q_{3\mu} & 0 & 0 & 0 & 0 \\ 0 & 0 & iQ_{2\mu} + Q_{3\mu} & -\Lambda_\mu - Q_{1\mu} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\Lambda_\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Lambda_\mu - Q_{1\mu} & -iQ_{2\mu} - Q_{3\mu} \\ 0 & 0 & 0 & 0 & 0 & 0 & iQ_{2\mu} - Q_{3\mu} & \Lambda_\mu + Q_{1\mu} \end{pmatrix}. \tag{7.7.22}$$

To simplify the later presentation we use the previously defined matrix Q_μ and write

$$B_\mu = \left(\begin{array}{cc|c|cc|c} 0 & 0 & & & & \\ 0 & -2\Lambda_\mu & & & & \\ \hline & & -\Lambda_\mu \mathbb{I} + Q_\mu & & & \\ \hline & & & 0 & 0 & \\ & & & 0 & 2\Lambda_\mu & \\ \hline & & & & & \Lambda_\mu \mathbb{I} - \overline{Q}_\mu \end{array}\right). \tag{7.7.23}$$

To calculate the inner fluctuations of the finite Dirac operator D_F we use $\phi = \sum_j a_j\,[D_F, b_j]$. We define a_j and b_j by

$$a_j = \left(\begin{array}{cccc|cccc} \lambda_j & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \overline{\lambda}_j & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_j & \beta_j & 0 & 0 & 0 & 0 \\ 0 & 0 & -\overline{\beta}_j & \alpha_j & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \lambda_j & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_j & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_j & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_j \end{array}\right) \quad b_j = \left(\begin{array}{cccc|cccc} \lambda'_j & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \overline{\lambda}'_j & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha'_j & \beta'_j & 0 & 0 & 0 & 0 \\ 0 & 0 & -\overline{\beta}'_j & \alpha'_j & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \lambda'_j & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda'_j & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda'_j & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda'_j \end{array}\right). \tag{7.7.24}$$

Note that the b_j matrices have components marked with a prime. From before we have

$$D_F = \begin{pmatrix} S & T^* \\ T & S' \end{pmatrix} = \left(\begin{array}{cccc|cccc} 0 & 0 & \overline{Y}_\nu & 0 & \overline{Y}_R & 0 & 0 & 0 \\ 0 & 0 & 0 & \overline{Y}_e & 0 & 0 & 0 & 0 \\ Y_\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & Y_e & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline Y_R & 0 & 0 & 0 & 0 & 0 & Y_\nu & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & Y_e \\ 0 & 0 & 0 & 0 & \overline{Y}_\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \overline{Y}_e & 0 & 0 \end{array}\right). \tag{7.7.25}$$

Only the S sub-matrix of D_F contributes to the commutator. We also see that we sum over matching indices, which means it is just the sum of the separate matrix expressions for each j. Using these input we see that the whole calculation can be summarized by defining the sub-matrix $\omega = \sum_j a'_j \left[S, b'_j\right]$. Here a'_j and b'_j denotes the upper left (4×4)-submatrices of a_j and b_j. We calculate that

$$\omega = \begin{pmatrix} 0 & 0 & \overline{Y}_\nu \phi'_1 & \overline{Y}_\nu \phi'_2 \\ 0 & 0 & -\overline{Y}_e \overline{\phi}'_2 & \overline{Y}_e \overline{\phi}'_1 \\ Y_\nu \phi_1 & -Y_e \overline{\phi}_2 & 0 & 0 \\ Y_\nu \phi_2 & Y_e \overline{\phi}_1 & 0 & 0 \end{pmatrix}, \tag{7.7.26}$$

where we have introduced four new scalar fields by defining

$$\phi_1 = \sum_j \alpha_j(\lambda'_j - \alpha'_j) + \beta_j \overline{\beta}'_j \tag{7.7.27}$$

$$\phi'_1 = \sum_j \lambda_j(\alpha'_j - \lambda'_j) \tag{7.7.28}$$

$$\phi_2 = \sum_j \overline{\alpha}_j \overline{\beta}'_j - \overline{\beta}_j(\lambda'_j - \alpha'_j) \tag{7.7.29}$$

$$\phi'_2 = \sum_j \lambda_j \beta'_j. \tag{7.7.30}$$

The complete ϕ-matrix can be expressed as

$$\phi = \sum_j a_j \left[D_F, b_j\right] = \begin{pmatrix} \omega & 0 \\ 0 & 0 \end{pmatrix}. \tag{7.7.31}$$

We require ϕ to be self-adjoint which implies that ω must be self-adjoint. This means that $\phi_2 = \overline{\phi}'_2$ and $\phi_1 = \overline{\phi}'_1$. There are now only two complex degrees of freedom in ω. We modify ω so that it is explicitly self-adjoint, and we also calculate $\overline{\omega}$ which we will need shortly. We find that

$$\omega = \begin{pmatrix} 0 & 0 & \overline{Y}_\nu \overline{\phi}_1 & \overline{Y}_\nu \overline{\phi}_2 \\ 0 & 0 & -\overline{Y}_e \phi_2 & \overline{Y}_e \phi_1 \\ Y_\nu \phi_1 & -Y_e \overline{\phi}_2 & 0 & 0 \\ Y_\nu \phi_2 & Y_e \overline{\phi}_1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \overline{\omega} = \begin{pmatrix} 0 & 0 & Y_\nu \phi_1 & Y_\nu \phi_2 \\ 0 & 0 & -Y_e \overline{\phi}_2 & Y_e \overline{\phi}_1 \\ \overline{Y}_\nu \overline{\phi}_1 & -\overline{Y}_e \phi_2 & 0 & 0 \\ \overline{Y}_\nu \overline{\phi}_2 & \overline{Y}_e \phi_1 & 0 & 0 \end{pmatrix}. \tag{7.7.32}$$

We can now find Φ from the general expression $\Phi = D_F + \phi + J_F \phi J_F^*$. A short calculation shows that

$$\Phi = D_F + \phi + J_F \phi J_F^* \tag{7.7.33}$$

$$= \begin{pmatrix} S & T^* \\ T & \overline{S} \end{pmatrix} + \begin{pmatrix} \omega & 0 \\ 0 & 0 \end{pmatrix} + J_F \begin{pmatrix} \omega & 0 \\ 0 & 0 \end{pmatrix} J_F^* \tag{7.7.34}$$

$$= \begin{pmatrix} S & T^* \\ T & \overline{S} \end{pmatrix} + \begin{pmatrix} \omega & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \overline{\omega} \end{pmatrix} \tag{7.7.35}$$

$$= \begin{pmatrix} S + \omega & T^* \\ T & \overline{S + \omega} \end{pmatrix}. \tag{7.7.36}$$

The explicit form of this matrix is given by

$$\Phi = \left(\begin{array}{cc|cc} 0 & (Y+Y_0)^* & \multicolumn{2}{c}{\multirow{2}{*}{T^*}} \\ Y+Y_0 & 0 & & \\ \hline \multicolumn{2}{c|}{\multirow{2}{*}{T}} & 0 & (Y+Y_0)^T \\ & & \overline{Y+Y_0} & 0 \end{array}\right), \tag{7.7.37}$$

where we have used

$$Y = \begin{pmatrix} Y_\nu & 0 \\ 0 & Y_e \end{pmatrix} \quad Y_0 = \begin{pmatrix} Y_\nu \phi_1 & -Y_e \overline{\phi}_2 \\ Y_\nu \phi_2 & Y_e \overline{\phi}_1 \end{pmatrix} \quad T = \left(\begin{array}{cc|cc} Y_R & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right). \tag{7.7.38}$$

Field strengths and traces

We have now defined and computed the necessary gauge fields. The next step is to define the field strengths. The general definition of the field strength of B_μ is given as

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + i\left[B_\mu, B_\nu\right]. \tag{7.7.39}$$

To clarify the meaning of equation 7.7.39 we first observe that the term B_μ represents four fields taking values in the domain of 8×8 matrices. Equation 7.7.39 therefore defines $4 \times 4 = 16$ matrices which each have $8 \times 8 = 64$ components. We could write this more explicitly as

$$F^{ij}_{\mu\nu} = \partial_\mu B^{ij}_\nu - \partial_\nu B^{ij}_\mu + i\left[B^{ij}_\mu, B^{ij}_\nu\right], \tag{7.7.40}$$

where $i, j \in \{1, \ldots, 8\}$. We wish to express the field strength of B_μ in terms of the field strengths of Λ_μ and Q_μ, and hence we also define

$$\Lambda_{\mu\nu} = \partial_\mu \Lambda_\nu - \partial_\nu \Lambda_\mu \tag{7.7.41}$$

$$Q_{\mu\nu} = \partial_\mu Q_\nu - \partial_\nu Q_\mu + i\left[Q_\mu, Q_\nu\right]. \tag{7.7.42}$$

The symbol $\Lambda_{\mu\nu}$ is a set of 16 numbers while $Q_{\mu\nu}$ is a set of 16 matrices. We do not write the commutator term for $\Lambda_{\mu\nu}$ since it is always zero. The explicit form of $F_{\mu\nu}$ is found by a straight forward calculation to be

$$F_{\mu\nu} = \left(\begin{array}{cc|c|cc|c} 0 & 0 & & & & \\ 0 & -2\Lambda_{\mu\nu} & & & & \\ \hline & & -\Lambda_{\mu\nu}\mathbb{I}_2 + Q_{\mu\nu} & & & \\ \hline & & & 0 & 0 & \\ & & & 0 & 2\Lambda_{\mu\nu} & \\ \hline & & & & & \Lambda_{\mu\nu}\mathbb{I}_2 - \overline{Q}_{\mu\nu} \end{array}\right). \tag{7.7.43}$$

The next expression we need to evaluate is $F_{\mu\nu}F^{\mu\nu}$. Recall that each $F_{\mu\nu}$ is an 8×8 matrix such that this is an expression for a new 8×8 matrix resulting from the sum of 16 matrix products. Each

matrix $F_{\mu\nu}$ is block diagonal so the result can be read off directly as

$$F_{\mu\nu}F^{\mu\nu} = \tag{7.7.44}$$

$$\left(\begin{array}{cc|c|cc|c} 0 & 0 & & & & \\ 0 & 4\Lambda_{\mu\nu}\Lambda^{\mu\nu} & & & & \\ \hline & & \Lambda_{\mu\nu}\Lambda^{\mu\nu}\mathbb{I}_2 + Q_{\mu\nu}Q^{\mu\nu} & & & \\ & & -\Lambda_{\mu\nu}Q^{\mu\nu} - Q_{\mu\nu}\Lambda^{\mu\nu} & & & \\ \hline & & & 0 & 0 & \\ & & & 0 & 4\Lambda_{\mu\nu}\Lambda^{\mu\nu} & \\ \hline & & & & & \Lambda_{\mu\nu}\Lambda^{\mu\nu}\mathbb{I}_2 + \overline{Q}_{\mu\nu}\overline{Q}^{\mu\nu} \\ & & & & & -\Lambda_{\mu\nu}\overline{Q}^{\mu\nu} - \overline{Q}_{\mu\nu}\Lambda^{\mu\nu} \end{array}\right).$$

To calculate the trace of this matrix, we note that the trace of $Q^{\mu\nu}$ is zero and therefore all cross-terms have trace zero. Furthermore we find that $\mathrm{tr}(Q_{\mu\nu}Q^{\mu\nu}) = \mathrm{tr}\big(\overline{Q}_{\mu\nu}\overline{Q}^{\mu\nu}\big)$. The matrix $\mathbb{I}_2$ contributes a factor of 2. By using these results we can read the trace directly from the matrix and we conclude that

$$Tr\left(F_{\mu\nu}F^{\mu\nu}\right) = 12\Lambda_{\mu\nu}\Lambda^{\mu\nu} + 2Tr\left(Q_{\mu\nu}Q^{\mu\nu}\right). \tag{7.7.45}$$

We have to calculate the traces of Φ^2 ,Φ^4 and $D_\mu\Phi\ D^\mu\Phi$. We do not give the full derivation. We note that $D_\mu = \partial_\mu + [B_\mu, \Phi]$. Repeating Φ and calculating Φ^2 we get

$$\Phi = \begin{pmatrix} S+\omega & T^* \\ T & \overline{S+\omega} \end{pmatrix} \quad \Phi^2 = \begin{pmatrix} (S+\omega)(S+\omega) + TT^* & (S+\omega)T^* + T^*\overline{S+\omega} \\ T(S+\omega) + \overline{S+\omega}T & TT^* + \overline{(S+\omega)}\ \overline{(S+\omega)} \end{pmatrix}. \tag{7.7.46}$$

The calculations are straight forwards but cumbersome to present. The resulting traces involving Φ are

$$\mathrm{Tr}\left(\Phi^2\right) = 4(|Y_\nu|^2 + |Y_e|^2)|H|^2 + 2|Y_R|^2 \tag{7.7.47}$$

$$= 4a|H|^2 + 2c \tag{7.7.48}$$

$$\mathrm{Tr}\left(\Phi^4\right) = 4(|Y_\nu|^4 + |Y_e|^4)|H|^4 + 8(|Y_R|^2|Y_\nu|^2)|H|^2 + 2|Y_R|^4 \tag{7.7.49}$$

$$= 4b|H|^4 + 8e|H|^2 + 2d \tag{7.7.50}$$

$$\mathrm{Tr}\left(D_\mu\Phi\ D^\mu\Phi\right) = 4a|\tilde{D}_\mu H|^2. \tag{7.7.51}$$

Here we have combined the two complex scalar fields ϕ_1 and ϕ_2 into the complex Higgs column vector

$$H = \begin{pmatrix} \phi_1 + 1 \\ \phi_2 \end{pmatrix}, \tag{7.7.52}$$

and we have also defined the following constants

$$a = |Y_\nu|^2 + |Y_e|^2 \tag{7.7.53}$$

$$b = |Y_\nu|^4 + |Y_e|^4 \tag{7.7.54}$$

$$c = |Y_R|^2 \tag{7.7.55}$$

$$d = |Y_R|^4 \tag{7.7.56}$$

$$e = |Y_R|^2|Y_\nu|^2. \tag{7.7.57}$$

We have also defined $\tilde{D}_\mu$ as

$$\tilde{D}_\mu = \partial_\mu - iQ^a_\mu\sigma^a + i\Lambda_\mu. \tag{7.7.58}$$

Bosonic Lagrangian

All the components necessary for the bosonic Lagrangian have been calculated. We can write the full expression as

$$S_b = \int \sqrt{g}\, d^4x\, L_{Total} \tag{7.7.59}$$

$$L_{Total} = 8L_{Gravity} + L_{Gauge} + L_{Higgs} \tag{7.7.60}$$

$$L_{Gravity} = \frac{f_4\Lambda^4}{2\pi^2} - \frac{f_2\Lambda^2}{24\pi^2}R + \frac{f(0)}{480\pi^2}\left(R^2 - 3R_{\mu\nu}R^{\mu\nu}\right) \tag{7.7.61}$$

$$L_{Gauge} = \frac{f(0)}{24\pi^2}\left(12\Lambda_{\mu\nu}\Lambda^{\mu\nu} + 2Tr\left(Q_{\mu\nu}Q^{\mu\nu}\right)\right) \tag{7.7.62}$$

$$L_{Higgs} = -\frac{2f_2\Lambda^2}{4\pi^2}(4a|H|^2 + 2c) + \frac{f(0)}{8\pi^2}(4b|H|^4 + 8e|H|^2 + 2d) \tag{7.7.63}$$

$$+ \frac{f(0)}{24\pi^2}\Delta(4a|H|^2 + 2c) + \frac{f(0)}{48\pi^2}R(4a|H|^2 + 2c) + \frac{f(0)}{8\pi^2}4a|\tilde{D}_\mu H|^2. \tag{7.7.64}$$

This completes the mathematical derivation of the bosonic Lagrangian. To connect this with the conventional electroweak Lagrangian found in particle physics, we need to perform some rewrites. We first step is to rename the gauge fields while at the same time "factoring out" coupling constants. This is accomplished by setting

$$\Lambda_\mu = \frac{1}{2}g_1 B_\mu \tag{7.7.65}$$

$$Q^a_\mu = \frac{1}{2}g_2 W^a_\mu. \tag{7.7.66}$$

The field W^a_μ is the usual $\mathrm{SU}(2)$ gauge fields and B_μ is the usual $\mathrm{U}(1)$ field. We then apply some conventional definitions given by

$$c_w = \cos\theta_w = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} \tag{7.7.67}$$

$$s_w = \sin\theta_w = \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \tag{7.7.68}$$

$$W_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu + iW^2_\mu\right) \tag{7.7.69}$$

$$W^*_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu - iW^2_\mu\right) \tag{7.7.70}$$

$$Z_\mu = c_w W^3_\mu - s_w B_\mu \tag{7.7.71}$$

$$A_\mu = s_w W^3_\mu + c_w B_\mu. \tag{7.7.72}$$

We also need to transform the Higgs field to a form that allows a perturbative interpretation, to see how the Higgs field gives masses (and mixings) to bosons and fermions. We rescale and parametrize the Higgs fields by setting

$$H = \begin{pmatrix}\phi_1 + 1\\ \phi_2\end{pmatrix} \to H = \sqrt{\frac{af(0)}{\pi^2}}\begin{pmatrix}\phi_1 + 1\\ \phi_2\end{pmatrix} = \begin{pmatrix}v + h + i\phi^0\\ i\sqrt{2}\phi^-\end{pmatrix}. \tag{7.7.73}$$

For the complex field ϕ^- we also define the notation $\phi^+ = \overline{(\phi^-)}$. By gauge transforming the Higgs field we can restrict the form to be

$$H(x) = \begin{pmatrix}v + h(x)\\ 0\end{pmatrix}. \tag{7.7.74}$$

Rewriting the above Lagrangian with these equations produces the bosonic part of leptonic electroweak theory. We do not present the full derivation of this.

Fermionic Lagrangian

The fermionic part of the Lagrangian follows from the fermionic action principle

$$S_f = \frac{1}{2}\langle J\xi, D_A\xi\rangle, \tag{7.7.75}$$

where $\xi \in \mathcal{H}^+$. The generic form of the Dirac operator on an almost commutative manifold is given by

$$D = D_M \otimes \mathbb{I} + \gamma^\mu \otimes B_\mu + \gamma_M \otimes \Phi \tag{7.7.76}$$

We have derived the form of B_μ and Φ for the electroweak theory in the previous subsections. A generic element of H^+ can be written as

$$\begin{aligned}\xi =&\psi_{e_L^\lambda} \otimes e_L^\lambda + \psi_{e_R^\lambda} \otimes e_R^\lambda + \psi_{\overline{e}_R^\lambda} \otimes \overline{e}_L^\lambda + \psi_{\overline{e}_L^\lambda} \otimes \overline{e}_R^\lambda \\ &+\psi_{\nu_L^\lambda} \otimes \nu_L^\lambda + \psi_{\nu_R^\lambda} \otimes \nu_R^\lambda + \psi_{\overline{\nu}_R^\lambda} \otimes \overline{\nu}_L^\lambda + \psi_{\overline{\nu}_L^\lambda} \otimes \overline{\nu}_R^\lambda\end{aligned} \tag{7.7.77}$$

Note that the ψ's are classical spinors that are treated as Grassmann variables. Furthermore the subscripts and the overbars on the ψ's are just names. The L and R designate that the ψ belongs to $\mathcal{H}_M^+$ or $\mathcal{H}_M^-$ independent of the overbar.

To determine the Lagrangian we need to calculate the set of terms

$$\frac{1}{2}\left\langle J\xi\,, D\xi\right\rangle = \frac{1}{2}\left\langle J\xi\,, D_M \otimes \mathbb{I}\xi\right\rangle + \frac{1}{2}\left\langle J\xi\,, \gamma^\mu \otimes B_\mu\xi\right\rangle + \frac{1}{2}\left\langle J\xi\,, \gamma_M \otimes \Phi\xi\right\rangle. \tag{7.7.78}$$

The first term, involving $D_M \otimes \mathbb{I}$, produces all the kinetic terms for the fermions. The term, involving $\gamma^\mu \otimes B_\mu$, produces all the gauge-fermion terms. The third term, involving $\gamma_M \otimes \Phi$, produces the Yukawa couplings between Higgs and fermions. The Yukawa couplings are determined by Y_e and Y_ν, as well as the fermion Majorana mass terms determined by Y_R.

The full derivation of the fermionic part will be presented in the next section when constructing the standard model (see subsection 7.8).

7.8 The standard model

The standard model is the accepted current theory for particle physics. There are no known deviations from this model in collider experiments. The SM is based on the gauge group $U(1) \oplus SU(2) \oplus SU(3)$. In the present context we take an extended view of the standard model that includes right-handed neutrinos as well as neutrino masses and mixings. Most of the structure concerning the electroweak interactions of leptons, as presented in the previous section, will be carried over to the standard model unchanged. For ease of reading we have decided to include some parts of the reasoning of the previous section in this section as well.

Hilbert space and algebra representation

The starting point for deriving the standard model is the algebra $\mathcal{A} = \mathbb{C} \oplus \mathbb{H} \oplus M_3(\mathbb{C})$. We write the elements of this algebra as $(\lambda, (\alpha, \beta), m)$, with $m \in M_3(\mathbb{C})$ and otherwise the notation is the same as for the electroweak theory. Sometimes we will also write $(m_1, \ldots, m_9)$ instead of just m. To get started we first define the Hilbert space that we will represent this algebra on. The Hilbert space will be (3×32) -dimensional, and presentation of the full fledged matrices will be typographically inconvenient. We will instead display matrices for lower dimensional subspaces of the full Hilbert spaces. We will switch between one generation (1G) and three generation (3G) according to necessity and presentational convenience. We will on occasion use the same symbol

for an operator for both the 1G and the 3G form. The Hilbert space for the leptonic particles (1G) will be the same as the one used for the presentation of the electroweak theory. The two four-dimensional spaces used are given by

$$\mathcal{H}_F^l = \begin{pmatrix} \nu_R \\ e_R \\ \nu_L \\ e_L \end{pmatrix} \quad \mathcal{H}_F^{\bar{l}} = \begin{pmatrix} \overline{\nu}_R \\ \overline{e}_R \\ \overline{\nu}_L \\ \overline{e}_L \end{pmatrix}. \tag{7.8.1}$$

For the quark states (1G) we will use the two 12-dimensional spaces given by

$$\mathcal{H}_F^q = \begin{pmatrix} u_R \\ d_R \\ u_L \\ d_L \end{pmatrix} \otimes \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad \mathcal{H}_F^{\overline{q}} = \begin{pmatrix} \overline{u}_R \\ \overline{d}_R \\ \overline{u}_L \\ \overline{d}_L \end{pmatrix} \otimes \begin{pmatrix} R \\ G \\ B \end{pmatrix}. \tag{7.8.2}$$

We thus have a total Hilbert space (1G) given by

$$\mathcal{H}_F = \mathcal{H}_F^l \oplus \mathcal{H}_F^{\bar{l}} \oplus \mathcal{H}_F^q \oplus \mathcal{H}_F^{\overline{q}}. \tag{7.8.3}$$

This is a 32-dimensional space ($4 + 4 + 12 + 12 = 32$). For 3G the full Hilbert space is simply

$$\mathcal{H}_F = \left(\mathcal{H}_F^l \oplus \mathcal{H}_F^{\bar{l}} \oplus \mathcal{H}_F^q \oplus \mathcal{H}_F^{\overline{q}}\right)^{\oplus 3}. \tag{7.8.4}$$

We next present the representations of the algebra on the Hilbert space $\mathcal{H}_F$. The same technique of presenting the representation on the various subspace will be used.

The representation on $\mathcal{H}_F^l$ (1G) is given by

$$\mathbb{C} \oplus \mathbb{H} \oplus M_3(\mathbb{C}) \ni (\lambda, (\alpha, \beta), m) \mapsto \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \overline{\lambda} & 0 & 0 \\ 0 & 0 & \alpha & \beta \\ 0 & 0 & -\overline{\beta} & \alpha \end{pmatrix} \begin{vmatrix} \nu_R \\ e_R \\ \nu_L \\ e_L \end{vmatrix} \tag{7.8.5}$$

The representation on $\mathcal{H}_F^{\bar{l}}$ (1G) is given by

$$\mathbb{C} \oplus \mathbb{H} \oplus M_3(\mathbb{C}) \ni (\lambda, (\alpha, \beta), m) \mapsto \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix} \begin{vmatrix} \overline{\nu}_R \\ \overline{e}_R \\ \overline{\nu}_L \\ \overline{e}_L \end{vmatrix}. \tag{7.8.6}$$

The representation on $\mathcal{H}_F^q$ (1G) is given by

$$\mathbb{C} \oplus \mathbb{H} \oplus M_3(\mathbb{C}) \ni (\lambda, (\alpha, \beta), m) \mapsto \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \overline{\lambda} & 0 & 0 \\ 0 & 0 & \alpha & \beta \\ 0 & 0 & -\overline{\beta} & \alpha \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{vmatrix} u_R \\ d_R \\ u_L \\ d_L \end{vmatrix} \otimes \begin{vmatrix} R \\ G \\ B \end{vmatrix} \tag{7.8.7}$$

The representation on $\mathcal{H}_F^{\overline{q}}$ (1G) is given by

$$\mathbb{C} \oplus \mathbb{H} \oplus M_3(\mathbb{C}) \ni (\lambda, (\alpha, \beta), m) \mapsto \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & m_9 \end{pmatrix} \begin{vmatrix} \overline{u}_R \\ \overline{d}_R \\ \overline{u}_L \\ \overline{d}_L \end{vmatrix} \otimes \begin{vmatrix} R \\ G \\ B \end{vmatrix}. \tag{7.8.8}$$

Chirality and charge conjuagtion operators

The γ_F on the subspace $H_l \oplus H_{\bar{l}}$ will be the same as the one given for the electroweak theory. The same applies for the J_F operator. On the subspace $H_q \oplus H_{\bar{q}}$ we get (1G)

$$\gamma_F|_{H_F^q \oplus H_F^{\bar{q}}} = \begin{pmatrix} \mathbb{I}_{+2}^{-2} & 0 \\ 0 & \mathbb{I}_{-2}^{+2} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{7.8.9}$$

and

$$J_F|_{H_F^q \oplus H_F^{\bar{q}}} = \begin{pmatrix} 0 & \mathbb{I}_4 \\ \mathbb{I}_4 & 0 \end{pmatrix} C \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{7.8.10}$$

The finite Dirac operator

As in the electroweak theory, the finite Dirac operator is given by

$$D_F = \begin{pmatrix} S & T^* \\ T & \overline{S} \end{pmatrix}. \tag{7.8.11}$$

We list the various components on subspaces of the Hilbert space. The S sub-matrix is given as (3G)

$$S|_{(\mathcal{H}_F^l \oplus \mathcal{H}_F^{\bar{l}})^{\oplus 3}} = \left(\begin{array}{cc|cc} 0 & 0 & Y_\nu^* & 0 \\ 0 & 0 & 0 & Y_e^* \\ \hline Y_\nu & 0 & 0 & 0 \\ 0 & Y_e & 0 & 0 \end{array}\right) \quad S|_{(\mathcal{H}_F^q \oplus \mathcal{H}_F^{\bar{q}})^{\oplus 3}} = \left(\begin{array}{cc|cc} 0 & 0 & Y_u^* & 0 \\ 0 & 0 & 0 & Y_d^* \\ \hline Y_u & 0 & 0 & 0 \\ 0 & Y_d & 0 & 0 \end{array}\right) \otimes \mathbb{I}_3. \tag{7.8.12}$$

Where the Ys are 3×3 Yukawa coupling matrices. The given S submatrices operate on a 12- and a 36-dimensional subspace of the Hilbert space.

The Majorana masses for right-handed neutrinos are specified by

$$T|_{(\mathcal{H}_F^l \oplus \mathcal{H}_F^{\bar{l}})^{\oplus 3}} = \left(\begin{array}{cc|cc} \overline{Y}_R & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right). \tag{7.8.13}$$

Where the Y_R is a 3×3 Majorana mass matrix. The given T submatrix operate on 12-dimensional subspace of the Hilbert space. There are no Majorana masses for other particles which means that

$$T|_{(\mathcal{H}_F^q \oplus \mathcal{H}_F^{\bar{q}})^{\oplus 3}} = \left(\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \otimes \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{7.8.14}$$

Gauge fields

We can write the inner fluctuation of the continuous Dirac operator D_M as $A_\mu = -ia\partial_\mu b$, with the requirement that $A_\mu = A_\mu^*$. Let us define two elements of $\mathcal{A} = \mathbb{C} \oplus \mathbb{H} \oplus M_3(\mathbb{C})$ as $a = (\lambda, q, m)$

and $b = (\lambda', q', m')$. Since each matrix element of the representation is linear with respect to algebra components, and the representation is block diagonal we can define

$$\Lambda_\mu = -i\lambda\partial_\mu\lambda' \tag{7.8.15}$$
$$\Lambda'_\mu = -i\bar{\lambda}\partial_\mu\bar{\lambda}' \tag{7.8.16}$$
$$Q_\mu = -iq\partial_\mu q'. \tag{7.8.17}$$
$$V'_\mu = -im\partial_\mu m'. \tag{7.8.18}$$

(Note the prime on V'_μ which is to facilitate notation when we implement the unimodularity condition in the step after this one.) The matrix A_μ can be represented in terms of these fields. We require that A_μ is self-adjoint which means that its block diagonal components Λ_μ, Λ'_μ, Q_μ and V_μ must be self-adjoint. We impose self-adjointness of Λ_μ by requiring that

$$(\Lambda_\mu)^* = (\Lambda_\mu). \tag{7.8.19}$$

However, (Λ_μ) is just a complex number which means that it must be in the subset of real number. Hence we have $(\Lambda_\mu) \in \mathbb{R}$. Furthermore we see that

$$(\Lambda_\mu)^* = (-i\lambda\partial_\mu\lambda')^* = i\bar{\lambda}\partial_\mu\bar{\lambda}' = -\Lambda'_\mu. \tag{7.8.20}$$

We impose self-adjointness of Q_μ that

$$(Q_\mu)^* = (Q_\mu). \tag{7.8.21}$$

This last condition restrict Q_μ to be

$$Q_\mu = \begin{pmatrix} Q_{1\mu} & -iQ_{2\mu} + Q_{3\mu} \\ iQ_{2\mu} + Q_{3\mu} & -Q_{1\mu} \end{pmatrix}, \tag{7.8.22}$$

with $Q_{j\mu} \in \mathbb{R}$. Thus Q_μ is an element of the real algebra of $i\mathfrak{su}(n)$. We impose self-adjointness of V'_μ by requiring that

$$(V'_\mu)^* = (V'_\mu). \tag{7.8.23}$$

We know V'_μ is some subset of $M_3(\mathbb{C})$, and for $M_3(\mathbb{C})$ we know that the set of all skew-adjoint elements are given by $\mathfrak{u}(3)$. Hence V'_μ must be a member of $i\mathfrak{u}(3)$. At this point we impose the unimodularity condition, which amounts to

$$\mathrm{tr}(A_\mu) = 0. \tag{7.8.24}$$

To actually calculate this we need to set up the matrices resulting from the previous definitions. We find that

$$A_\mu|_{\mathcal{H}_F^l} = \begin{pmatrix} \Lambda_\mu & 0 & & \\ 0 & -\Lambda_\mu & & \\ & & Q_\mu & \end{pmatrix} \qquad A_\mu|_{\mathcal{H}_F^{\bar{l}}} = \begin{pmatrix} \Lambda_\mu & 0 & & \\ 0 & \Lambda_\mu & & \\ & & \Lambda_\mu & \\ & & & \Lambda_\mu \end{pmatrix} \tag{7.8.25}$$

$$A_\mu|_{\mathcal{H}_F^q} = \begin{pmatrix} \Lambda_\mu & 0 & \\ 0 & -\Lambda_\mu & \\ & & Q_\mu \end{pmatrix} \otimes \mathbb{I}_3 \quad A_\mu|_{\mathcal{H}_F^{\bar{q}}} = \mathbb{I}_4 \otimes \left(V'_\mu\right). \tag{7.8.26}$$

The trace of A_μ is the sum of the traces of the listed submatrices. We note that Q_μ is traceless and that Λ_μ is just a number. This means that the lepton and quark submatrices are both traceless. The

non-zero part of the trace results just from the anti-lepton and the anti-quark submatrices. The trace can be evaluated by inspection to be

$$\mathrm{tr}(A_\mu) = 4\Lambda_\mu - 4tr(V'_\mu). \tag{7.8.27}$$

Setting this trace to zero we get $tr(V'_\mu) = -\Lambda_\mu$. If we define a new 3×3 matrix by setting $V_\mu = V'_\mu + \frac{1}{3}\Lambda_\mu \mathbb{I}_3$ or $\bar{V}_\mu = -V'_\mu - \frac{1}{3}\Lambda_\mu \mathbb{I}_3$ get a traceless matrix. We decide to use $V'_\mu = -(\bar{V}_\mu + \frac{1}{3}\Lambda_\mu \mathbb{I}_3)$. We can now write the gauge fields in their final forms as

$$A_\mu|_{\mathcal{H}_F^l} = \begin{pmatrix} \Lambda_\mu & 0 & \\ 0 & -\Lambda_\mu & \\ & & Q_\mu \end{pmatrix} \qquad A_\mu|_{\mathcal{H}_F^{\bar{l}}} = \begin{pmatrix} \Lambda_\mu & 0 & & \\ 0 & \Lambda_\mu & & \\ & & \Lambda_\mu & \\ & & & \Lambda_\mu \end{pmatrix} \tag{7.8.28}$$

$$A_\mu|_{\mathcal{H}_F^q} = \begin{pmatrix} \Lambda_\mu & 0 & \\ 0 & -\Lambda_\mu & \\ & & Q_\mu \end{pmatrix} \otimes \mathbb{I}_3. \qquad A_\mu|_{\mathcal{H}_F^{\bar{q}}} = -\mathbb{I}_4 \otimes \left(\overline{V}_\mu + \frac{1}{3}\Lambda_\mu \mathbb{I}_3)\right) \tag{7.8.29}$$

To complete the fluctuated Dirac operator we need an expression for $J_F A_\mu J_F^*$. This sandwiching between J_F and J_F^* results in exchanging the upper left block for the lower left block and taking the complex conjugate at the same time. On the lepton-anti-lepton subspace we get the matrix

$$J_F A_\mu J_F^*|_{\mathcal{H}_F^l \oplus \mathcal{H}_F^{\bar{l}}} = \tag{7.8.30}$$

$$\begin{pmatrix} \Lambda_\mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \Lambda_\mu & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \Lambda_\mu & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Lambda_\mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Lambda_\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\Lambda_\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & Q_{1\mu} & iQ_{2\mu} + Q_{3\mu} \\ 0 & 0 & 0 & 0 & 0 & 0 & -iQ_{2\mu} + Q_{3\mu} & -Q_{1\mu} \end{pmatrix}.$$

Combining the above results by matrix subtraction we find an expression for B_μ the lepton-anti-lepton subspace given as

$$B_\mu|_{\mathcal{H}_F^l \oplus \mathcal{H}_F^{\bar{l}}} = \tag{7.8.31}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2\Lambda_\mu & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\Lambda_\mu + Q_{1\mu} & -iQ_{2\mu} + Q_{3\mu} & 0 & 0 & 0 & 0 \\ 0 & 0 & iQ_{2\mu} + Q_{3\mu} & -\Lambda_\mu - Q_{1\mu} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\Lambda_\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Lambda_\mu - Q_{1\mu} & -iQ_{2\mu} - Q_{3\mu} \\ 0 & 0 & 0 & 0 & 0 & 0 & iQ_{2\mu} - Q_{3\mu} & \Lambda_\mu + Q_{1\mu} \end{pmatrix}.$$

To simplify the later presentation we use the previously defined matrix Q_μ and write

$$B_\mu|_{\mathcal{H}_F^l \oplus \mathcal{H}_F^{\bar{l}}} = \left(\begin{array}{cc|c|cc|c} 0 & 0 & & & & \\ 0 & -2\Lambda_\mu & & & & \\ \hline & & -\Lambda_\mu \mathbb{I}_2 + Q_\mu & & & \\ \hline & & & 0 & 0 & \\ & & & 0 & 2\Lambda_\mu & \\ \hline & & & & & \Lambda_\mu \mathbb{I}_2 - \overline{Q}_\mu \end{array}\right). \tag{7.8.32}$$

The same logic applies to the quark-antiquark subspace. The challenge with quark-antiquark subspace is how to notate the color aspect. Because of the linearity of the tensor product we can exchange numerical factors between the left-side and the right-side of the $\otimes$ symbol. We write

$$A_\mu|_{\mathcal{H}_F^q \oplus \mathcal{H}_F^{\overline{q}}} = \tag{7.8.33}$$

$$\left(\begin{array}{cc|c|cc|c} \Lambda_\mu \mathbb{I}_3 & 0 & & & & \\ 0 & -\Lambda_\mu \mathbb{I}_3 & & & & \\ \hline & & Q_\mu \otimes \mathbb{I}_3 & & & \\ \hline & & & -\overline{V}_\mu - \frac{1}{3}\Lambda_\mu \mathbb{I}_3 & 0 & \\ & & & 0 & -\overline{V}_\mu - \frac{1}{3}\Lambda_\mu \mathbb{I}_3 & \\ \hline & & & & & -\mathbb{I}_2 \otimes \overline{V}_\mu - \frac{1}{3}\Lambda_\mu \mathbb{I}_2 \otimes \mathbb{I}_3 \end{array}\right).$$

To complete the fluctuated Dirac operator need an expression for $J_F A_\mu J_F^*$. This sandwiching between J_F and J_F^* results in exchanging the upper left block for the lower left block and taking the complex conjugate at the same time.

$$JA_\mu J^*|_{\mathcal{H}_F^q \oplus \mathcal{H}_F^{\overline{q}}} = \tag{7.8.34}$$

$$\left(\begin{array}{cc|c|cc|c} -V_\mu - \frac{1}{3}\Lambda_\mu \mathbb{I}_3 & 0 & & & & \\ 0 & -V_\mu - \frac{1}{3}\Lambda_\mu \mathbb{I}_3 & & & & \\ \hline & & -\mathbb{I}_2 \otimes V_\mu - \frac{1}{3}\Lambda_\mu \mathbb{I}_2 \otimes \mathbb{I}_3 & & & \\ \hline & & & \Lambda_\mu \mathbb{I}_3 & 0 & \\ & & & 0 & -\Lambda_\mu \mathbb{I}_3 & \\ \hline & & & & & \overline{Q}_\mu \otimes \mathbb{I}_3 \end{array}\right).$$

We can now calculate B_μ to be

$$B_\mu|_{\mathcal{H}_F^q} = \left(\begin{array}{cc|c} V_\mu + \frac{4}{3}\Lambda_\mu \mathbb{I}_3 & 0 & \\ 0 & V_\mu - \frac{2}{3}\Lambda_\mu \mathbb{I}_3 & \\ \hline & & \begin{array}{c}(Q_\mu + \frac{1}{3}\Lambda_\mu \mathbb{I}_2) \otimes \mathbb{I}_3 \\ +\mathbb{I}_2 \otimes V_\mu\end{array} \end{array}\right) \tag{7.8.35}$$

$$B_\mu|_{\mathcal{H}_F^{\overline{q}}} = \left(\begin{array}{cc|c} -\overline{V}_\mu - \frac{4}{3}\Lambda_\mu \mathbb{I}_3 & 0 & \\ 0 & -\overline{V}_\mu + \frac{2}{3}\Lambda_\mu \mathbb{I}_3 & \\ \hline & & \begin{array}{c}-(\overline{Q}_\mu + \frac{1}{3}\Lambda_\mu \mathbb{I}_2) \otimes \mathbb{I}_3 \\ -\mathbb{I}_2 \otimes \overline{V}_\mu\end{array} \end{array}\right). \tag{7.8.36}$$

The structure of the ϕ field is very similar to what we found for the electroweak theory. To calculate the inner fluctuations of the finite Dirac operator D_F we use $\phi = \sum_j a_j [D_F, b_j]$. We

define the upper right parts of a_j and b_j by

$$a_j^l = a_j|_{(\mathcal{H}_F^l)} = \begin{pmatrix} \lambda_j & 0 & 0 & 0 \\ 0 & \overline{\lambda}_j & 0 & 0 \\ 0 & 0 & \alpha_j & \beta_j \\ 0 & 0 & -\overline{\beta}_j & \alpha_j \end{pmatrix} \quad b_j^l = b_j|_{(\mathcal{H}_F^l)} = \begin{pmatrix} \lambda'_j & 0 & 0 & 0 \\ 0 & \overline{\lambda}'_j & 0 & 0 \\ 0 & 0 & \alpha'_j & \beta'_j \\ 0 & 0 & -\overline{\beta}'_j & \alpha'_j \end{pmatrix}. \tag{7.8.37}$$

$$a_j^q = a_j|_{(\mathcal{H}_F^q)} = \begin{pmatrix} \lambda_j & 0 & 0 & 0 \\ 0 & \overline{\lambda}_j & 0 & 0 \\ 0 & 0 & \alpha_j & \beta_j \\ 0 & 0 & -\overline{\beta}_j & \alpha_j \end{pmatrix} \otimes \mathbb{I}_3 \quad b_j^q = b_j|_{(\mathcal{H}_F^q)} = \begin{pmatrix} \lambda'_j & 0 & 0 & 0 \\ 0 & \overline{\lambda}'_j & 0 & 0 \\ 0 & 0 & \alpha'_j & \beta'_j \\ 0 & 0 & -\overline{\beta}'_j & \alpha'_j \end{pmatrix} \otimes \mathbb{I}_3. \tag{7.8.38}$$

Note that the b_j matrices have components marked with a prime. Only the S sub-matrix of D_F contributes to the commutator. From before we have

$$S_l = S|_{(\mathcal{H}_F^l)^{\oplus 3}} = \left(\begin{array}{cc|cc} 0 & 0 & Y_\nu^* & 0 \\ 0 & 0 & 0 & Y_e^* \\ \hline Y_\nu & 0 & 0 & 0 \\ 0 & Y_e & 0 & 0 \end{array}\right) \quad S_q = S|_{(\mathcal{H}_F^q)^{\oplus 3}} = \left(\begin{array}{cc|cc} 0 & 0 & Y_u^* & 0 \\ 0 & 0 & 0 & Y_d^* \\ \hline Y_u & 0 & 0 & 0 \\ 0 & Y_d & 0 & 0 \end{array}\right) \otimes \mathbb{I}_3, \tag{7.8.39}$$

where we have added the shorthand notation S_l and S_q. In $\phi = \sum_j a_j [D_F, b_j]$ we sum over matching indices, which means it is just the sum of the separate matrix expressions for each j. Using these input we see that the whole calculation can be summarized by defining the sub-matrices $\omega_l = \sum_j a_j^l \left[S_l, b_j^l\right]$ and $\omega_q = \sum_j a_j^q \left[S_q, b_j^q\right]$. In the remainder of this subsection we suppress the $\otimes \mathbb{I}_3$ factors. We calculate that

$$\omega_l = \begin{pmatrix} 0 & 0 & \overline{Y}_\nu \phi'_1 & \overline{Y}_\nu \phi'_2 \\ 0 & 0 & -\overline{Y}_e \overline{\phi}'_2 & \overline{Y}_e \overline{\phi}'_1 \\ Y_\nu \phi_1 & -Y_e \overline{\phi}_2 & 0 & 0 \\ Y_\nu \phi_2 & Y_e \overline{\phi}_1 & 0 & 0 \end{pmatrix}, \tag{7.8.40}$$

and

$$\omega_q = \begin{pmatrix} 0 & 0 & \overline{Y}_u \phi'_1 & \overline{Y}_u \phi'_2 \\ 0 & 0 & -\overline{Y}_d \overline{\phi}'_2 & \overline{Y}_d \overline{\phi}'_1 \\ Y_u \phi_1 & -Y_d \overline{\phi}_2 & 0 & 0 \\ Y_u \phi_2 & Y_d \overline{\phi}_1 & 0 & 0 \end{pmatrix}. \tag{7.8.41}$$

We have introduced four new scalar fields by defining

$$\phi_1 = \sum_j \alpha_j(\lambda'_j - \alpha'_j) + \beta_j \overline{\beta}'_j \tag{7.8.42}$$

$$\phi'_1 = \sum_j \lambda_j(\alpha'_j - \lambda'_j) \tag{7.8.43}$$

$$\phi_2 = \sum_j \overline{\alpha}_j \overline{\beta}'_j - \overline{\beta}_j(\lambda'_j - \alpha'_j) \tag{7.8.44}$$

$$\phi'_2 = \sum_j \lambda_j \beta'_j. \tag{7.8.45}$$

We define the matrix ω as

$$\omega = \begin{pmatrix} \omega_l & 0 \\ 0 & \omega_q \end{pmatrix}. \tag{7.8.46}$$

The complete ϕ-matrix can be expressed as

$$\phi = \sum_j a_j \left[D_F, b_j\right] = \begin{pmatrix} \omega & 0 \\ 0 & 0 \end{pmatrix}. \tag{7.8.47}$$

We require ϕ to be self-adjoint which implies that ω must be self-adjoint. This means that $\phi_2 = \overline{\phi}_2'$ and $\phi_1 = \overline{\phi}_1'$. There are now only two complex degrees of freedom in ω. We modify ω_l and ω_q so that they are explicitly self-adjoint. We find that

$$\omega_l = \begin{pmatrix} 0 & 0 & \overline{Y}_\nu\overline{\phi}_1 & \overline{Y}_\nu\overline{\phi}_2 \\ 0 & 0 & -\overline{Y}_e\phi_2 & \overline{Y}_e\phi_1 \\ Y_\nu\phi_1 & -Y_e\overline{\phi}_2 & 0 & 0 \\ Y_\nu\phi_2 & Y_e\overline{\phi}_1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \omega_q = \begin{pmatrix} 0 & 0 & \overline{Y}_u\overline{\phi}_1 & \overline{Y}_u\overline{\phi}_2 \\ 0 & 0 & -\overline{Y}_d\phi_2 & \overline{Y}_d\phi_1 \\ Y_u\phi_1 & -Y_d\overline{\phi}_2 & 0 & 0 \\ Y_u\phi_2 & Y_d\overline{\phi}_1 & 0 & 0 \end{pmatrix}. \tag{7.8.48}$$

We can now find Φ from the general expression $\Phi = D_F + \phi + J_F\phi J_F^*$. A short calculation shows that

$$\begin{aligned} \Phi &= D_F + \phi + J_F\phi J_F^* && (7.8.49)\\ &= \begin{pmatrix} S & T^* \\ T & \overline{S} \end{pmatrix} + \begin{pmatrix} \omega & 0 \\ 0 & 0 \end{pmatrix} + J_F \begin{pmatrix} \omega & 0 \\ 0 & 0 \end{pmatrix} J_F^* && (7.8.50)\\ &= \begin{pmatrix} S & T^* \\ T & \overline{S} \end{pmatrix} + \begin{pmatrix} \omega & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \overline{\omega} \end{pmatrix} && (7.8.51)\\ &= \begin{pmatrix} S+\omega & T^* \\ T & \overline{S+\omega} \end{pmatrix}. && (7.8.52) \end{aligned}$$

The explicit form of the top left part of the Φ matrix is given by

$$\Phi|_{\mathcal{H}_F^l \oplus \mathcal{H}_F^q} = \left(\begin{array}{cc|cc} 0 & (Y_l + Y_{0,l})^* & & \\ Y_l + Y_{0,l} & 0 & & \\ \hline & & 0 & (Y_q + Y_{0,q})^* \\ & & Y_q + Y_{0,q} & 0 \end{array}\right), \tag{7.8.53}$$

where we have used

$$Y_l = \begin{pmatrix} Y_\nu & 0 \\ 0 & Y_e \end{pmatrix} \quad Y_{0,l} = \begin{pmatrix} Y_\nu\phi_1 & -Y_e\overline{\phi}_2 \\ Y_\nu\phi_2 & Y_e\overline{\phi}_1 \end{pmatrix} \quad Y_q = \begin{pmatrix} Y_u & 0 \\ 0 & Y_d \end{pmatrix} \quad Y_{0,q} = \begin{pmatrix} Y_u\phi_1 & -Y_d\overline{\phi}_2 \\ Y_u\phi_2 & Y_d\overline{\phi}_1 \end{pmatrix} \tag{7.8.54}$$

$$T_l = \left(\begin{array}{cc|cc} Y_R & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad T_q = \left(\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) T = \begin{pmatrix} T_l & 0 \\ 0 & T_q \end{pmatrix}. \tag{7.8.55}$$

Field strengths and traces

We have now defined and computed the necessary gauge fields. The next step is to define the field strengths of these gauge field. We define the field strengths as

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + i\left[B_\mu, B_\nu\right] \tag{7.8.56}$$
$$\Lambda_{\mu\nu} = \partial_\mu \Lambda_\nu - \partial_\nu \Lambda_\mu \tag{7.8.57}$$
$$Q_{\mu\nu} = \partial_\mu Q_\nu - \partial_\nu Q_\mu + i\left[Q_\mu, Q_\nu\right] \tag{7.8.58}$$
$$V_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu + i\left[V_\mu, V_\nu\right] \tag{7.8.59}$$
$$\tag{7.8.60}$$

We calculate $F_{\mu\nu}$ to be

$$F_{\mu\nu}|_{\mathcal{H}_F^l \oplus \mathcal{H}_F^{\bar{l}}} = \left(\begin{array}{cc|c|cc|c} 0 & 0 & & & & \\ 0 & -2\Lambda_{\mu\nu} & & & & \\ \hline & & -\Lambda_{\mu\nu}\mathbb{I} + Q_{\mu\nu} & & & \\ \hline & & & 0 & 0 & \\ & & & 0 & 2\Lambda_{\mu\nu} & \\ \hline & & & & & \Lambda_{\mu\nu}\mathbb{I} - \overline{Q}_{\mu\nu} \end{array}\right). \tag{7.8.61}$$

$$F_{\mu\nu}|_{\mathcal{H}_F^q} = \left(\begin{array}{cc|c} V_{\mu\nu} + \frac{4}{3}\Lambda_{\mu\nu}\mathbb{I}_3 & 0 & \\ 0 & V_{\mu\nu} - \frac{2}{3}\Lambda_{\mu\nu}\mathbb{I}_3 & \\ \hline & & (Q_{\mu\nu} + \frac{1}{3}\Lambda_{\mu\nu}\mathbb{I}_2) \otimes \mathbb{I}_3 \\ & & +\mathbb{I}_2 \otimes V_{\mu\nu} \end{array}\right) \tag{7.8.62}$$

$$F_{\mu\nu}|_{\mathcal{H}_F^{\bar{q}}} = \left(\begin{array}{cc|c} -\overline{V}_{\mu\nu} - \frac{4}{3}\Lambda_{\mu\nu}\mathbb{I}_3 & 0 & \\ 0 & -\overline{V}_{\mu\nu} + \frac{2}{3}\Lambda_{\mu\nu}\mathbb{I}_3 & \\ \hline & & -(\overline{Q}_{\mu\nu} + \frac{1}{3}\Lambda_{\mu\nu}\mathbb{I}_2) \otimes \mathbb{I}_3 \\ & & -\mathbb{I}_2 \otimes \overline{V}_{\mu\nu} \end{array}\right). \tag{7.8.63}$$

We will not write an explicit expression for $F_{\mu\nu}F^{\mu\nu}$ since the information we need can extracted directly from the matrices we already have written down. We square every cell and ignore crossterms since they will be zero in the trace. The leptonic trace is unchanged from the electroweak calculation except for a factor of 3 coming from three generations. The quark trace is

$$\left(\frac{4}{3}\right)^2 \times 3 + \left(-\frac{2}{3}\right)^2 \times 3 \left(\frac{1}{3}\right)^2 \times 2 \times 6 = \frac{16}{3} + \frac{4}{3} + \frac{2}{3} = \frac{22}{3}. \tag{7.8.64}$$

The results for antiquarks are exactly the same as only the sign of the prefactor changes. In addition we need a factor of 3 for the three generations. This gives a total quark contribution of $44\Lambda_{\mu\nu}\Lambda^{\mu\nu}$. From the leptonic side we have $12\Lambda_{\mu\nu}\Lambda^{\mu\nu}$ multiplied by three which is $36\Lambda_{\mu\nu}\Lambda^{\mu\nu}$. Together this gives total of $80\Lambda_{\mu\nu}\Lambda^{\mu\nu}$. By simple counting we see that per generation we get $8Q_{\mu\nu}Q^{\mu\nu}$ and

$8V_{\mu\nu}V^{\mu\nu}$. With three generations this becomes $24Q_{\mu\nu}Q^{\mu\nu}$ and $24V_{\mu\nu}V^{\mu\nu}$. We can thus write the expression for the trace as

$$Tr\left(F_{\mu\nu}F^{\mu\nu}\right) = 80\Lambda_{\mu\nu}\Lambda^{\mu\nu} + 24Tr\left(Q_{\mu\nu}Q^{\mu\nu}\right) + 24Tr\left(V_{\mu\nu}V^{\mu\nu}\right). \tag{7.8.65}$$

The calculation of Φ^2, Φ^4 and $D_\mu\Phi D^\mu\Phi$ is similar as for the electroweak case. We do not present the full derivation. The traces involving Φ are formally the same as for the electroweak case and are given as

$$\mathrm{Tr}\left(\Phi^2\right) = 4a|H|^2 + 2c \tag{7.8.66}$$
$$\mathrm{Tr}\left(\Phi^4\right) = 4b|H|^4 + 8e|H|^2 + 2d \tag{7.8.67}$$
$$\mathrm{Tr}\left(D_\mu\Phi D^\mu\Phi\right) = 4a|\tilde{D}_\mu H|^2. \tag{7.8.68}$$

The constants are now redefined slightly to include quarks and three generation and are given as

$$a = \mathrm{Tr}\left(Y_\nu^* Y_\nu + Y_e^* Y_e + 3Y_u^* Y_u + 3Y_d^* Y_d\right) \tag{7.8.69}$$
$$b = \mathrm{Tr}\left((Y_\nu Y_\nu^*)^2 + (Y_e Y_e^*)^2 + 3(Y_u Y_u^*)^2 + 3(Y_d Y_d^*)^2\right) \tag{7.8.70}$$
$$c = \mathrm{Tr}(Y_R Y_R^*) \tag{7.8.71}$$
$$d = \mathrm{Tr}\left((Y_R Y_R^*)^2\right) \tag{7.8.72}$$
$$e = \mathrm{Tr}\left((Y_R Y_R^*)(Y_\nu Y_\nu^*)\right). \tag{7.8.73}$$

Bosonic Lagrangian

The standard model bosonic Lagrangian can now be presented. Following the usual formula we get

$$S_b = \int \sqrt{g}\, d^4x\, L_{Total} \tag{7.8.74}$$
$$L_{Total} = 96L_{Gravity} + L_{Gauge} + L_{Higgs} \tag{7.8.75}$$
$$L_{Gravity} = \frac{f_4\Lambda^4}{2\pi^2} - \frac{f_2\Lambda^2}{24\pi^2}R + \frac{f(0)}{480\pi^2}\left(R^2 - 3R_{\mu\nu}R^{\mu\nu}\right) \tag{7.8.76}$$
$$L_{Gauge} = \frac{f(0)}{24\pi^2}\left(80\Lambda_{\mu\nu}\Lambda^{\mu\nu} + 24Tr\left(Q_{\mu\nu}Q^{\mu\nu}\right) + 24Tr\left(V_{\mu\nu}V^{\mu\nu}\right)\right) \tag{7.8.77}$$
$$L_{Higgs} = -\frac{2f_2\Lambda^2}{4\pi^2}(4a|H|^2 + 2c) + \frac{f(0)}{8\pi^2}(4b|H|^4 + 8e|H|^2 + 2d) \tag{7.8.78}$$
$$+ \frac{f(0)}{24\pi^2}\Delta(4a|H|^2 + 2c) + \frac{f(0)}{48\pi^2}R(4a|H|^2 + 2c) + \frac{f(0)}{8\pi^2}4a|\tilde{D}_\mu H|^2 \tag{7.8.79}$$

This completes the formal derivation of the bosonic Lagrangian. To complete the connection with ordinary Lagrangians found in particle physics one needs to perform some rewrites. These are basically the same as the one suggested for the electroweak theory. We will not repeat them nor will we specify the minor changes that are required.

Fermionic Lagrangian

The fermionic part of the Lagrangian follows from the fermionic action principle

$$S_f = \frac{1}{2}\langle J\xi, D_A\xi\rangle, \tag{7.8.80}$$

where $\xi \in \mathcal{H}^+$. The generic form of the Dirac operator on an almost commutative manifold is given by

$$D = D_M \otimes \mathbb{I} + \gamma^\mu \otimes B_\mu + \gamma_M \otimes \Phi \tag{7.8.81}$$

We have derived the form of B_μ and Φ for the SM in the previous subsections. A generic element of $\mathcal{H}^+$ can be written as

$$\begin{aligned}\xi =&\psi_{e_L^\lambda}\otimes e_L^\lambda+\psi_{e_R^\lambda}\otimes e_R^\lambda+\psi_{\overline{e}_R^\lambda}\otimes \overline{e}_L^\lambda+\psi_{\overline{e}_L^\lambda}\otimes \overline{e}_R^\lambda\\ &+\psi_{\nu_L^\lambda}\otimes \nu_L^\lambda+\psi_{\nu_R^\lambda}\otimes \nu_R^\lambda+\psi_{\overline{\nu}_R^\lambda}\otimes \overline{\nu}_L^\lambda+\psi_{\overline{\nu}_L^\lambda}\otimes \overline{\nu}_R^\lambda \qquad (7.8.82)\\ &+\psi_{u_L^{\lambda c}}\otimes u_L^{\lambda c}+\psi_{u_R^{\lambda c}}\otimes u_R^{\lambda c}+\psi_{\overline{u}_R^{\lambda c}}\otimes \overline{u}_L^{\lambda c}+\psi_{\overline{u}_L^{\lambda c}}\otimes \overline{u}_R^{\lambda c}\\ &\psi_{d_L^{\lambda c}}\otimes d_L^{\lambda c}+\psi_{d_R^{\lambda c}}\otimes d_R^{\lambda c}+\psi_{\overline{d}_R^{\lambda c}}\otimes \overline{d}_L^{\lambda c}+\psi_{\overline{d}_L^{\lambda c}}\otimes \overline{d}_R^{\lambda c} \qquad (7.8.83)\end{aligned}$$

The index $\lambda\in\{1,2,3\}$ labels the generation and $c\in\{R,G,B\}$ labels the color. Note that all the ψ's are classical spinors treated as Grassmann variables. Furthermore the subscripts and the overbars on the ψ's are just names. The L and R designate that ψ belongs to $\mathcal{H}_M^+$ or $\mathcal{H}_M^-$ respectively. This is independent of the overbar on the particle name.

We then need to calculate

$$\begin{aligned}S_f &= S_{kinetic}+S_{gauge}+S_{Yukawa} \qquad (7.8.84)\\ &= \frac{1}{2}\langle J\xi\,, D_M\otimes\mathbb{I}\xi\rangle+\frac{1}{2}\langle J\xi\,,\gamma^\mu\otimes B_\mu\xi\rangle+\frac{1}{2}\langle J\xi\,,\gamma_M\otimes\Phi\xi\rangle\,. \qquad (7.8.85)\end{aligned}$$

The first term, involving $D_M\otimes\mathbb{I}$, produces all the kinetic terms for the fermions. The term, involving $\gamma^\mu\otimes B_\mu$, produces all the gauge-fermion terms. The third term, involving $\gamma_M\otimes\Phi$, produces the Yukawa couplings between Higgs and fermions. The Yukawa couplings are determined by Y_e and Y_ν, as well as the fermion Majorana mass terms determined by Y_R.

We start by evaluating

$$S_{kinetic}=\frac{1}{2}\langle J\xi\,, D_M\otimes\mathbb{I}\xi\rangle\,. \qquad (7.8.86)$$

All the basis vectors of $\mathcal{H}_F$ are orthonormal. For this first part of the action, this orthonormality condition combined with the action of the finite part of $J=J_M\otimes J_F$ produces terms which match up $J_M\psi_{\overline{x}}$ with $D_M\psi_x$. We illustrate the logic behind the derivation for the case of neutrinos. It is clear that only neutrino "related" vectors will be relevant for this term. We therefore start with a less general ξ vector consisting of only neutrino terms for this illustration.

$$\begin{aligned}&\frac{1}{2}\langle J\xi\,, D_M\otimes\mathbb{I}\xi\rangle \qquad (7.8.87)\\ =&\frac{1}{2}\langle J_M\otimes J_F\left(\psi_{\nu_L^\lambda}\otimes\nu_L^\lambda+\psi_{\nu_R^\lambda}\otimes\nu_R^\lambda+\psi_{\overline{\nu}_R^\lambda}\otimes\overline{\nu}_L^\lambda+\psi_{\overline{\nu}_L^\lambda}\otimes\overline{\nu}_R^\lambda\right)\\ &\quad, D_M\otimes\mathbb{I}\left(\psi_{\nu_L^\lambda}\otimes\nu_L^\lambda+\psi_{\nu_R^\lambda}\otimes\nu_R^\lambda+\psi_{\overline{\nu}_R^\lambda}\otimes\overline{\nu}_L^\lambda+\psi_{\overline{\nu}_L^\lambda}\otimes\overline{\nu}_R^\lambda\right)\rangle \qquad (7.8.88)\\ =&\frac{1}{2}\langle J_M\otimes\mathbb{I}\left(\psi_{\nu_L^\lambda}\otimes\overline{\nu}_L^\lambda+\psi_{\nu_R^\lambda}\otimes\overline{\nu}_R^\lambda+\psi_{\overline{\nu}_R^\lambda}\otimes\nu_L^\lambda+\psi_{\overline{\nu}_L^\lambda}\otimes\nu_R^\lambda\right)\\ &\quad, D_M\otimes\mathbb{I}\left(\psi_{\nu_L^\lambda}\otimes\nu_L^\lambda+\psi_{\nu_R^\lambda}\otimes\nu_R^\lambda+\psi_{\overline{\nu}_R^\lambda}\otimes\overline{\nu}_L^\lambda+\psi_{\overline{\nu}_L^\lambda}\otimes\overline{\nu}_R^\lambda\right)\rangle \qquad (7.8.89)\\ =&\frac{1}{2}\langle J_M\psi_{\nu_L^\lambda}\,, D_M\psi_{\overline{\nu}_R^\lambda}\rangle+\frac{1}{2}\langle J_M\psi_{\nu_R^\lambda}\,, D_M\psi_{\overline{\nu}_L^\lambda}\rangle+\frac{1}{2}\langle J_M\psi_{\overline{\nu}_R^\lambda}\,, D_M\psi_{\nu_L^\lambda}\rangle\\ &\quad+\frac{1}{2}\langle J_M\psi_{\overline{\nu}_L^\lambda}\,, D_M\psi_{\nu_R^\lambda}\rangle \qquad (7.8.90)\\ =&\frac{1}{2}\langle J_M\psi_{\nu^\lambda}\,, D_M\psi_{\overline{\nu}^\lambda}\rangle+\frac{1}{2}\langle J_M\psi_{\overline{\nu}^\lambda}\,, D_M\psi_{\nu^\lambda}\rangle \qquad (7.8.91)\\ =&\langle J_M\psi_{\overline{\nu}^\lambda}\,, D_M\psi_{\nu^\lambda}\rangle \qquad (7.8.92)\end{aligned}$$

Where we have used the orthogonality of L and R spinors and the symmetry of the expression of the type $\langle J_M\psi_{\overline{\nu}^\lambda}\,, D_M\psi_{\nu^\lambda}\rangle$. The total set of terms from $\langle J\xi\,, D_M\otimes\mathbb{I}\xi\rangle$ follows the same pattern. Suppressing generation and color indices they are given as

$$\begin{aligned} S_{kinetic} &= \frac{1}{2}\langle J\xi\,, D_M\otimes\mathbb{I}\xi\rangle && (7.8.93)\\ &= \langle J\psi_{\overline{\nu}}\,, D_M\otimes\mathbb{I}\psi_\nu\rangle + \langle J\psi_{\overline{e}}\,, D_M\otimes\mathbb{I}\psi_e\rangle && (7.8.94)\\ &+ \langle J\psi_{\overline{u}}\,, D_M\otimes\mathbb{I}\psi_u\rangle + \langle J\psi_{\overline{d}}\,, D_M\otimes\mathbb{I}\psi_d\rangle\,. && (7.8.95)\end{aligned}$$

We next consider the term

$$S_{gauge} = \frac{1}{2}\langle J\xi\,, \gamma^\mu\otimes B_\mu\xi\rangle\,. \tag{7.8.96}$$

We illustrate the derivation of these terms by showing how it works for the leptonic sector. Basically the terms gives all the matrix elements of the matrix B_μ. We start by repeating B_μ with explicit components.

$$B_\mu|_{\mathcal{H}_l} = \left(\begin{array}{cc|cc|c} \overline{\nu}_R & \overline{e}_R & \overline{\nu}_L & \overline{e}_L & \\ \hline 0 & 0 & 0 & 0 & \nu_R \\ 0 & \Lambda_\mu & 0 & 0 & e_R \\ \hline 0 & 0 & W^3_\mu+\Lambda_\mu & W^1_\mu - iW^2_\mu & \nu_L \\ 0 & 0 & W^1_\mu + iW^2 & -W^3_\mu+\Lambda_\mu & e_L \end{array}\right) \tag{7.8.97}$$

We then want to rewrite this using the physical fields.

$$B_\mu|_{\mathcal{H}_l} = \left(\begin{array}{cc|cc|c} \overline{\nu}_R & \overline{e}_R & \overline{\nu}_L & \overline{e}_L & \\ \hline 0 & 0 & 0 & 0 & \nu_R \\ 0 & A_\mu+Z_\mu & 0 & 0 & e_R \\ \hline 0 & 0 & A_\mu+Z_\mu & W^*_\mu & \nu_L \\ 0 & 0 & W_\mu & A_\mu+Z_\mu & e_L \end{array}\right) \tag{7.8.98}$$

The terms involving A_μ will appear while the other will involve only the left-hand part of leptons and the right-hand part of anti-leptons. In the end we get

$$\begin{aligned} S_{gauge} &= \frac{1}{2}\langle J\xi\,, \gamma^\mu\otimes B_\mu\xi\rangle \\ &= s_w g_2 A_\mu\left(-\langle J_M\overline{e}^\lambda, \gamma^\mu e^\lambda\rangle\frac{2}{3}\langle J_M\overline{u}^\lambda, \gamma^\mu u^\lambda\rangle\right) \\ &- \frac{1}{3}\langle J_M\overline{d}^\lambda, \gamma^\mu d^\lambda\rangle + \frac{g_2}{4c_w}Z_\mu\left(\langle J_M\overline{e}^\lambda, \gamma^\mu e^\lambda\rangle + \langle J_M\overline{e}^\lambda, \gamma^\mu e^\lambda\rangle\right) \\ &+ \frac{g_2}{4c_w}W_\mu\left(\langle J_M\overline{e}^\lambda, \gamma^\mu e^\lambda\rangle + \langle J_M\overline{e}^\lambda, \gamma^\mu e^\lambda\rangle\right) \\ &+ \frac{g_2}{4c_w}W^*_\mu\left(\langle J_M\overline{e}^\lambda, \gamma^\mu e^\lambda\rangle + \langle J_M\overline{e}^\lambda, \gamma^\mu e^\lambda\rangle\right) \\ &+ \frac{g_2}{4c_w}G_\mu\left(\langle J_M\overline{e}^\lambda, \gamma^\mu e^\lambda\rangle + \langle J_M\overline{e}^\lambda, \gamma^\mu e^\lambda\rangle\right) \end{aligned} \tag{7.8.99}$$

Finally we consider the term

$$S_{Yukawa} = \frac{1}{2}\left\langle J\xi\,, \gamma_M \otimes \Phi\xi \right\rangle. \tag{7.8.100}$$

Again this basically picks out the various matrix elements of Φ. We define mass matrices m_x and m_R by setting

$$m_x = i\frac{\pi v}{\sqrt{af(0)}}Y_x \quad m_R = iY_R. \tag{7.8.101}$$

We rescale and parametrize the Higgs fields by setting

$$H = \begin{pmatrix}\phi_1 + 1\\ \phi_2\end{pmatrix} \to H = \sqrt{\frac{af(0)}{\pi^2}}\begin{pmatrix}\phi_1 + 1\\ \phi_2\end{pmatrix} = \begin{pmatrix}v + h + i\phi^0\\ i\sqrt{2}\phi^-\end{pmatrix}. \tag{7.8.102}$$

For the complex field ϕ^- we also define the notation $\phi^+ = \overline{(\phi^-)}$. Writing everything out we get

$$\begin{aligned}
S_{Yukawa} &= \frac{1}{2}\left\langle J\xi\,, \gamma_M \otimes \Phi\xi\right\rangle \\
&= i(1+\frac{h}{v})\left(\langle J_M\overline{\nu}^\lambda, m_\nu\nu^\lambda\rangle + \langle J_M\overline{e}^\lambda, m_e e^\lambda\rangle + \langle J_M\overline{u}^\lambda, m_u u^\lambda\rangle + \langle J_M\overline{d}^\lambda, m_d d^\lambda\rangle\right) \\
&+ \phi^0\left(\langle J_M\overline{\nu}^\lambda, m_\nu\nu^\lambda\rangle + \langle J_M\overline{e}^\lambda, m_e e^\lambda\rangle + \langle J_M\overline{u}^\lambda, m_u u^\lambda\rangle + \langle J_M\overline{d}^\lambda, m_d d^\lambda\rangle\right) \\
&+ \phi^-\left(\langle J_M\overline{\nu}^\lambda, m_\nu\nu^\lambda\rangle + \langle J_M\overline{e}^\lambda, m_e e^\lambda\rangle\right) \\
&+ \phi^+\left(\langle J_M\overline{\nu}^\lambda, m_\nu\nu^\lambda\rangle + \langle J_M\overline{e}^\lambda, m_e e^\lambda\rangle\right) \\
&+ \phi^-\left(\langle J_M\overline{\nu}^\lambda, m_\nu\nu^\lambda\rangle + \langle J_M\overline{e}^\lambda, m_e e^\lambda\rangle\right) \\
&+ \phi^+\left(\langle J_M\overline{\nu}^\lambda, m_R\nu^\lambda\rangle + \langle J_M\overline{e}^\lambda, m_R e^\lambda\rangle\right) \\
&+ \langle J_M\overline{\nu}^\lambda, m_R\nu^\lambda\rangle + \langle J_M\overline{e}^\lambda, m_R e^\lambda\rangle.
\end{aligned} \tag{7.8.103}$$

All the terms involving ϕ_0, ϕ^+ and ϕ^- disappear when we use the commonly used unitary gauge. Thus, only the first and last line in the above expression is usually shown when using this gauge.

We have derive the fermionic part of the Lagrangian of the form

$$S_f = S_{kinetic} + S_{gauge} + S_{Yukawa} \tag{7.8.104}$$

$$= \frac{1}{2}\left\langle J\xi\,, D_M \otimes \mathbb{I}\xi\right\rangle + \frac{1}{2}\left\langle J\xi\,, \gamma^\mu \otimes B_\mu\xi\right\rangle + \frac{1}{2}\left\langle J\xi\,, \gamma_M \otimes \Phi\xi\right\rangle. \tag{7.8.105}$$

We obtained expressions that were in the form of inner products on $L^2(M,S)$. We can also express this in terms of the fiberwise inner product. One example of this would be

$$\left\langle J\psi_{\overline{e}}\,, D_M \otimes \mathbb{I}\psi_e\right\rangle = \int \sqrt{g}\, d^4x\, (J\psi_{\overline{e}}, D_M \otimes \mathbb{I}\psi_e). \tag{7.8.106}$$

By this manner we obtain the usual formulation of the action as an integral. This completes the derivation of the fermionic part of the Lagrangian for the standard model.

7.9 The Pati-Salam model

The direct implementation of the standard model in NCG is quite economical but leads to problems with the observed Higgs mass [140]. This can be dealt by including an extra scalar field, which couples to the Higgs field, which had been previously excluded for various reasons [140].

A more satisfying interpretation of the origin of this extra scalar field comes from redoing the algebra derivation with the benefit of these later insights. The current interpretation of the derivation of the standard model is that one or more unnecessary (or unwarranted) assumptions have been used in the derivation thereby restricting the choice of algebras too much. Dropping the first-order assumption leads instead to the algebra $M_2(\mathbb{H}) \oplus M_4(\mathbb{C})$ breaking to $\mathbb{H} \oplus \mathbb{H} \oplus M_4(\mathbb{C})$. This leads to a semi-GUT with the symmetry $SU(2) \oplus SU(2) \oplus SU(4)$. This model is known as the Pati-Salam model and was first presented (independent of any NCG context) in [129]. The full derivation within NCG can be found in [130]. Se also [123] for a further analysis of model with respect coupling unification at high energies.

The first idea of the Pati-Salam model is to unify leptons with quarks by considering leptons as a "fourth color" of quarks [129]. The second idea is to treat right-handed and left-handed fermions symmetrically [129]. So, in simple terms, the SU(4) represents the new quark-lepton color symmetry while the SU(2)xSU(2) represents the symmetric treatment of the left and right with respect to the weak interaction. We also mention that Pati-Salam can be conveniently embedded in SO(10). Several breaking patterns from Pati-Salam down to the standard model are available [130][123]. The gauge group is not simple but it still represents an attractive model in terms of elegance and economy.

Due to space and time limitations we will not present the derivation of this model, or a more detailed presentation of its properties.

7.10 Notes

Gravity terms

The bosonic terms that are of the type we could call gravity terms and the bosonic terms that are of the type we could call gauge and scalar terms, do not separate completely in the bosonic action principle. There are extra gravitational terms added when we develop the primarily gauge and scalar related terms. These terms are terms of the exact same form as those already found in the EH-action plus, or are extra terms which can be ignored at low energies. To present the appearance of the SM in CLC-models most readily we suppress such "extra" terms (which end up being absorbed or ignored in most analyses).

IV Extended theories II - Loop quantum gravity

8. Loop quantum gravity

> The most conservative approach to quantum gravity is to seek a theory that combines the best features of general relativity and quantum field theory. To do this, we must try to find a background-free quantum theory with local degrees of freedom propagating causally.
>
> Baez

Summary

LQG is both a specific quantization of GR, and a general set of methods for quantizing connection based theories in a background free manner. LQG uses a formulation of GR based on connections smeared along one-dimensional curves. The classical theory is defined solely by constraints. Following Dirac's method for quantizing fully constrained systems, we first quantize the unconstrained theory, and then implement the constraints after quantization. LQG leads to a mathematically rigorous picture of quantum geometry, where area and volume operators can be defined and take discrete values. In LQG there is no spacetime at the fundamental level. Spacetime is an emergent phenomenon, most likely appearing as a specific phase in the classical macroscopic limit of the theory. As there is no time variable, the dynamics of the theory must be expressed by relational constructions. Application of LQG to cosmology yields a bounce model where the Big Bang singularity is eliminated.

8.1 Introduction

Constructing a quantum theory of gravity is one of the major unsolved problems in theoretical physics. The difficulty stems from the unique properties of gravity that set it apart from other forces (see Chapter 4). The opposition between gravity (GR) and other forces (QFT) can be seen

in the top panel of figure 8.1. It is not just the obvious conflict between classical concepts and quantum concepts, but also the strong tension between background dependent theories (QFT) and background free theories (GR). In the experimentally accessible regions of physics, gravity effects and quantum effects are quite well separated (see the left-hand and middle part of figure 8.2). However, in the natural domain of fundamental physics they are both important (see the right-hand part of figure 8.2). The various traditional approaches to quantization are still available in the gravity case, but their straightforward application results in only partial insights into a possible quantum theory. Applying the typical perturbative expansion around an arbitrarily chosen background metric yields a non-renormalizable theory, and this also breaks (or hides) the background free property of the theory. Approaches such as supergravity theory have made progress, but in addition to requiring new particles, extra dimensions, and supersymmetry, these theories are still perturbative, divergent and non-renormalizable (with the possible exception of $\mathcal{N} = 8$ $D = 4$ which could be finite) [141]. String theory is suggested to be a finite UV-completion of perturbative gravity [142][143]. It is certainly a rich and powerful theory that includes spin-2 gravity and avoids divergences. However, its most well-known formulation relies on the presence of a fixed background metric that is alien to GR concepts. String theory has led to many new insights, but it has perhaps not yet fully tackled the problem of defining a quantum geometry (see chapter 9 and 10 for further discussions).

In this chapter we will present the theory of loop quantum gravity (LQG). LQG is a distinct approach to quantizing gravity by literally combining the principles of quantum physics and GR (see figure 8.1). LQG constructs a non-perturbative background independent QFT based on GR. It does so in a more minimalistic fashion than other approaches. LQG is compatible with supersymmetry, extra dimensions and extra matter fields, but it does not require them [144][77]. LQG is a theory of pure quantum gravity that does not (per se) involve unification of gravity with other forces. LQG is compatible with the matter found in the SM, but does not yet contribute any insights on the matter aspect [145][146][77].

Constructing LQG requires a number of steps, several of which can involve unfamiliar mathematical techniques. Our presentation of LQG will follow the outline presented in figure 8.3, and we now give a brief overview.

GR is often formulated as a theory for determining the metric of spacetime [45]. The strategy of LQG is to reformulate classical GR in a manner which is closer to ordinary gauge theory. This involves using the connection as the primary variable instead of the metric [147][25]. It also uses the Hamiltonian formulation of GR, as this allows the use of canonical quantization techniques. While using a Hamiltonian formulation makes diffeomorphism invariance less apparent, it does preserve background independence [25].

The reformulation of the classical theory consists of three steps (which we cover in section 8.4). The first step is the Legendre transformation of the Einstein-Hilbert action. This produces the Arnowitt-Deser-Misner action (ADM action), which employs the three-dimensional metric and its conjugate momentum as phase space variables. The second step transforms the ADM formulation into a formulation using the extrinsic curvature and the densitized triads as variables. Using triads introduces an additional $\mathrm{SO}(3)$ (or $\mathrm{SU}(2)$) rotational redundancy. In the third and final step the extrinsic curvature is exchanged for the Ashtekar-Barbero connection, while the densitized triads are kept unchanged. This formulation is called the Ashtekar formulation, and makes GR look (kinematically) like standard Yang-Mills $\mathrm{SU}(2)$ gauge theory with extra constraints. The constraints are called the Gauss constraint, the diffeomorphism constraint, and the Hamiltonian constraint.

Direct quantization of the Ashtekar formulation using connections and triads has not been successful. Instead, new smeared versions of the variables are introduced, and these are then used for quantization (section 8.5). The connection variables are smeared along graphs, and the triads are smeared over two-dimensional surfaces. These smearings, which consists of integrations, are defined without using a background metric. The smeared versions of the connection and the triads

are referred to as holonomies and fluxes respectively.

Having obtained a suitable classical formulation, in the form of the Poisson algebra of holonomies and fluxes, the theory is then canonically quantized (section 8.6). This is done by constructing a Hilbert space and defining an action of these holonomies and fluxes on the Hilbert space states. After first quantizing the theory without enforcing any of the constraints, we move on to applying the quantum version of the constraints one-by-one (section 8.7). The Gauss constraint, the diffeomorphism constraint, and the Hamiltonian constraint are tackled in that order. When solving the Gauss constraint we find a solution space consisting of $\mathrm{SU}(2)$ invariant spin-networks states. These states are fundamentally important quantum geometric states with discrete valued volumes and areas. Next, we use the diffeomorphism constraint to find diffeomorphism invariant states referred to as s-knot states. These are closely related to knot invariants of knot theory. Finally we arrive at the major outstanding problem of LQG, the issue of implementing the Hamiltonian constraint and thereby defining the true physical states of the theory. Several techniques exist, but this problem is not yet completely solved. Having thus in principle determined the physical states of the theory, we then discuss how to use relational methods to extract actual dynamics from these states (section 8.8).

The inability to explicitly solve the Hamiltonian constraint in the general setting, limits our ability to apply the theory to arbitrary problems. We therefore seek simpler settings, to see if progress can be made. Cosmology provides one such simplified setting (section 8.11), and we use it to give a concrete example of solving the theory and expressing actual dynamics [148].

Before starting the construction of LQG we review some general issues regarding classical theories with constraints (section 8.2) and their quantization (section 8.3)

8.2 Classical theories with constraints

Ordinary gauge theories and GR are examples of theories with redundancies. We will refer to the redundancies in both theories as gauge redundancies. Such theories require a delicate handling, both in the Lagrangian and the Hamiltonian formulation. In the Lagrangian case redundancies leads to constraint equations and under-determined Cauchy problems. In the Hamiltonian case it leads to a non-invertible Legendre transformations and to constraint equations. We need to handle gauge redundancies when developing LQG, and we start out with a general analysis of how to address such complications. There are few properly mathematical presentations of analytical mechanics, especially when it comes to field theory. We base our presentation on [25, chapter 24], [26, chapter 1], [27], [28] and [29].

We have already stated our preferred approach to physical systems in chapter 2. This algebraic view of physics will be a center point of our understanding of LQG. However, the abstract formalism of chapter 2 will now be implemented in a more concrete setting in the Hamiltonian framework. The Hamiltonian framework builds on the Lagrangian framework. To establish the proper setting for doing physics in the algebraic manner, we need to review both the Lagrangian and the Hamiltonian approach in more detail. Despite our arguments (in chapter 2) that we should allow general probability distributions as the proper states of classical mechanics , in this section we will focus on the special case of states that can be represented by points (Dirac measures).

Lagrangian mechanics

First we will discuss the mathematical setup for Lagrangian mechanics. Lagrangian mechanics describes the state of a system by points in a smooth n-dimensional manifold M. We let points of M be denoted by $q = (q_1, q_2, ..., q_n)$, and we define $C = M$ as the configuration space of the system. The coordinates of the points of M are called generalized coordinates. A parametrized

Before combining

QFT
- Basic structure
  - Noncommutative algebra
  - States (Hilbert space)
  - Global time parameter
- Geometry and symmetry
  - Fixed geometry
  - Global O(1,3) symmetry
  - Not diffeomorphism invariant

GR
- Basic structure
  - Commutative algebra
  - States (phase space)
  - No time parameter
- Geometry and symmetry
  - No (prior) geometry
  - No isometries
  - Diffeomorphism invariant

After combining

QFT
- Basic structure
  - Noncommutative algebra (operators)
  - States (Hilbert space)
  - ~~Global time parameter~~
- Geometry and symmetry
  - ~~Fixed geometry~~
  - ~~Global O(1,3) symmetry~~
  - ~~Not diffeomorphism invariant~~

GR
- Basic structure
  - ~~Commutative algebra~~
  - ~~States (Phase space)~~
  - No time parameter
- Geometry and symmetry
  - No (prior) geometry
  - No isometries
  - Diffeomorphism invariant

Figure 8.1: The diagram shows the basic structure of QFT and GR. The top row shows the structure of each of the two theories by themselves (before combining them). The bottom row shows which parts should (ideally) be chosen in a theory combining these two structures. Note that many tools from fixed geometry physics are excluded by this choice. Much of the utility of QFT is based on having these tools available. This fact partly explains why building quantum gravity by this direct approach requires several new or extended tools.

path is a map

$$\gamma\colon I \to M. \tag{8.2.1}$$

The derivatives $d\gamma(t)/dt$ of a parametrized path γ are called generalized velocities. It is natural to describe the set of coordinates and velocities by using the the tangent bundle $TC = TM$. We usually think of this bundle in terms of some local trivialization, and view this space as $TC = TM = C \times C = M \times M$. A point in TC is denoted in the standard fashion by $(q_1, q_2, ..., q_n, v_1, v_2, ..., v_n)$. A generic path in TC is a map

$$\gamma'\colon I \to TC. \tag{8.2.2}$$

We let prime or no prime distinguish between paths in TC and C respectively. We consider only paths γ' in TC that are lifts of paths γ in M (a path where the second set of coordinates represent the derivative of the path). For a given path

$$\gamma\colon I \to M, \quad \gamma\colon t \mapsto q(t) \tag{8.2.3}$$

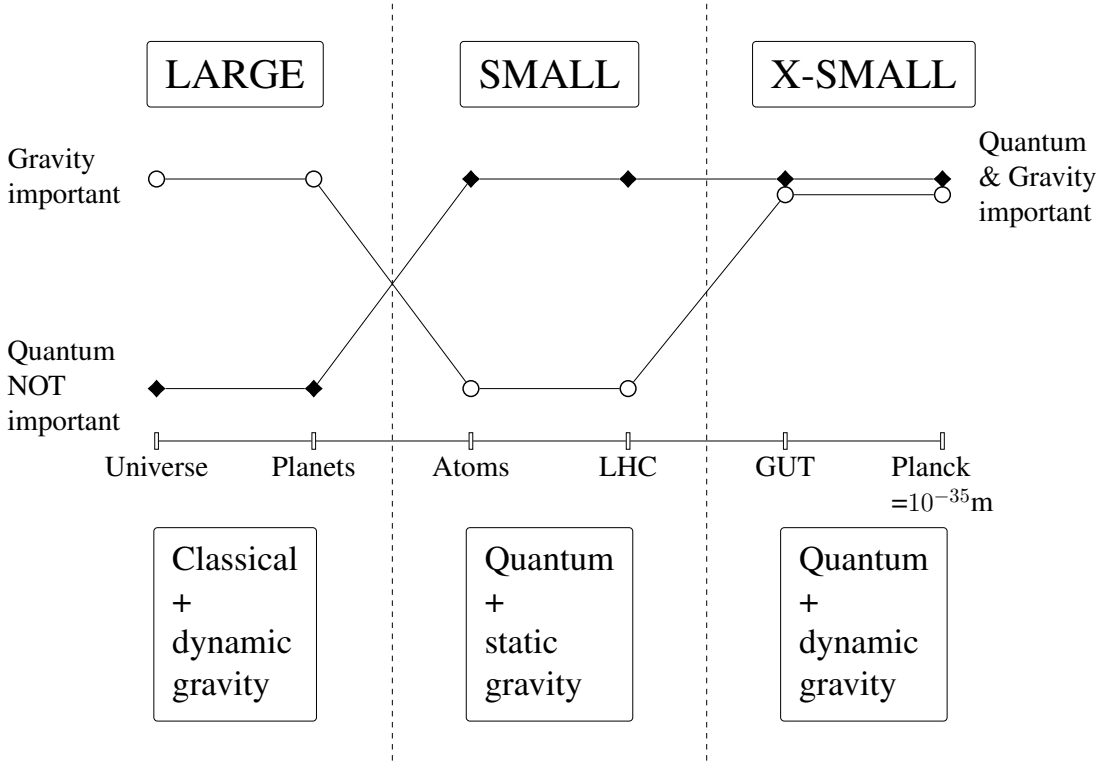


Figure 8.2: The diagram shows how quantum effects and gravity effects separate quite nicely in the experimentally accessible domains of physics, denoted here as the large and small domain. However, in the natural domain of fundamental physics, denoted here as the extra small domain (x-small), there is no separation and both quantum and gravity effects are important. Note that the curves are illustrations of some basic principles and the "data points" are not to be taken literally. The red line with diamond markings refer to quantum effects, and the blue line with circles refer to gravity effects. A low position is meant to indicate low importance, and a high position indicates high importance. By static gravity we mean situations where the dynamic aspect of gravity can be ignored. This is usually described as "gravity can be ignored", but since gravity is spacetime structure, denoting this as static gravity seems more accurate. The universe is taken to be the current (and very big) universe. The universe at very early times is of course quite another matter.

we define the lifted path in TC to be

$$I \ni t \mapsto (q(t), \dot{q}(t)) \in TC \quad \text{with} \quad \dot{q}(t) = \frac{\mathrm{d}q}{\mathrm{d}t}. \tag{8.2.4}$$

A Lagrangian is a map

$$L(q, v) : TC \to \mathbb{R}.$$

The composition of a lifted path γ' with the Lagrangian L is a map

$$\left(L \circ \gamma'\right) : I \to \mathbb{R} \tag{8.2.5}$$

$$\left(L \circ \gamma'\right) : t \mapsto L\left(\gamma'(t)\right). \tag{8.2.6}$$

We now set $I = [t_0, t_1]$. The path space $P(C)$ of C, with fixed endpoints (q_0, t_0) and (q_1, t_1), is defined as

$$P(C)^{q_1,t_1}_{q_0,t_0} = \{\gamma : [t_0, t_1] \to C \mid \gamma(t_0) = q_0, \gamma(t_1) = q_1\}. \tag{8.2.7}$$

A variation (with fixed endpoints) Γ of the path $\gamma(t)$ is a family of paths $\gamma_\epsilon(t) = \Gamma(t, \epsilon)$ parameterized by $\epsilon \in [-\epsilon_0, \epsilon_1]$. We can express this family of paths as a map

$$\Gamma : [t_0, t_1] \times [-\epsilon_0, \epsilon_1] \to C, \tag{8.2.8}$$

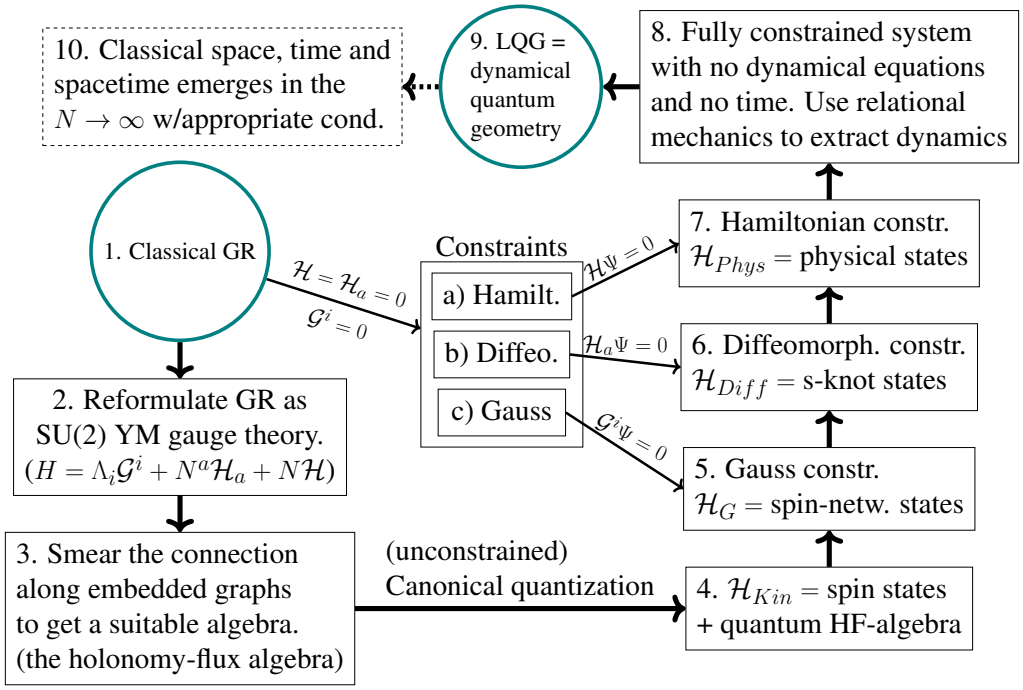


Figure 8.3: The diagram shows how loop quantum gravity (LQG) is constructed from general relativity (GR). The progression is a counterclockwise circle starting with classical GR in step 1, and ending up with LQG as a theory of dynamical quantum geometry in step 9. For more on step 1 see chapter 4. Step 2 (see section 8.4) performs the Ashtekar reformulation of GR as an SU(2) Yang-Mills (YM) gauge theory. The resulting Hamiltonian is a sum of constraints as shown in the parenthesis. Then, in step 3 (see section 8.5), the SU(2) gauge connection variables are smeared along one-dimensional embedded graphs to obtain the classical holonomy-flux (HF) algebra. The HF-algebra is more suitable for quantization than the connection itself. The HF-algebra is canonically quantized, disregarding the constraints, in step 4 (see section 8.3 and 8.6) . The constraints are then successively applied to the Hilbert space as quantum constraints in steps 5, 6 and 7 (see section 8.7). Spin-states, spin-network states, s-knot states and physical states, are semi-official names for the resulting quantum states at the various stages of applying the constraints. Finally, in step 8 (see section 8.8), we use relational methods to extract dynamics since the Hamiltonian vanishes on the constraint surface, and there is no physical time parameter. The inclusion of the extra step, step 10 (see section 8.9), underscores the importance of being able to recover the classical theory in the appropriate limit. The equations above and below the lines to and from the box of constraints in the middle, are the classical constraint equations on the left, and the corresponding quantum constraint equations on the right. On the left $\mathcal{H}, \mathcal{H}, \mathcal{G}$ represent functions, and on the right they represent the corresponding operators. These symbols are also used in the expression for the Hamiltonian in box 2. The symbols Λ_i, N^a, N denote (arbitrary) Lagrange multipliers. By $N \to \infty$ we mean the macroscopic limit of infinitely many "quanta". Abbreviations: YM= Yang Mills, Hamilt. = Hamiltonian constraint, diffeo. = diffeomorphism constraint, constr. = constraint, cond. = conditions w = with

such that $\Gamma(t, 0) = \gamma(t)$ and $\Gamma(t_0, \epsilon) = q_0$ and $\Gamma(t_1, \epsilon) = q_1$. The action functional is a map from paths into the real numbers.

Definition 8.2.1 The action functional for a given Lagrangian L with path space

$$P = P\left(C\right)_{q_0,t_0}^{q_1,t_1} \tag{8.2.9}$$

is a map $S_L \colon P \to \mathbb{R}$ defined by

$$S_L(\gamma) = \int_{t_0}^{t_1} L(\gamma')dt. \tag{8.2.10}$$

Note the prime in definition 8.2.1 that indicates that γ' is the lift of γ. The classical paths of a Lagrangian system are defined to be the paths in C that are stationary points of the action functional. A path γ is a stationary point of the action if for all variations γ_ϵ (with fixed endpoints) of the path γ we have

$$\left.\frac{\mathrm{d}S_L(\gamma_\epsilon)}{\mathrm{d}\epsilon}\right|_{\epsilon=0} = 0. \tag{8.2.11}$$

For a Lagrangian that depends only on the path and its corresponding lifted path we get the following expression for the classical path(s).

Theorem 8.2.2 The stationary paths of a Lagrangian system are given by solutions of

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = 0. \tag{8.2.12}$$

The boundary conditions are as for the path space $P = P\,(C)^{q_1,t_1}_{q_0,t_0}$.

Lagrangian field theory

Lagrangian mechanics applied to fields is in principle quite similar to the point particle case described above. The difficulty lies in the fact that we are now dealing with systems with an infinite number of degrees of freedom. This means that to be rigorous we have to deal with infinite dimensional manifolds. We shall not dive too far into this area, but rather emphasize some aspects of this, to make the structure of theory as clear as possible. We shall at times rely on the typical physicist method of just imagining that everything is finite dimensional, and that it is therefore well defined, and then proceeding by analogy to the infinite dimensional case without worrying to much about the problems that infinite dimensional structures present. We will use complex (or real) scalar fields in four spacetime dimensions as our prime example to keep things as simple as possible. Let us denote the space portion of our chosen spacetime (in some give frame of reference) by M. We will not worry about relativistic covariance properties as our focus here is on a well defined algebraic description of the system. The configuration space of field theory is now the set of all possible scalar fields on M.

Definition 8.2.3 — Configuration space. The configuration space C of a Lagrangian complex scalar field theory on the space M is the function space

$$C = C^\infty(M, \mathbb{C}). \tag{8.2.13}$$

Paths and their lifts are defined as before. A path is a map $\gamma : I \to C$. and a lifted path is a map $\gamma : I \to TC$. For TC we always imagine dealing with a local or global trivialization, which means that $TC = C \times C$. A Lagrangian for a field theory is a map

$$L(q, v) : TC \to \mathbb{R}, \tag{8.2.14}$$

just like it is for particle theory. The variational problem is dealt with analogously to the particle case. By using a more complex definition of the variational problem we end up with a similar solution. For a Lagrangian that depends only on the path at its lift we get the following expression for the stationary path(s).

Theorem 8.2.4 The stationary paths of a Lagrangian system are given by

$$\frac{\delta L}{\delta \Phi} - \partial_\mu \frac{\delta L}{\delta(\partial_\mu \Phi)} = 0. \tag{8.2.15}$$

We will not enter into discussion of a proper definition of functional derivatives like $\frac{\delta L}{\delta \Phi}$.

Hamiltonian mechanics

For our review of Hamiltonian mechanics we will mostly consider the particle version. We need two definitions to derive Hamiltonian mechanics from a given Lagrangian system. Note that this transition does not depend on time derivatives (v is just a coordinate on TC).

Definition 8.2.5 — Conjugate momentum. Given a Lagrangian system with N generalized coordinates $q_1, \ldots, q_N$ associated velocities $v_1, \ldots, v_N$ the conjugate momentum of q_i is denoted p_i and is defined as

$$p_i = \frac{\partial \mathcal{L}}{\partial v^i}. \tag{8.2.16}$$

Definition 8.2.6 — Hamiltonian. Given a Lagrangian system with N generalized coordinates $q_1, \ldots, q_N$ and associated velocities $v_1, \ldots, v_N$. Let the conjugate moment $p_1, \ldots, p_N$ be defined as given by definition 8.2.5. We define the Hamiltonian of the system by

$$\mathcal{H} = p_i v^i(p,q) - \mathcal{L}\left(p,q\right). \tag{8.2.17}$$

Here $v^i(p,q)$ is the the solution of $p_i = \frac{\partial \mathcal{L}}{\partial v^i}$ with respect to v^i. (Note that we are assuming invertibility.)

Hamiltonian mechanics is the Legendre transformed version Lagrangian mechanics. The Legendre transformation is a map

$$\text{Legendre}\colon TM \to T^*M \tag{8.2.18}$$

$$\text{Legendre}\colon (q,v) \mapsto (q,p). \tag{8.2.19}$$

We are now doing physics on the cotangent bundle instead of on the tangent bundle. We call $\mathrm{Ph} = T^*M$ the phase space of the system. (We will again assume that we can operate on the local trivialization of this bundle where $\mathrm{Ph} = T^*M = C \times C$.) States are now points in phase space, or more generally they are probability measures on phase space (see chapter 2). The set of all observables is the function space

$$\mathcal{O} = C^\infty(\mathrm{Ph}, \mathbb{R}). \tag{8.2.20}$$

The set of observables $\mathcal{O}$ forms an algebra defined by point-wise addition, multiplication, and scalar multiplication in the codomain of the functions. We will call the algebra just defined the associative algebra part of the classical algebra of observables. We usually define an extra (non-associative) algebraic operation on $\mathcal{O}$, called the Poisson bracket. The Poisson bracket is a binary operation

$$\{,\} : \mathcal{O} \times \mathcal{O} \to \mathcal{O}, \tag{8.2.21}$$

that gives $\mathcal{O}$ a Lie algebra structure. The Poisson bracket is also required to be a derivation with respect to the associative algebra. The Poisson Lie algebra operation is non-associative and anti-symmetric (as are all Lie algebras).

By introducing the Poisson bracket we have implicitly made T^*M into a symplectic manifold, with an associated commutative Poisson algebra A. Often one will go in the opposite direction and derive the Poisson structure from a symplectic structure. As emphasized in chapter 2 the Poisson algebraic structure is the true fundamental object of the theory. The Poisson bracket is a derivation so that an arbitrary function $f \in \mathcal{O}$ will implicitly define a Hamiltonian vector field by using exterior derivation and the one-form vector duality map (see appendix F.4 for more details). Any such Hamiltonian vector field can be integrated to give the corresponding integral curves that determines the flows corresponding to the chosen function. The dynamics of a system is usually specified by designating a function as the so called Hamiltonian of the system, and taking the flows defined by Hamiltonian vector fields of the Hamiltonian to represent dynamic change. Any function may be arbitrarily designated as the Hamiltonian of the system, but usually the choice of such a function is based on a Lagrangian, which itself is based on physical principles. We shall shortly see that the flow defined by the Hamiltonian does not always correspond to the proper physical time-evolution of the system, but that sometimes it just generates the non-physical time-*coordinate* flow of the system.

In the Hamiltonian system we do not consider lifted paths. A path in the Hamiltonian system is a map

$$\gamma \colon I \to \mathrm{Ph} \tag{8.2.22}$$

$$\gamma \colon t \mapsto (q(t), p(t)). \tag{8.2.23}$$

The paths in the Hamiltonian system corresponding to stationary paths of the associated Lagrangian system are given by solutions of Hamilton's equations. We can think of this in the following manner. The stationary paths of the associated Lagrangian system define vector fields on $T(TC)$ whose integral curves are those same stationary paths. Hamilton's equations define vector fields on $T(T^*C)$ whose integral curves are the Legendre push to T^*C of the stationary paths on TC. Equivalently we could make the correspondence directly on the level of vector fields. This makes no different as there is a one-to-one map between the flows defined by the corresponding vector fields and the vector fields.

Definition 8.2.7 — Hamilton's equations. Hamilton's equations are the equations given by

$$\frac{\mathrm{d}q(t)}{\mathrm{d}t} = \{H, q\} \tag{8.2.24}$$

$$\frac{\mathrm{d}p(t)}{\mathrm{d}t} = \{H, q\}\,. \tag{8.2.25}$$

In our definition these equations specify a vector field on T^*C, which means a section of $T(T^*C)$, given by

$$X_H = \frac{\partial H}{\partial p}\frac{\partial}{\partial q} - \frac{\partial H}{\partial q}\frac{\partial}{\partial p}. \tag{8.2.26}$$

We will use traditional notation, which hides these aspects, with the understanding that what we have detailed here is implied.

When defining the Hamiltonian above, we assumed that the equation for the conjugate momentum could be solved to find $v_i(p, q)$. This is not always the case. It depends on the Hessian which is defined by

$$\text{Hessian} = \det\left[\frac{\partial p_i}{\partial \dot{q^j}}\right] = \det\left[\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}\right]. \tag{8.2.27}$$

When the Hessian is zero the equation defining the conjugate momenta is singular, and cannot be solved for the velocity in terms of coordinates and momenta. We call the Lagrangian a singular Lagrangian when the Hessian is zero. A singular Lagrangian leads to constraint equations in Hamiltonian formalism. The conjugate momenta defined by using a singular Lagrangian are not all independent variables, and we get equations of the form

$$\phi_k(q_i, p_i) = 0. \tag{8.2.28}$$

This means that we can express some of the momenta as functions of other coordinates and momenta. If the dimension of the configuration space is m and the rank of the Hessian is r then there are r independent momenta. There are then $m - r$ dependent momenta. Let us order the momenta so that the independent momenta are given as p_A with $A \in \{1, \dots r\}$, and the dependent momenta are p_i with $i \in \{r + 1, \dots m\}$. Let a take values in the full range $\{1, \dots m\}$. Let us define the expression for the dependent momenta in terms of the other variables as

$$p_i = \pi_i(q^a, p_A). \tag{8.2.29}$$

We can then make an explicit constraint equation

$$\phi_k(q_i, p_i) \coloneqq p_i - \pi_i(q^a, p_A). \tag{8.2.30}$$

These equations are called primary constraints and represent conditions (or limitations) on the states that are allowed in the theory.

Definition 8.2.8 — Primary constraints. Primary constraints are constraints that originate directly from the definition of conjugate momenta without using the equations of motion.

Definition 8.2.9 — Secondary constraints. Secondary constraints are constraints that originate from the equations of motion. Specifically they are constraints that ensure that the time derivatives of all constraints are zero. This ensures that time evolution preserves all constraints.

We interpret constraints as being caused by describing the system by more variables than there are actual corresponding physical degrees of freedom. The phase space of the theory is in a certain sense to large. The actual dynamics of the system takes places in a subspace of phase space where the constraints are met. If there are K constraints, and they are all independent, the constraint surface will be a $2N - K$ dimensional submanifold (where N is the number of generalized coordinates). This subspace is called the (primary) constraint surface.

Before moving on let us interject with a short terminology note. We will continue to refer to $\mathcal{O}$ as observables in the case of constrained systems, even though the elements are no longer proper observables. The proper observables will be called Dirac observables. In the same manner we will still call the full unconstrained phase-space just the phase-space.

Definition 8.2.10 — Constraint surface. The surface defined by the equations

$$\phi_i = 0, \tag{8.2.31}$$

is called the constraint surface Γ_P. (With subscript P denoting primary constraints.)

Constraints in the finite-dimensional case are just a finite set of ordinary phase space functions, and they do not necessarily vanish everywhere, they only vanish (by definition) on the constraint surface. This presents some linguistic challenges. We need the concept of something being equal modulo constraints. We clarify this concept in the following definitions.

Definition 8.2.11 — Weakly vanishing. A function is said to be weakly vanishing when it is equal to zero (modulo constraints) on Γ_P. We write f is weakly vanishing as $f \approx 0$, and we define the meaning of this notation as

$$f \approx 0 \Leftrightarrow f|_{\Gamma_P} = \lambda_k \phi_k|_{\Gamma_P} . \tag{8.2.32}$$

Definition 8.2.12 — Weakly equivalent. Two functions f and g are said to be weakly equivalent when they are equal (modulo constraints) on Γ_P. We write this as $f \approx g$, and we define the meaning of this notation as

$$f \approx g \Leftrightarrow f|_{\Gamma_P} = g|_{\Gamma_P} . \tag{8.2.33}$$

Using these definitions we can say that $\phi_k \approx 0$ (and as we will see below, $H - H_T \approx 0$).

When a system has constraints, the usual prescription for writing down the Hamiltonian breaks down. Following the usual prescription and then writing down Hamilton's equations will not properly capture the dynamics since this will ignore the constraints. To proceed we first note that the Hamiltonian is weakly equivalent to the Hamiltonian plus any linear combination of the constraints. We write this as

$$H' \approx H + \sum \lambda_k \phi_k. \tag{8.2.34}$$

Using this modified Hamiltonian produces the modified equations of motions given by

$$\dot{q}_i \approx \frac{\partial H}{\partial p_i} + \dot{q}_\alpha \frac{\partial \phi_\alpha}{\partial p_i} \tag{8.2.35}$$

$$\dot{p}_i \approx -\frac{\partial H}{\partial q_i} - \dot{q}_\alpha \frac{\partial \phi_\alpha}{\partial q_i}. \tag{8.2.36}$$

The velocities $\dot{q}_\alpha$ are arbitrary functions, and we introduce (arbitrary) variables (called Lagrange multipliers) to correspond to these velocities. To make the formalism more compact, we use the above results and define a new Hamiltonian, called the total Hamiltonian. The total Hamiltonian includes a Lagrange multiplier multiplied by the relevant constraint for all singular velocities. We define the total Hamiltonian H_{total} as

$$H_{total} = H + \sum \lambda_k \phi_k. \tag{8.2.37}$$

Writing this expression in explicit form we get

$$H_{total} = \sum p_i \dot{q}_i(q_i, p_i) - \mathcal{L}(q_i, p_i) + \sum \lambda_k \phi_k(q_i, p_i). \tag{8.2.38}$$

Using the ordinary equations of motion together with the total Hamiltonian produces the correct result. For any function f we write

$$\dot{f} \approx \{f, H_T\} = \{f, H\} + \{f, \lambda_k \phi_k\} \tag{8.2.39}$$

The constraints themselves are also observables (i.e. they are functions), and the evolution equation must also apply to them. We write this as

$$\dot{\phi}_i \approx \{\phi_i, H_T\} = \{\phi_k, H\} + \{\phi_i, \lambda_k \phi_k\} . \tag{8.2.40}$$

The primary constraints we are dealing with are part of the definition of the Hamiltonian systems we are constructing. These constraints must therefore be true at all times, which means their time

derivatives must be zero. This means that from the above evolution equations we get the following restriction

$$\dot{\phi_k} \approx \{\phi_k, H\} + \{f, \lambda_k \phi_k\} \approx 0. \tag{8.2.41}$$

This equation can either lead to an equation just involving the λ^k's, or it can lead to new equations amongst the q_i's and p_i's. In the latter case these equation define secondary constraints Φ^j which must be met to ensure that $\dot{\phi_k} = 0$. The Φ^j's themselves must also have a zero time derivatives which could lead to another level of constraints. Any further levels of constraints beyond primary constraints, are all referred to as secondary constraints.

We have divided constraints into primary and secondary constraints. A second way of categorizing constraint are by the terms first and second class. We define these terms for any phase space functions as follows.

Definition 8.2.13 — First class function. A first class function f is a function that have weakly vanishing Poisson brackets with all constraints ϕ_i. We write

$$\forall i \, \{f, \phi_i\} \approx 0. \tag{8.2.42}$$

Definition 8.2.14 — Second class function. A second class functions is a function that does not have weakly vanishing Poisson brackets with all constraints. (I.e. second class functions are all functions that are not first class functions.) We write

$$\exists i \, \{f, \phi_i\} \not\approx 0. \tag{8.2.43}$$

This classification also applies to constraints (i.e. the constraint functions). First class constraints are therefore constraints that have weakly vanishing Poisson brackets with all constraints.

It may be the case that no velocities are invertible, and in this case we call the system totally constrained. The Hamiltonian is then just a sum of Lagrange multipliers multiplying the corresponding constraints. The constraints ϕ_i represents constraint equations of the form $\phi_i = 0$. The Hamiltonian of a totally constrained system is weakly vanishing. This lead to a situation where the derivative with respect to the Hamiltonian flow parameter of any observable (function) is zero. This situation is called "the problem of time" and it occurs in any Hamiltonian system that has time reparametrization invariance. We will return to how to deal with this situation in section **??**.

We can relate first class constraints to gauge transformations. To see this note that the constraints restrict the phase-space to the constraint surface, but like for any other $f \in \mathcal{O}$ they also generate flows on this surface. We repeat the the general evolution equation for an observable f, which is

$$\dot{f} \approx \{f, H_T\} = \{f, H\} + \{f, \lambda_k \phi_k\}. \tag{8.2.44}$$

These flow generated by the constraints are found in the second part of the Hamilton's equation for constrained systems. First note that λ_k is an arbitrary function. Second note that

$$\{f, \lambda_k \phi_k\} = \lambda_k \{f, \phi_k\} + \{f, \lambda_k\} \phi_k \approx \lambda_k \{f, \phi_k\}, \tag{8.2.45}$$

since the last term in the second expression is proportional to a constraint. This means that

$$\dot{f} \approx \{f, H_T\} = \{f, H\} + \lambda_k \{f, \phi_k\}. \tag{8.2.46}$$

Recall now that λ_k are arbitrary functions which means that any evolution with regards to a given Hamiltonian can be altered by an arbitrary multiple of $\{f, \phi_k\}$. This means that any Hamiltonian evolution is only determined up to the flows generated by first class constraints ϕ_k. These flows

are the equivalent of gauge transformations. The orbits of these flows are a set of mathematically equivalent points with regards to flows generated by the total Hamiltonian. Thus the proper physical phase space is the quotient space Γ_{red} of the constraint surface Γ under the equivalence relation defined by the flows generated by the first class constraints (see figure 8.4).

Since first class constraints generate flows that represent gauge transformation this means that proper observables must be independent of such flows. Proper observables must thus be functions of quotient space Γ_{red}. In other words proper observables must commute with all the first class constraints. Observables that weakly Poisson commute with all the first class constraints are called Dirac observables. For fully constrained systems this implies that Dirac observable also weakly Poisson commute with the Hamiltonian. Dirac observables are identical with what we have previously defined as first class functions.

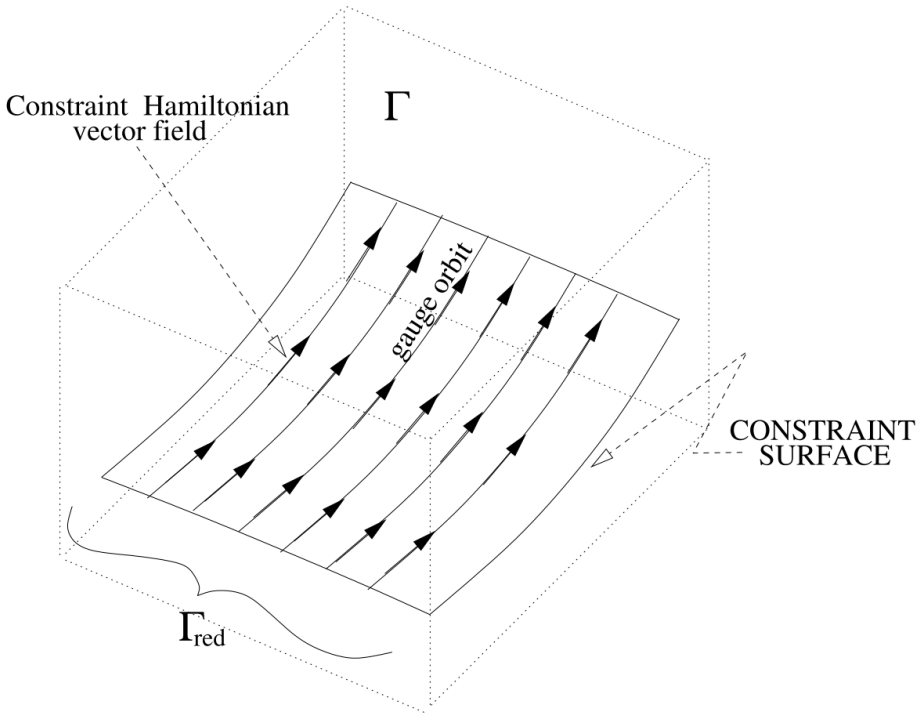


Figure 8.4: The constraint surface Γ defined by setting the constraints functions equal to zero. On the constraint surface the various constraint functions define gauge orbits. The physical phase space is the quotient Γ_{red} of the constraint surface with respect to the gauge orbits. Illustration from [149].

Hamiltonian field theory

We will not repeat the full analysis for the field theory case. We limit ourselves to a few remarks. The phase-space, the algebra of observables and the Poisson brackets are crucial elements in the construction of LQG. It is useful to have clear definitions of these structures. We define the phase space of a Hamiltonian scalar field theory defined on spatial manifold M to be

$$\mathrm{Ph} = C^\infty(M, \mathbb{R}) \times C^\infty(M, \mathbb{R}). \tag{8.2.47}$$

(Again we are assuming we can work on a bundle trivialization.) The algebra of observables is defined as

$$\mathcal{O} = C^\infty(\mathrm{Ph}, \mathbb{R}). \tag{8.2.48}$$

Constraints in the field theory case are infinite families of functions $\phi_x \in \mathcal{O}$, parameterized by $x \in M$, that we require to be 0. Such constraints again define a submanifold Γ_F of Ph. Once again the Poisson brackets adds another binary operation to algebra $\mathcal{O}$ by defining

$$\{\,,\} : \mathcal{O} \times \mathcal{O} \to \mathcal{O}. \tag{8.2.49}$$

This is in principle the same as in the finite-dimensional case except that operations on infinite dimensional spaces are more involved to define. (Which of course means that they are not really the same at all.)

Poisson algebra structure

Since the Poisson algebraic structure of $\mathcal{O}$ is the key structure when building LQG we want to emphasize a couple of aspects regarding this. We shall need the concept of elementary functions.

Definition 8.2.15 — Elementary functions. Given a phase-space Ph and a set of observables $\mathcal{O}$. A subset $S \subset \mathcal{O}$ is called a set of elementary functions if

1. The functions in S separate the elements of Ph
2. All elements of $\mathcal{O}$ can be expressed by elements of S.
3. The set S is minimal.

Referring to a given such set S we will denote any given member functions as an elementary function. (We note this since a function being an elementary function in itself, without referring to the set S, is not strictly a meaningful concept.)

Our goal is to illustrate some points with regards to the Poisson bracket in Hamiltonian field theory, but these aspects are best understood in analogy with finite dimensional case. We therefore once more return to the finite-dimensional case. In the finite-dimensional case we have a finite set of elementary functions $X_i, P_i \in \mathcal{O}$ where

$$X_i \colon \mathrm{Ph} \to \mathbb{R} \tag{8.2.50}$$

$$P_i : \mathrm{Ph} \to \mathbb{R}. \tag{8.2.51}$$

The maps X_i, P_i can be considered to be projections that are defined by

$$X_i \colon (x_1, \ldots, x_m, p_1, \ldots, p_m) \mapsto x_i \tag{8.2.52}$$

$$P_i : (x_1, \ldots, x_m, p_1, \ldots, p_m) \mapsto p_i. \tag{8.2.53}$$

In ordinary physics notation we use the same symbol x_i for both $x_i \in M$ and $x_i \in \mathcal{O}$. This is conceptually quite unfortunate. We define the Poisson bracket between two generic functions as

$$\{,\} : \mathcal{O} \times \mathcal{O} \to \mathcal{O} \tag{8.2.54}$$

$$\{,\} : (f, g) \mapsto \sum_{i=1}^{m} \frac{\partial f}{\partial X_i}\frac{\partial g}{\partial P_i} - \frac{\partial g}{\partial X_i}\frac{\partial f}{\partial P_i}. \tag{8.2.55}$$

We calculate the Poisson brackets for the fundamental functions X_i, P_i as

$$\{X_i, P_j\} = \sum_{k=1}^{m} \frac{\partial X_i}{\partial X_k}\frac{\partial P_j}{\partial P_k} - \frac{\partial P_j}{\partial X_k}\frac{\partial X_i}{\partial P_k} = \sum_{k=1}^{m} \delta_{ik}\delta_{jk} - 0 \cdot 0 = \delta_{ij} \tag{8.2.56}$$

$$\{X_i, X_j\} = \sum_{k=1}^{m} \frac{\partial X_i}{\partial X_k}\frac{\partial X_j}{\partial P_k} - \frac{\partial X_j}{\partial X_k}\frac{\partial X_i}{\partial P_k} = \sum_{k=1}^{m} \delta_{ik} \cdot 0 - \delta_{jk} \cdot 0 = 0 \tag{8.2.57}$$

$$\{P_i, P_j\} = \sum_{k=1}^{m} \frac{\partial P_i}{\partial X_k}\frac{\partial P_j}{\partial P_k} - \frac{\partial P_j}{\partial X_k}\frac{\partial P_i}{\partial P_k} = \sum_{k=1}^{m} 0 \cdot \delta_{jk} - 0 \cdot \delta_{ik} = 0. \tag{8.2.58}$$

We now return to field theory. In the infinite-dimensional case (with only a single field) we have infinite families of fundamental functions parameterized by $x \in M$. For a scalar field theory we have

$$\Theta_x \colon \mathrm{Ph} \to \mathbb{R} \tag{8.2.59}$$
$$\Pi_x \colon \mathrm{Ph} \to \mathbb{R}. \tag{8.2.60}$$

These functions are seldom discussed or explicitly defined. These functions assign a real number to every function $C^\infty(M \to \mathbb{R})$. Unlike in the finite-dimensional case there is not really a well established notation for signifying that a function is to be considered as a variable. Although the idea is exactly the same as before, there is no equivalent notation denoting the same conceptual separation between x as a variable and a as a constant. We temporarily use notation $[f]$ to denote that the function f is intended to serve as a function variable. We define explicitly the families of fundamental functions as

$$\Theta_x \colon ([\phi], [\pi]) \mapsto \phi(x) \tag{8.2.61}$$
$$\Pi_x \colon ([\phi], [\pi]) \mapsto \pi(x). \tag{8.2.62}$$

This notation means that Θ_x assigns to any function $\phi \in \mathcal{O}$ the value $\phi(x) \in \mathbb{R}$. In standard physics notation $\phi(x)$ is used for both the parameterized family of functions $\{\Phi_x\} \subset \mathcal{O}$, and for the image of the function ϕ applied to x, which gives the number $\phi(x) \in \mathbb{R}$. This standard physics notation is somewhat confusing. The Poisson brackets are defined as

$$\{F, G\} = \int \mathrm{d}^3x \, \frac{\delta F}{\delta \Theta_x} \frac{\delta G}{\delta \Pi_x} - \frac{\delta G}{\delta \Theta_x} \frac{\delta F}{\delta \Pi_x}. \tag{8.2.63}$$

Note that we are integrating over the family of elementary functions (Θ_x, Π_x), not over the domain of the involved functions (F, G). The brackets for the elementary functions are

$$\{\Theta_x, \Pi_y\} = \int \mathrm{d}^3z \, \frac{\delta \Theta_x}{\delta \Theta_z} \frac{\delta \Pi_y}{\delta \Pi_z} - \frac{\delta \Pi_y}{\delta \Theta_z} \frac{\delta \Theta_x}{\delta \Pi_z} = \delta(x, y) \tag{8.2.64}$$
$$\{\Theta_x, \Theta_y\} = \int \mathrm{d}^3z \, \frac{\delta \Theta_x}{\delta \Theta_z} \frac{\delta \Theta_y}{\delta \Pi_z} - \frac{\delta \Theta_y}{\delta \Theta_z} \frac{\delta \Theta_x}{\delta \Pi_z} = 0 \tag{8.2.65}$$
$$\{\Pi_x, \Pi_y\} = \int \mathrm{d}^3z \, \frac{\delta \Pi_x}{\delta \Theta_z} \frac{\delta \Pi_y}{\delta \Pi_z} - \frac{\delta \Pi_y}{\delta \Theta_z} \frac{\delta \Pi_x}{\delta \Pi_z} = 0. \tag{8.2.66}$$

We see that the right hand side of equation **??** will be problematic. The Dirac delta $\delta(x, y)$ is not a function but a distribution. Therefore $\delta(x, y) \notin \mathcal{O}$ which is in contradiction the requirements stated in the definition of the Poisson bracket. When we get to this aspect in the development of LQG we will deal with this by smearing, which means integrating over subfamilies of elementary functions.

We observe one final aspect of describing dynamics in the Hamiltonian framework. In chapter 2 we used a one-parameter group of algebra automorphism to describe dynamics. To connect this to the present case we note that the algebra in question is the algebra of observables, and the automorphisms are those determined by the flows determined by the equations

$$F_H = \{F, H\}. \tag{8.2.67}$$

The Poisson bracket is a derivation (i.e. a tangent vector) on the algebra and it brings along a flow as we have previously discussed. The is a slight subtlety in all of this. In our previous discussion of dynamic evolution (see definition 8.2.7) we took the tangent vectors to be tangents on Ph and not on $\mathcal{O}$. We are now placing the tangent vectors on $\mathcal{O}$. The explanation for this change is as follows. We are always talking about elementary functions, or observables that can be expressed in terms

of them. An elementary function has a double interpretation as a member of $\mathcal{O}$ as well as being an element of Ph. E.g. the function X_i that we defined above, and the object x_i are basically two alternative descriptions of the same entity. This means that flows on Ph also define flows on $\mathcal{O}$ and vice versa. We can easily go back and forth between these pictures, and we will do so without further notice. These two picture of dynamics correspond to the Schrodinger and the Heisenberg picture when these two pictures are understood in the appropriate generalized sense that we have outlined here. Only the combination of observables and states can have a real world interpretation, and therefore only the combination of the two have any dynamical meaning. It is therefore to be expected that dynamics can be represented as changes in either one of these components.

8.3 Quantization of theories with constraints

In the previous section we have established the basic structure of classical theories with constraints. We now turn to the question of quantization. There is in general no unique and well defined path from classical theories to quantum theories, but for systems with a single degree of freedom, with some additional restrictions, the correspondence is unique. For other systems, including field theories, there exists a canonical formal procedure that forms an important reference for discussing and understanding the subject. We need to have an understanding of this procedure that is closer to being rigorous mathematics than what is usually presented.

Our goal in this section will be to spell out in some detail a more explicit and well defined framework for quantization. This will give us a deeper understanding of quantum theories and the specific relation to their classical starting point. We proceed along the lines of chapter 2 where we built a common algebraic understanding of both classical and quantum system. It is therefore natural and necessary that our understanding of quantization will be in the form of a map between such algebraic system. Specifically, we will define quantization to be a Lie algebra homomorphism between algebras (see section 8.13 – in the subsection GNS construction for some further clarifications). The relevant algebras will be subalgebras of the algebra of observables for the classical system and the quantum system.

To fix our terminology and clarify our perspective we review quantization in the more well known cases of QM and QFT not involving constraints, before proceeding to the more complex cases involving constrained systems.

Quantization in QM

In canonical quantization we quantize a classical Hamiltonian theory. The essential structure of the classical theory is expressed by giving its Poisson algebra. Quantizing this system means finding a Lie algebra representation of a suitable subalgebra $\mathfrak{A}$ of the Lie algebra part of the classical Poisson algebra $\mathcal{O}$ [12][150]. The representation will be in the form of an operator algebra over a complex Hilbert space $\mathcal{H}$. The Lie algebra part of the classical Poisson algebra is mapped by a homomorphism denoted $\mathcal{Q}$, to a Lie algebra defined as follows. Let $L(\mathcal{H})$ be the complex algebra of bounded linear operators on $\mathcal{H}$. Define the quantum Poisson bracket $\{\ ,\ \}_Q$ on $L(\mathcal{H})$ by [1]

$$\left\{\hat{f},\hat{g}\right\}_Q = i\hbar(\hat{f}\hat{g}-\hat{g}\hat{f}). \tag{8.3.1}$$

[1] In this formula we have temporarily reinstated $\hbar$ for easier reading. We have done the same ting in figure 8.5. Outside of these formulas we set $\hbar = 1$. Note that the presence of $\hbar$ defines a one-parameter family of quantization maps $Q_\hbar$.

This makes $L(\mathcal{H})$ a Lie algebra. We require that the quantization map $\mathcal{Q}\colon \mathfrak{A} \to L(\mathcal{H})$ must be a Lie algebra homomorphism, which means that

$$\mathcal{Q}\colon \alpha f + \beta g \mapsto \alpha\mathcal{Q}(f) + \beta\mathcal{Q}(g) \tag{8.3.2}$$

$$\mathcal{Q} : (\{f, g\}) \mapsto \{\mathcal{Q}(f), \mathcal{Q}(g)\}_Q \tag{8.3.3}$$

$$(\mathcal{Q}\colon \mathbb{I} \mapsto \hat{\mathbb{I}}). \tag{8.3.4}$$

The last condition only applies for unital Lie algebras. Note that for this homomorphism to make sense we must be dealing with either two real Lie algebras or two complex Lie algebras. If we restrict $L(\mathcal{H})$ to be the subalgebra of symmetric linear operators we can treat this as a real Lie algebra. Otherwise we have to either define the classical algebra of observable to be the algebra of complex functions over the complex numbers, or create the complexification of the algebra of real functions over the real numbers. For the complex case we have to add the condition that

$$\mathcal{Q}\colon f^* \mapsto \mathcal{Q}(f)^*. \tag{8.3.5}$$

In general one uses the Gelfand-Naimark-Segal (GNS) construction to find and catalog representations [13][151]. The GNS construction takes a C*-algebra $\mathcal{A}$ and a single state ω as input, and from this it generates a representation on a Hilbert space made from $\mathcal{A}$ (see section 8.13 - subsection GNS construction for the details of the proper use of the GNS construction). The GNS construction also defines the required inner product on this vector space to make it a Hilbert space. Because of the existence of a map between states and representations provided by the GNS constructions, representations can be studied by classifying states. Note that the GNS construction gives representation of a C*-algebra. This means that we cannot apply it to the classical Poisson sub-algebra $\mathfrak{A}$ since we only want to preserve the Lie algebra part of the Poisson algebra. Instead we must first find a Lie algebra (iso)morphisms to an abstract noncommutative algebra $\mathfrak{B}$ that implements the Lie algebra multiplication as a commutator. Then we can apply the GNS construction to the algebra $\mathfrak{B}$ to find its representations (see [25] for a discussion of this in the context of LQG). We refer the details of the GNS construction to the appendices (see appendixH) and instead we give an example and some caveats.

Let C be the configuration space for a system with phase space Ph, then the Hilbert space will be $\mathcal{H} = L^2(C, dx)$. Note that these are functions only on the configuration space C, while the algebra of observables are functions on the phase space Ph. Specifically, for a one-dimensional system the Hilbert space is given as $\mathcal{H} = L^2(\mathbb{R}, dx)$, which is a complete infinite dimensional inner-product space with a countable orthonormal basis.

We have stated that the Lie algebra we are representing is a subalgebra of the full algebra of observables. Let us exemplify this. For the simplest case of a one-dimensional system described classically by $(q, p) \in \mathbb{R}^2$, the full Poisson algebra is given by $\mathcal{O} = C^\infty(\mathbb{R}^2, \mathbb{R})$. To performs the quantization we must find some sub-algebra $\mathfrak{A} \subset \mathcal{O}$. The most common choice is the space defined as the linear span of the set $\{1, q, p\}$. The elements of this space are all order one polynomials in q and p. The sub-algebra can be extended to second order polynomials without breaking the Lie algebra homomorphism requirements, but it cannot be extended beyond that. This result is stated in the Groenewold-van Howe theorem [150][14].

Unless we restrict M to be compact, q and p will be unbounded functions. Such functions map to unbounded operators under $\mathcal{Q}$. Unbounded operators are not defined on all of $\mathcal{H}$ and therefore each have their own domain D with $D \subsetneq \mathcal{H}$. There might exist some dense set on which all relevant such operators are defined, but in general this is not true. Either we must assume and use such a common domain to define the quantum algebra, or we must get rid of the unbounded operators. The Weyl method maps the unbounded functions to bounded functions, and quantizes this modified algebra of functions instead [12]. This avoids any domain issues. We will implicitly assume that

such Weyl countermeasures have been implemented, and that when we are referring to the classical algebra in the context of the map $\mathcal{Q}$ we are actually referring to the Weyl modified classical algebra.

A very important question regarding the representation is to which degree it is unique. Properly different representation would represent distinct and different quantum theories, and we would have to indicate which "QM" we were talking about in any specific calculation. For the countable-dimensional case, the Stone-von Neumann theorem specifies that the representation is unique, given the assumptions of being unitary, irreducible and weakly continuous [147].

Quantization in field theory

We quickly repeat the analogous steps for a field theory. Again the Poisson algebra $\mathcal{O}$ will represent the essential structure of the classical theory. From $\mathcal{O}$ we must select a suitable subalgebra $\mathfrak{A}$. This subalgebra must them be transformed to the abstract quantum algebra $\mathfrak{B}$. We will the use the GNS construction to find a representation of $\mathfrak{B}$ (see section 8.13 - subsection GNS construction for the details of the proper use of the GNS construction). When quantizing this theory in the Hamiltonian framework, we need a Hilbert space that consist of functions on the space of functions from M to $\mathbb{R}$. This is the space of square integrable functions on the configuration space $C = C^\infty(M, \mathbb{R})$. We write this as

$$\mathcal{H} = L^2(C, d\mu). \tag{8.3.6}$$

(Note that it is not obvious what the measure $\mathrm{d}\mu$ should be.) The functions in $L^2(C, d\mu)$ assign a value to each element (field configuration) in the configuration space. The value of a function for a given field configuration will eventually come to represent the probability amplitude of that field configuration.

In the infinite dimensional case there are usually many inequivalent representations, and we have to add more criteria to determine which representation to use. This could be a criteria like respecting a certain symmetry or preserving some other relevant aspect.

Refined algebraic quantization

Having established some of the ground rules for quantization we now adapt our approach so as to be applicable to constrained theories. This adaption is called refined algebraic quantization (RAQ) or Dirac quantization. The basic question is how to handle the constraints. One could imagine implementing the constraints on the classical phase space before quantization, and then try to identify a proper set of (Dirac) observables. This approach could potentially be very difficult. The difficuly is not just that the constraints are difficult to solve, but also because the resulting algebra of (Dirac) observables can be difficult to represent [25]. In Dirac quantization the constraints are ignored when performing the first step of quantization, and then implemented in the quantum theory after quantization. For LQG we will follow this procedure. Except for adding techniques to handle the constraints, the method is basically the same as the canonical quantization we have described above. Nevertheless, we will write down the procedure for RAQ as a four step procedure, and give some details about each step (see [25] for further details).

Definition 8.3.1 — RAQ / Dirac quantization. For a given classical system, containing only first class constraints and being represented by a classical *-Poisson algebra, the Dirac quantization procedure consist of the following four steps:

1. Identify a suitable classical *-Poisson subalgebra $\mathfrak{A}$.
2. Define an abstract *-algebra $\mathfrak{B}$.
3. Find a representation of $\mathfrak{B}$ as operators on a Hilbert space.
4. Implement the constraints as operators on the Hilbert space.

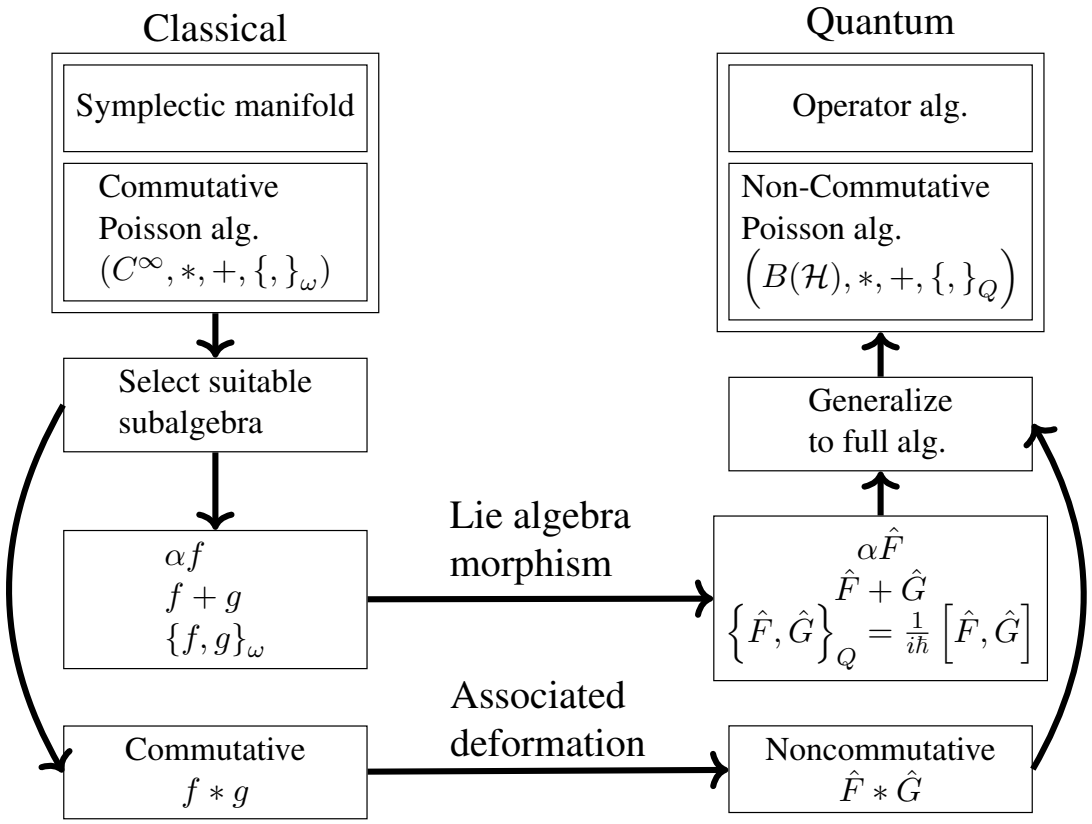


Figure 8.5: The diagram shows a simplified view of the process of quantization. Note that compared to definition 8.3.1 we are skipping some of the intermediate steps (see section 8.13 - subsection GNS construction for a clarification). On the left hand side are the classical theory that forms the starting point. On the right hand side are the quantum theory that is the result of quantization. The two lowermost boxes on both sides represent the four algebraic structures found in the respective Poisson algebras. The Lie algebra part of the classical Poisson algebra is mapped by a Lie algebra morphism to the quantum Lie algebra. Because of this morphism there is an accompanying deformation of the commutative product of the classical Poisson algebra to a non-commutative product in the quantum Poisson algebra. We recover the ordinary definition of commutator brackets from the definition of the quantum Lie algebra product (quantum Poisson bracket). Notice that the active quantization map in this approach only involves a subalgebra of the Lie algebra part of the classical Poisson algebra. Other setups for quantization puts the deformation part in the central position, but we will not use such techniques in this text.

Step 1

We assume we are given a classical Hamiltonian theory with a set of first class constraints. Step one highlights several things to be aware of for a successful quantization. The first thing to notice is that we cannot hope to quantize the full algebra of functions. The Groenewold-van Hove theorem tells us that such a quantization is not possible. If we let M be the phase space, we must select a proper subalgebra $\mathfrak{A}$ of the full algebra of observables $\mathcal{O}$. Note that quantization of different subalgebras in principle can give different quantum theories. The second point is that the algebra $\mathfrak{A}$ must be suitable. By suitable we mean that the Poisson bracket is the canonical one, or something equally simple. We also mean that the elementary functions are such that they transform neatly under the transformations generated by the constraints. Finally, we need to make sure that the elementary functions are such that they can be represented by operators with a clearly defined (and preferably simple) action on Hilbert space states.

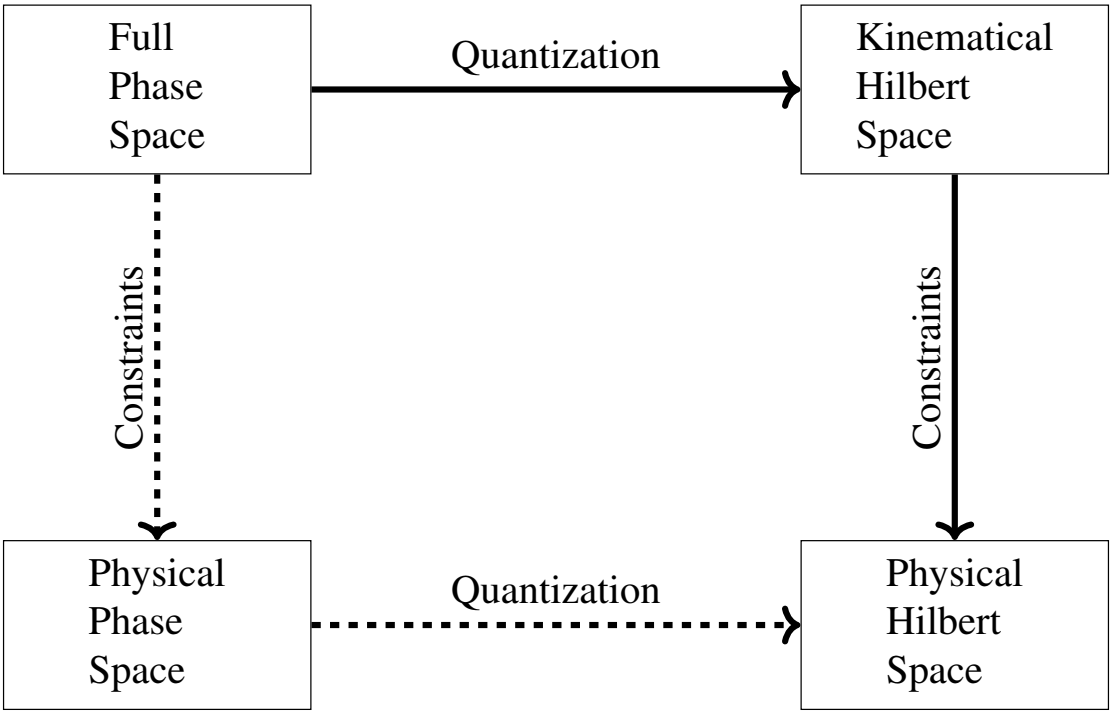


Figure 8.6: The diagram shows the possible orderings of constraint implementation and quantization. Dirac quantization means following the full drawn lines. It can not be assumed that the diagram is commutative. Quantization by the non-Dirac path may in many case be very difficult to accomplish.

Step 2

In the second step we want to transform the original Lie algebra $\mathcal{A}$ to its enveloping algebra. We will not present the details of this construction (see [25]). Note that to avoid domain problems with unbounded operators, one usually consider the algebra modified by Weyl procedure rather than the original classical algebra $\mathcal{A}$ (see [25]). In the Weyl algebra one uses bounded functions of the original variables instead of the unbounded originals.

Step 3

In step 2 we construct a representation of the algebra $\mathfrak{B}$. (We assume as before that unbounded functions have been dealt with in a Weyl procedure.) A representation is a *-morphism $\pi : \mathfrak{B} \to \mathcal{L}(\mathcal{H})$. In finite-dimensional QM this representation is unique by the Stone-von Neumann theorem. In the infinite-dimensional case there are many inequivalent representations. Additional requirements such as enforcing certain symmetries and supporting simple representation of relevant constraints can lead to a unique representation. The GNS construction is the major tool to find and study such representations (see section 8.13 - subsection GNS construction for the details of the proper use of the GNS construction).

Step 4

The quantization procedure we are using quantizes the system without regards to the constraints. These constraints must now be implemented on the Hilbert space. The constraint equations $\phi(q,p) = 0$ will now become equations $\hat{\phi}\,|\psi\rangle = 0$. It is not strictly required to implement the constraints by solving this operator equation. One important alternative technique is to implement constraints by group averaging (see section 8.7). We cannot assume that the new space defined by implementing constraints always can be implemented by subsets of the original Hilbert space. Sometimes a modified space must be defined (see [152] for examples). The constraint procedure can also require defining a new inner product.

8.4 The Ashtekar formulation of classical GR

In 1986 Ashtekar gave a new formulation of GR which marks the start of LQG [153][154]. The unique insight behind the Ashtekar formulation was the construction of a Hamiltonian version of GR which is very similar to Yang-Mills theories. This made available more of the tools used in quantization of Yang-Mills theories, and started the path towards loop quantization of GR [155] [156].

There are two approaches that are used to derive the gauge like Hamiltonian system from the Einstein-Hilbert Lagrangian. The first approach performs a Legendre transformation on the Einstein-Hilbert action, and then does several canonical transformations on these variable to arrive at the desired endpoint [157] [25]. The other approach starts with the Palatini-Holst action, which is classically equivalent to the EH-action, and performs a Legendre transformation to arrive directly at the same Ashtekar variables as in the first approach [77], [158]. In the following we use $\kappa = 8\pi G, c = 1, \hbar = 1$ and the four-dimensional metric is $(-,+,+,+)$. See table 8.1 for the rest of our notational conventions.

Approach 1: Hamiltonian formulation of the Einstein-Hilbert action

The Hamiltonian formulation splits spacetime provisionally into space and time. This split is seemingly at odds with the focus on background independence. It turns out however, that the split does not break background independence, but it does make it non-manifest [159]. This is a drawback with the Hamiltonian approach but using canonical methods for quantization have so far been more successful than Lagrangian path-integral techniques. All-in-all the benefits so far go in favor of the Hamiltonian approach.

Step 1: ADM action

The first step is to foliate spacetime into a one-parameter family of three-dimensional leafs [160]. To ensure that this foliation is possible we need the following theorem [161].

Theorem 8.4.1 — Geroch's theorem. If a spacetime manifold M is globally hyperbolic it is diffeomorphic to $\mathbb{R} \times \sigma$, where σ is a three-dimensional manifold.

Proof. See [161] for a proof. ■

Given a diffeomorphism $X : \mathbb{R} \times \sigma \to M$ this implies that there is a family of embeddings

$$X_t : \sigma \to M \quad \text{with} \quad X_t = X(t, \,). \tag{8.4.1}$$

The image of σ by the map X_t is the hypersurface $\Sigma_t = X_t(\sigma)$ in M. We refer to X as a foliation of M. We refer to Σ_t as a slice or a leaf (of the foliation). We denote the tangent bundle for a leaf as $T\Sigma_t$, and the the tangent space at $p \in \Sigma_t$ as $T_p\Sigma_t$.

If we fix a specific diffeomorphism $X : \mathbb{R} \times \sigma \to M$, this breaks diffeomorphism invariance. However, if we operate with the assumption of permitting an arbitrary diffeomorphism $X : \mathbb{R} \times \sigma \to M$, then diffeomorphism invariance is preserved.

Lemma 8.4.2 — Foliations and diffeomorphisms. Let $X : \mathbb{R} \times \sigma \to M$ and $Y : \mathbb{R} \times \sigma \to M$ be two foliations. Then $X \circ Y^{-1}$ and $Y \circ X^{-1}$ are diffeomorphism and there exists a diffeomorphism such that $Y = \phi \circ X$.

Proof. For part one, we see that the map $X^{-1} \circ Y$ is a map $M \to M$ composed of diffeomorphism, hence it is a diffeomorphism. The case $Y^{-1} \circ X$ follows directly from this. For part two, let $\phi = Y \circ X^{-1}$ then $\phi \circ X = (Y \circ X^{-1}) \circ X = Y \circ (X^{-1} \circ X) = Y$ since composition is associative and all involved maps are bijections. ■

Item	**Symbol**	**Comment**
Manifold	M, σ, Σ_t	4d, 3d, 3d-embedding
Foliation	X	$X\colon \mathbb{R}\times\sigma\to M$ $X\colon (t,x^a)\mapsto x^\mu$
Fol. param.	n^μ, N, N^a	normal vector, lapse, shift vector
Coord. indices	$\mu,\nu,\ldots,a,b,\ldots$	4d, 3d
Orth. indices	$I,J,\ldots,i,j,\ldots$	4d-tetrad, 3d-triad
Orth. bases	e^I_μ, e^a_i, E^a_i	4d-tetrad, 3d-triad, 3d-densitized
Tensors	$S^{\mu\nu}, T^{\mu\nu}, T^{ab}, T^{ij}, T^{ai}$	4d-cord,4d-spat., 3d-coord,3d-tri., 3d-mix.
Metrics	$g_{\mu\nu}, g_{IJ}, q_{\mu\nu}, q_{ab}, \delta_{ij}$	4d, 4d-tetr., 4d-spat., 3d-coord, 3d-tri.
Metric det.	g, q	4d,3d
Cov. deriv.	∇_μ, D_μ, D_a	4d, 4d-spat., 3d-coord.
Conn. comp.	$\Gamma^\mu_{\nu\sigma}, \Gamma^a_{bc}, \Gamma^a_{ic}, A^i_a$	4d-coord., 3d-coord., 3d-spin, 3d-Asht.
Ext. curv.	$K_{\mu\nu}, K_{ab}, K^i_a$	4d-spat, 3d-coord., 3d-mix.
Conj. momenta	P^{ab}, Π, Π^a	of metric, of lapse, of shift
Lagrange multipl.	λ, λ^a	For momenta Π, Π^a
Space of conn.	$\mathcal{A}, \bar{\mathcal{A}}$	smooth conn., generalized conn.
Conn.	A	$A\in\mathcal{A}$
Smearing	S, e, γ	surface, edge, graph
Hol.	$h_e, h_e[A], h_e[\mathcal{A}]_{\alpha\beta}, h^{(j)}[c,A]$	e = edge
Hol. op.	$\hat{h}_e, \hat{h}_e[A], \hat{h}_e[A]_{\alpha\beta}, \hat{h}^{(j)}[c,A]$	e = edge
Flux	$F^{fa}_S[A]$	S = surface
Flux op.	$\hat{F}^{fa}_S[A]$	S = surface
Intertw.	ι	
Rep. matrix	$D^{(j_1)}_{mn}(g)$	$g\in SU(2)$
Constraints	$\mathcal{H}, \mathcal{H}_a, \mathcal{G}_i$	Hamiltonian, diffeomorphism, Gauss
Lagrange multipl.	N, N^a, Λ^i	Hamiltonian, diffeomorphism, Gauss
Immirzi param.	β	
Spc. of cyl. func.	$Cyl_\gamma, Cyl, \overline{Cyl}$	γ = graph, generalized, completion
Paths	$\mathcal{P}$	Spc. of eq.cl. of paths (see text)
Proj. fam.	$X_\infty, \overline{X}, X_l$	Proj. family (see text)
Groupoids	$\mathcal{L}, l(\gamma)$	Groupoids (see text)

Table 8.1: LQG notation. Abbreviations: orth. = orthonormal,coord. = coordinate, cov.=covariant, deriv. = derivative, multipl. = multiplier, intertw. = intertwiner, det. = determinant, fol. = foliation, hol. = holonomy, conn. = connection, conj. = conjugated, op. = operator, param. = parameter, ext. = extrinsic, curv. = curvature, spc. = space, comp. = components, cyl. = cylindrical, func. = functions, tetr. = tetrad, tri. = triad, spat. = spatial, mix.= mixed, Asht. = Ashtekar w. = with, rep. = representation, diff. = diffeomorphism

Theorem 8.4.3 — Foliation invariance. Let S_{3+1} be a (3+1)-foliated action such that

$$S_{3+1}\left[X^*g\right] = S_{EH}\left[g\right] \tag{8.4.2}$$

for some foliation X. Then for any other foliation Y we have

$$S_{3+1}\left[Y^*(g)\right] = S_{EH}\left[g\right]. \tag{8.4.3}$$

Proof. Let $\phi\colon M \to M$ be a diffeomorphism. That EH is a Diff-invariant action means that $S_{EH}\left[g\right] = S_{EH}\left[\phi^*g\right]$. Since we postulate that $S_{3+1}\left[X^*g\right] = S_{EH}\left[g\right]$ we find that $S_{3+1}\left[X^*(\phi^*g)\right] = S_{EH}\left[\phi^*g\right] = S_{EH}\left[g\right]$ But $X^*(\phi^*g) = (\phi \circ X)^*(g)$ and $Y = \phi \circ X$ is just a different foliation. For any two foliation X and Y there is a diffeomorphism such that $Y = \phi \circ X$ by lemma 8.4.2. Thus we have shown that $S_{3+1}\left[Y^*(g)\right] = S_{3+1}\left[X^*g\right] = S_{EH}\left[g\right]$. ■

Note however that the actual foliations used for ADM will be required to have leaves that are spatial hypersurfaces. This means that the foliations are not completely arbitrary and demonstrating that this also respects diffeomorphism invariance is beyond the scope of this text.

For each slice Σ_t we define a unit length normal vectorfield n which is orthogonal to the slice.

Definition 8.4.4 — Unit normal of a folliation. The unit normal vector field n of a foliation Σ_t of the manifold M with metric tensor g is defined by

$$g(n,n) = -1 \tag{8.4.4}$$

$$(\forall v \in T_p\Sigma_t)\, g\left(n(p), v\right) = 0. \tag{8.4.5}$$

The last equation should be understood to hold for each point $p \in \Sigma_t$ and for each leaf Σ_t.

The idea we wish to pursue is to express any relevant four-dimensional geometric expression in terms of entities which are defined in terms of objects on the individual three-dimensional slices [25][162]. This will enable us to re-express the four dimensional EH-action as a 3+1 action. We then perform a Legendre transformation on the 3+1 action, and proceed to find the correct Hamiltonian by using techniques adapted to constrained systems.

We start by realizing that any tangent spaces of T_pM of M can be decomposed into two subspaces as

$$T_pM = N_p\Sigma_t \oplus T_p\Sigma_t. \tag{8.4.6}$$

The space $N_p\Sigma_t$ is the normal space at p. The vector T^μ is defined as the tangent vector of the coordinate lines of t (defined by the foliation X). This vector can be decomposed into a normal component and a tangent component as

$$T^\mu = \frac{\partial X^\mu}{\partial t} = N(x)n^\mu(x) + N^\mu(x). \tag{8.4.7}$$

Here $N^\mu(x)$ is a vector field (called the shift vector) orthogonal to n^μ. The field $N^\mu(x)$ is defined by

$$N^\mu(x) \coloneqq T^\mu - N(x)n^\mu(x). \tag{8.4.8}$$

The scalar field $N(x)$ (called the lapse) is the component of T^μ along the unit normal. We can write this as

$$N(x) = -T \cdot n \tag{8.4.9}$$

This induces a decomposition of any four-dimensional tensor. For the four-dimensional metric this decomposition is given by

$$g_{\mu\nu} = q_{\mu\nu} - n_\mu n_\nu. \tag{8.4.10}$$

Note that $q_{\mu\nu}$ is a map

$$q_{\mu\nu} \colon (N_p\Sigma_t \oplus T_p\Sigma_t) \times (N_p\Sigma_t \oplus T_p\Sigma_t) \to \mathbb{R}, \tag{8.4.11}$$

and that

$$q_{\mu\nu} \colon (N_p\Sigma_t \oplus T_p\Sigma_t) \times (N_p\Sigma_t \oplus T_p\Sigma_t) \ni ((u,m),(v,n)) \mapsto q_{ab}(m,n). \tag{8.4.12}$$

Here q_{ab} is the three-dimensional pullback of $q_{\mu\nu}$ to Σ_t. Note that this is the same as the pullback to σ.

We can now use $q_{\mu\nu}$ and n^μ to define general projections of any tensor onto two tensors acting on the decomposition (or multiple copies thereof)

$$T_pM = N_p\Sigma_t \oplus T_p\Sigma_t, \tag{8.4.13}$$

in such a way as one is zero on the normal component, and one is zero on the tangential component. The spatial projection tensor projects vectors from T_pM to $T_p\Sigma_t$. When applied to tensor indices it produces a tensor that operates on $T_p\Sigma_t$ (or Cartesian products thereof).

Definition 8.4.5 — Spatial projection tensor. The spatial projection tensor is defined as

$$q^\mu_\nu = g^{\mu\sigma} q_{\sigma\nu}. \tag{8.4.14}$$

Lemma 8.4.6 — Components of spatial projection tensor. The components of the spatial projection tensor are given by

$$q^\mu_\nu = \delta^\mu_\nu + n^\mu n_\nu. \tag{8.4.15}$$

Proof. This follows from a simple computation. ■

The orthogonal (or timelike) projection tensor projects vectors from T_pM to $N_p\Sigma_t$. When applied to tensor indices it produces a tensor that operates on $N_p\Sigma_t$ (or Cartesian products thereof).

Definition 8.4.7 — Orthogonal projection tensor. The orthogonal projection tensor is given by

$$N^\mu_\nu = g_{\mu\sigma} n^\sigma n^\nu. \tag{8.4.16}$$

Lemma 8.4.8 — Components of the orthogonal projection tensor. The components of the orthogonal projection tensor are defined as

$$N^\mu_\nu = -n^\mu n_\nu \tag{8.4.17}$$

Proof. This follows from a simple computation. ■

Spatial tensors are tensors that pull back to Σ_t "without change". These tensors do of course change under pullback in the strict sense of the word, but what we mean is that the action of spatial tensor and its pullback are "the same" in the sense of equation 8.4.12.

Definition 8.4.9 — Spatial tensor. A tensor is a spatial tensor if the result is zero when any of its indices are contracted with the unit normal vector of the foliation.

Before proceeding let us observe an important aspect of our setup. The manifold σ is diffeomorphic to Σ_t. This means that they are mathematically the same manifold, and that any paths or tensors can be pulled back or pushed forward between them. The only difference is that Σ_t is also a submanifold of M which means four-dimensional tensors M are also defined as four-dimensional tensors on Σ_t. At the same time we can also regard Σ_t as embedded in M by the identity map. By this embedding we can pullback four-dimensional forms and pushforward three-dimensional vectors. Because of the orthogonal projection we can in addition also pullback four-dimensional vectors and pushforward three-dimensional one-forms. To see that the orthogonal projection enables us to make this connection note that the spatial projection tensor is a map from T_pM to $N_p\Sigma_t$, and that we can define this map as the pullback map we are seeking. By the usual methods for defining the pushforward map, the pullback of vectors we just defined determines a corresponding pushforward map of functions of vectors (i.e. forms). The results of pullbacks are simple to specify in coordinates that follow the foliation. It simply amounts to setting the all components involving the t-coordinate to 0. Parts of the LQG literature introduce tensor equations which mix three-dimensional and four-dimensional indices in the same equation. We will use the implicit relations indicated in this paragraph, but we will avoid such notational practice.

The projection tensor can be used to define differential geometry on the embedded hyper-surface Σ_t. We have already defined the metric $q_{\mu\nu}$ and we need to define related objects such as the connection and the curvature. We start with the connection (in most situations we treat the words covariant derivative and connection as synonyms).

Definition 8.4.10 — Spatial covariant derivative. The spatial covariant derivative $D_\tau T$ of an (m,n)-tensor T is defined on the basis of the four-dimensional covariant derivative $\nabla_\kappa T$ as

$$D_\tau T^{\alpha_1\dots\alpha_m}_{\beta_1\dots\beta_n} = q^{\alpha_1}_{\rho_1}\dots q^{\alpha_m}_{\rho_m} q^{\sigma_1}_{\beta_1}\dots q^{\sigma_n}_{\rho_n} q^\kappa_\tau \nabla_\kappa T^{\rho_1\dots\rho_m}_{\sigma_1\dots\sigma_n} \tag{8.4.18}$$

We uses the projection operator to define the spatial covariant derivative. The next theorem establishes that this definition corresponds to what we would get from using the Levi-Civita connection derived from the spatial metric.

Proposition 8.4.11 — Spatial Levi-Civita connection. The spatial covariant derivative is the Levi-Civita connection (with regards to the spatial metric $q_{\mu\nu}$) for spatial tensors.

Proof. See [160, p47] for a proof. Note that in this reference D is defined as the Levi-Civita connection and then shown to be equivalent to the definition we have given. ■

Having defined the connection we want to define its curvature.

Definition 8.4.12 — Spatial curvature tensor. The (spatial) curvature tensor of the spatial Levi-Civita connection is defined as

$$R^\sigma_{\mu\nu\rho} u_\sigma = (D_\mu D_\nu - D_\nu D_\mu) u_\rho \tag{8.4.19}$$

Our next target is to express the curvature tensor of spatial Levi-Civita connection D in terms of the ordinary four-dimensional Levi-Civita connection ∇. We use to the spatial covariant derivative to write the first term of the curvature tensor as

$$D_\mu D_\nu u_\rho = q^\sigma_\mu q^\tau_\nu q^\kappa_\rho \nabla_\sigma D_\tau u_\kappa = q^\sigma_\mu q^\tau_\nu q^\kappa_\rho \nabla_\sigma q^\alpha_\tau q^\beta_\kappa \nabla_\alpha u_\beta. \tag{8.4.20}$$

We then write the second term in the same manner and start simplifying. It is useful to introduce the extrinsic curvature tensor.

Definition 8.4.13 — Extrinsic curvature. The extrinsic curvature $K_{\mu\nu}$ is given by

$$K_{\mu\nu} = q_\mu^\sigma q_\nu^\tau \nabla_\sigma n_\tau = \frac{1}{2}\mathcal{L}_n q_{\mu\nu}. \tag{8.4.21}$$

The extrinsic curvature tensor tell us how Σ_t is embedded in M. It contains the information about the curvature that is eliminated from the four-dimensional curvature tensor when defining the spatial curvature tensor. The extrinsic curvature tensor is a symmetric tensor, and it is also a spatial tensor. We write K_{ab} for the three-dimensional pullback to Σ_t.

A rather long derivation brings us from equation 8.4.20 to an expression that relates the four-dimensional to the three-dimensional Ricci tensor and the extrinsic curvature [163][147][25]. This is a classic theorem in differential geometry attributed to Gauss and Codazzi. We skip the derivation for now and just state the result.

Proposition 8.4.14 — Gauss-Codazzi equation. The relation between the four-dimensional and the three-dimensional Ricci tensor is given by

$$R^{(4)} = R^{(3)} + K^{\mu\nu}K_{\mu\nu} - (K_\nu^\nu)^2 - 2\nabla_\mu \left(n^\nu \nabla_\nu n^\mu - n^\mu K_\nu^\nu\right). \tag{8.4.22}$$

Proof. See [160] or [163] for a proof. ■

Note that both sides of this equation are scalars, and that all tensors involved are defined over all of M.

Lemma 8.4.15 — Gauss-Codazzi total derivative. The last term of the Gauss-Codazzi equation is a total derivative.

Proof. This is clear from inspection. ■

We assume the boundary conditions to be such that the final term can be ignored in the context of developing an action [25]. We now give a second version of the Gauss-Codazzi equation where we rewrite it using (the pulled back) three-dimensional tensors (and ignoring the boundary term). We assume without further proof that we can replace any spatial tensor with the corresponding pullback. Note that index free symbols like q become ambiguous when we introduce pullbacks. E.g. the symbol q refers both to the spatial tensor $q_{\mu\nu}$ and the pullback q_{ab} (and the determinants of these). We depend on context to decide between the alternatives. However, from this point on we will mostly be dealing with tensors pulled back to σ. For a more compact notation we define K as the trace K_a^a.

Proposition 8.4.16 — Gauss-Codazzi equation. The relation between the four-dimensional and the three-dimensional Ricci tensor is given by

$$R^{(4)} = R^{(3)} + K^{ab}K_{ab} - K^2. \tag{8.4.23}$$

Proof. Under the assumption that spatial tensors can be pulled back without change this follows by inspection. ■

To complete the action we need to relate the three-dimensional metric determinant to the four-dimensional metric determinant.

Proposition 8.4.17 — Three-dimensional determinant. For a foliation of a four-dimensional spacetime with three-dimensional spatial hypersurfaces as given above, the relation between the four-dimensional and the three-dimensional determinant is given by [160]

$$\sqrt{-det(g)} = N\sqrt{det(q)}, \tag{8.4.24}$$

where N is the lapse scalar field.

Proof. The proof is by a short calculation using Cramer's rule. See [160, p83]. ■

We define that g and q must also serve as symbols for the determinants of the tensors g and q (relying on context to resolve any ambiguity). The enables us to write the EH-action in 3+1 form [25][162].

Proposition 8.4.18 — The ADM action. The relation between the Einstein-Hilbert action written in terms of four-dimensional variables, and the one written in terms of three-dimensional variables is given as

$$S = \int_M \sqrt{-g}R^{(4)} \tag{8.4.25}$$

$$= \int_{\mathbb{R}}\int_\sigma N\sqrt{q}\left[R^{(3)} + K^{ab}K_{ab} - K^2\right] \tag{8.4.26}$$

$$= \int_{\mathbb{R}}\int_\sigma N\sqrt{q}\left[R^{(3)} + (q^{ac}q^{bd} - q^{ab}q^{cd})K_{ab}K_{cd}\right]. \tag{8.4.27}$$

Proof. This follows from propositions 8.4.17 and 8.4.16. ■

Based on this 3+1 action we are ready to perform a Legendre transformation [25][162]. Let us first define the Lagrangian of the ADM action as

$$L_{ADM} = N\sqrt{q}\left[R^{(3)} + K^{ab}K_{ab} - K^2\right] = N\sqrt{q}\left[R^{(3)} + (q^{ac}q^{bd} - q^{ab}q^{cd})K_{ab}K_{cd}\right]. \tag{8.4.28}$$

As a useful shorthand we define

$$G_{abcd} = q^{ac}q^{bd} - \frac{1}{2}q^{ab}q^{cd}. \tag{8.4.29}$$

Based on the ADM Lagrangian we calculate the conjugate momenta of N, N and q to be

$$\Pi = \frac{\partial L_{ADM}}{\partial \dot{N}} = 0 \tag{8.4.30}$$

$$\Pi_a = \frac{\partial L_{ADM}}{\partial \dot{N}^a} = 0 \tag{8.4.31}$$

$$P^{ab} = \frac{\partial L_{ADM}}{\partial \dot{q}_{ab}} = \frac{\sqrt{q}}{2\kappa}\left(q^{ac}q^{bd} - q^{ab}q^{cd}\right)K_{cd} = \frac{\sqrt{q}}{2\kappa}\left(K^{ab} - q^{ab}K\right). \tag{8.4.32}$$

The expression for Π and Π_a cannot be solved to give an expression for $\dot{N}$ and $\dot{N}^a$. This means that $\Pi = 0$ and $\Pi_a = 0$ are primary constraints and will be added to the Hamiltonian. This gives the action [25].

$$S = \frac{1}{2\kappa}\int dx^0 \int d^3x \dot{q}_{ab}P^{ab} + \dot{N}\Pi + \dot{N}^a\Pi_a - [\lambda\Pi + \lambda^a\Pi_a + L_{ADM}] \tag{8.4.33}$$

$$= \frac{1}{2\kappa}\int dx^0 \int d^3x \dot{q}_{ab}P^{ab} + \dot{N}\Pi + \dot{N}^a\Pi_a - [\lambda\Pi + \lambda^a\Pi_a + N^a\mathcal{H}_a + N\mathcal{H}], \tag{8.4.34}$$

with the ADM Lagrangian expressed as

$$L_{ADM}(N^a, \Pi_a, N, \Pi, q_{ab}, P^{ab}) = N^a \mathcal{H}_a + N\mathcal{H}. \tag{8.4.35}$$

Where we have defined

$$\mathcal{H}_a \left(P^{ab}, q_{cd}\right) = -2D_b P^b_a = -2q_{ac} D_b P^{bc} \tag{8.4.36}$$

$$\mathcal{H} \left(P^{ab}, q_{cd}\right) = \frac{2\kappa}{\sqrt{q}} \left(P^{ab}P_{ab} - \frac{1}{2}P^2\right) - \frac{\sqrt{q}}{2\kappa} R^{(3)} \tag{8.4.37}$$

$$= \frac{2\kappa}{\sqrt{q}} G_{abcd} \Pi^{cd} \Pi^{ab} - \frac{\sqrt{q}}{2\kappa} R^{(3)}. \tag{8.4.38}$$

The two primary constraints must be preserved under "time" evolution (Hamiltonian evolution). Calculating their Hamiltonian evolution leads to the two secondary constraints defined by [25]

$$\dot{\Pi} \ = \{H, \Pi\} = \mathcal{H} = 0 \tag{8.4.39}$$

$$\dot{\Pi}_a = \{H, \Pi_a\} = \mathcal{H}_a = 0. \tag{8.4.40}$$

These terms are already present in the action, but this shows that they are actually constraints and that N and N^a are their Lagrange multipliers. These constraints are called the Hamiltonian constraint and the diffeomorphism constraint. The new secondary constraints, as well as the previous primary constrains are all first class constraints. To gain further insight into the action we find the equation of motion for N and N^a. We get that

$$\frac{\partial N}{\partial t} \ = \frac{\delta H}{\delta \Pi} = \lambda \tag{8.4.41}$$

$$\frac{\partial N^a}{\partial t} = \frac{\delta H}{\delta \Pi_a} = \lambda^a. \tag{8.4.42}$$

We already know that λ and λ^a are arbitrary functions (Lagrange multipliers). These equations means that N and N^a are also arbitrary functions, since they are integrals of arbitrary functions. We can therefore just set Π and Π_a to zero and disregard all the terms involving them. We are left with P^{ab} and q_{cd} as the only dynamic variables. This simplifies the action to [157]

$$S = \frac{1}{2\kappa} \int dx^0 \int d^3x \left\{ \dot{q}_{ab} P^{ab} - \left[N^a \mathcal{H}_a(P^{ab}, q_{cd}) + N\mathcal{H}(P^{ab}, q_{cd}) \right] \right\}. \tag{8.4.43}$$

The Hamiltonian becomes a linear combination of two constraints and is given as [157]

$$H = \frac{1}{2\kappa} \int dx^0 \int d^3x \left\{ N^a \mathcal{H}_a(P^{ab}, q_{cd}) + N\mathcal{H}(P^{ab}, q_{cd}) \right\}. \tag{8.4.44}$$

To finish our presentation of the ADM formalism we list the Poisson brackets. We write them as [25]

$$\{P^{ab}(x), q_{cd}(y) \ \} = \kappa(\delta^a_c \delta^b_d - \delta^a_d \delta^b_c)\delta(x, y) \tag{8.4.45}$$

$$\{P^{ab}(x), P^{cd}(y)\} = 0 \tag{8.4.46}$$

$$\{q_{ab}(x), q_{cd}(y) \ \ \} = 0. \tag{8.4.47}$$

Step 2: Triad formulation

The next step in reformulating GR in the direction of gauge theory is to introduce orthonormal bases called triads. Such non-coordinate bases also appear in the Palatini-Holst formulation in the form of tetrads. We start out with some general aspect of a orthonormal basis using triads as our example. A triad field is a set of three vector fields that are orthonormal at each point. Such field are also called frame fields, and are defined over all of M (see [164] for a general introduction). The triad basis can be expressed using the coordinate basis as[2]

$$\vec{e}_i = e_i^a \vec{e}_a \quad (\vec{e}_a = \partial_a). \tag{8.4.48}$$

Here the symbol e_i^a represents the a-th component of the $\vec{e}_i$ triad basis in the coordinate basis. These symbols express the components of the orthonormal vector fields in the coordinate basis. We could equivalently say that it is the basis transformation matrices representing the transformation from the orthonormal basis to the coordinate basis. These transformation matrices are invertible and we can use the inverse matrix e_a^i to write

$$\vec{e}_a = e_a^i \vec{e}_i \quad (\vec{e}_a = \partial_a, e_a^i e_j^a = \delta_j^i). \tag{8.4.49}$$

This basis transformation field (matrix field) is often called the vierbein-field (dreibein-field) [164]. As for any basis transformation matrix, we can use this matrix to re-express tensor components in the new basis by writing

$$T_i = e_i^a T_a. \tag{8.4.50}$$

The definition of the components of the spatial metric in the coordinate basis is

$$q_{ab} = \langle \partial_a, \partial_b \rangle = \langle e_a^i \vec{e}_i, e_b^j \vec{e}_j \rangle = e_a^i e_b^j \langle \vec{e}_i, \vec{e}_j \rangle = \delta_{ij} e_a^i e_b^j. \tag{8.4.51}$$

The spatial metric can be written in the orthonormal basis as

$$g = \delta_{ij} e^i \otimes e^i. \tag{8.4.52}$$

This expression is perhaps unexpected for a dynamical metric, in the sense that the components of the metric are constant over all of M. This only means that the one-form basis e^i has been adjusted at every point so that equation 8.4.52 holds. Hence, the true variability of the metric is encoded in the coordinate basis components of the frame fields. From this equation we conclude that triad fields contain basically the same information as the metric in the sense that knowing the triad fields determines the metric field. The metric also determines the triads but not uniquely because introducing triads introduces yet another level of redundancies.

The metric derived from the triad fields are invariant under local $SO(3)$ rotations of the triads. One metric correspond to an infinite set of $SO(3)$ related triads. Triads in themselves are not directly physically relevant, they are only relevant as an alternate specification of metric information. Thus, when using triads as the basic variables, this induces a $SO(3)$ ($SU(2)$) gauge redundancy in the theory. Corresponding to this added redundancy in the fundamental variables comes a new set of constraints called the Gauss constraints denoted by $\mathcal{G}$ [165].

The terminology of raising and lowering of indices is nothing but a quick way to refer to the musical isomorphisms between vectors and one-forms induced by the metric tensor. When using orthonormal bases in addition to coordinate bases, tensors with a mixed set of indices are possible. The individual indices must be raised and lowered by the metric expressed in the corresponding basis.

[2]For the next few lines we will temporarily apply arrow notation to indicate that e represents a vector.

Special considerations also apply when calculating derivatives of tensors involving orthonormal bases. In the three-dimensional coordinate basis the derivative D_a can be expressed using the three-dimensional Christoffel symbol as [165]

$$D_a \mathrm{e}_b = \Gamma^c{}_{ab} \mathrm{e}_c. \tag{8.4.53}$$

We can express the covariant derivative of any tensor expressed in a coordinate basis by expanding this expression by applying the Leibnitz rule to tensor products. For a non-coordinate basis we need to use the spin connection. The spin connection is a matrix field of $n \times n$ one forms. We write this as

$$\omega^{ab} = \omega_i^{ab} e^i \tag{8.4.54}$$

An example of writing derivatives of mixed tensors is given by [164]

$$D_a V_i{}^b = \partial_a V_i{}^b + \omega_a{}^b{}_c V_i{}^c - \Gamma^c{}_{ia} V_c^b. \tag{8.4.55}$$

The derivatives of mixed tensors get a term with Christoffel symbol for each coordinate basis index, and a term with a spin-connection symbol for each orthonormal basis index.

We now switch to considering densitized triads E_a^i. The densitized triads E_a^i are simply the ordinary triad e_a^i multiplied by the square root of the determinant of the three-dimensional metric. In addition we change the indices on the extrinsic curvature to be mixed. The transformation are written as

$$e_a^i \to E_a^i = \sqrt{q} e_a^i \tag{8.4.56}$$

$$K_{jk} \to K_a^i = \delta^{ik} e_a^j K_{jk}. \tag{8.4.57}$$

Using these new variables the Poisson brackets becomes [25]

$$\left\{ E_i^a(x), K_b^j(y) \right\} = \kappa \delta_b^a \delta_j^i \delta(x, y) \tag{8.4.58}$$

$$\left\{ K_a^i(x), K_b^j(y) \right\} = 0 \tag{8.4.59}$$

$$\left\{ E_i^a(x), E_j^b(y) \right\} = 0. \tag{8.4.60}$$

We can restate the action in these new variables and we get [25]

$$S = \int E_i^a \dot{K}_a^i - (N_b \mathcal{H}^b + N\mathcal{H} + \Lambda^i \mathcal{G}_i). \tag{8.4.61}$$

We do not state the explicit form of the constraints in the new variables as we will not be needing them.

Step 3: Ashtekar variables

We have now formulated GR using densitized triads and the extrinsic curvature with mixed indices. The final step to reach the Ashtekar variables is to convert from the extrinsic curvature to Ashtekar-Barbero connection. The Ashtekar-Barbero connection is formed by adding a multiple of the extrinsic curvature to the spin connection [25][162].

$$\Gamma_a^i \to A_a^i = \Gamma_a^i + \beta K_a^i \tag{8.4.62}$$

Some authors also perform a rescaling of the densitized triads

$$E_a^i \to \frac{1}{\beta} E_a^i, \tag{8.4.63}$$

and we will do so too (we will use the same symbol for the rescaled triad). The rescaling in equation 8.4.63 will cause the β factor drop out of the right-hand side of the Poisson brackets and also the explicit expressions for the constraints are changed by factors involving β.

The connection and the triad as mathematical objects are defined by the equations

$$A = A_a^i \, \mathrm{d}x^a \otimes \tau_i \quad \text{with} \quad \tau_i \in \mathfrak{su}(2) \tag{8.4.64}$$

$$E = E_i^a \partial_a \otimes \tau^i \quad \text{with} \quad \tau^i \in \mathfrak{su}(2)^*. \tag{8.4.65}$$

By the dual of Lie algebra we mean the vector space dual. The relation $\tau^i(\tau_i) = \delta^i_j$ holds between the vector space basis and its dual basis. The Poisson brackets between these new variables can be calculated from the Poisson brackets given in step 2. The result of this calculation is

$$\left\{ A_a^i(x), E_j^b(y) \right\} = \kappa \delta_a^b \delta_j^i \delta(x,y) \tag{8.4.66}$$

$$\left\{ A_a^i(x), A_b^j(y) \right\} = 0 \tag{8.4.67}$$

$$\left\{ E_i^a(x), E_j^b(y) \right\} = 0. \tag{8.4.68}$$

The action becomes [25][157]

$$S = \int E_i^a \dot{A}_a^i - (N^a \mathcal{H}_a + N\mathcal{H} + \Lambda^i \mathcal{G}_i) \tag{8.4.69}$$

We determine the explicit form of the constraint in the Ashtekar variables to be [157][166]

$$\mathcal{H} = \beta^2 \frac{E_i^a E_j^b}{2\sqrt{q}} \epsilon^{ij}{}_k F_{ab}^k - \frac{1+\beta^2}{\sqrt{q}} K_{[a}^i K_{b]}^j E_i^a E_j^b \tag{8.4.70}$$

$$\mathcal{H}_a = F_{ab}^i E_i^b \tag{8.4.71}$$

$$\mathcal{G}_i = D_a E_i^a = \partial_a E_i^a + \epsilon_{ij}^k A_a^j E_k^a. \tag{8.4.72}$$

We have defined F_{ab}^i as

$$F_{ab}^i = 2D_{[a} A_{b]}^i = \partial_a A_b^i - \partial_a A_b^i + \epsilon_{jk}^i A_a^j A_b^k. \tag{8.4.73}$$

This concludes our derivation of Ashtekar formulation of GR.

Approach 2: Hamiltonian formulation of the Palatini-Holst action

The Palatini and the Palatini-Holst action produces the same equations of motions (for pure gravity). However, the Legendre transforms of these theories are quite different [158]. The Palatini action transforms into a metric non-connection based Hamiltonian using cotriads, while the Palatini-Holst action transforms directly to the Ashtekar connection based Hamiltonian [158]. The basic variables are given by the following objects. First we have orthonormal co-tetrad $e^I = e^I_\mu dx^\mu$, and its corresponding orthonormal tetrad $e_I = e^\mu_I e_\mu$. Second we have the connection 1-form field $\omega_\mu{}^I{}_J$. The curvature of the connection is represented by the curvature two-form Ω^{IJ}. The gravitational field previously described by the metric is now described using $\left(e^I, \omega_\mu{}^I{}_J\right)$. The Palatini-Holst action is given by

$$S_{PH}\left[e, \omega\right] = \frac{1}{2\kappa} \int \epsilon_{IJKL} e^I \wedge e^J \wedge \Omega^{KL} - \frac{1}{\kappa\beta} \int e^I \wedge e^J \wedge \Omega_{IJ}. \tag{8.4.74}$$

When we use the ADM foliation on this action we find that the spin connection decomposes into two $SU(2)$ connections. The cotetrad fields become cotriad fields plus the lapse and the shift

vector. By these means we arrive directly at the Ashtekar formulation of gravity [158]. This is not surprising since this was the motivation for constructing the Palatini-Holst action in the first place [167]. It is still interesting that such a simple modification of the Palatini action can lead so directly to the Ashtekar formulation. We will not present the details of this Legendre transformation (see e.g. [158] for a review).

8.5 The graph formulation of GR

We have now reformulated GR as a Yang-Mills theory with three sets of constraints. At this point one could attempt to quantize the theory by using methods used for gauge theories. This would represent canonical quantum gravity in the connection formulation. This approach has several problems and has not been successful.

Observe first that the Poisson brackets for Ashtekar variables are only formal expressions since they involve Dirac deltas like $\delta(x, y)$. The Dirac deltas are not functions (they are distributions), and therefore the Poisson bracket is not well defined as a binary operation on the function space. In addition to the distributional nature of the Poisson bracket, the infinite dimensional configuration space of Ashtekar variables, consisting of all possible smooth connections, is a space on which integration theory is hard to define [168]. Without a suitable measure for the configuration space we cannot make an inner product, and quantum theory is not possible.

To make the Poisson algebra mathematically well defined the Ashtekar variables have to be smeared over submanifolds of the spatial slice. Smearing over the full three-dimensional slice presents a problem, since there is no metric available to define the integration measure. Because of this such (brute-force) three-dimensional smearing will not be our chosen approach. Instead we will smear the gauge fields along one-dimensional edges, and smear the triad fields over two-dimensional surfaces. The resulting objects will be called holonomies and fluxes respectively. These smearings (i.e. integrations) are well defined in the absence of a metric. In addition to eliminating the distributional nature of the Poisson brackets, these smearings will also lead to a different configuration space where a measure can be constructed.

An essential feature of the suggested smearings is that the holonomies and fluxes that results from this will have quite simple behaviors under the transformations induced by the Gauss constraint and the diffeomorphism constraint. Under Gauss transformations only the endpoint of holonomies are affected. Under a diffeomorphism we can express the change in the holonomy by letting the diffeomorphism act on the associated one-dimensional edge.

Edges and graphs

We start by giving definitions of the basic objects needed for the smearing procedure [25].

Definition 8.5.1 — Curve. A curve is a continuous piecewise semi-analytic map

$$c\colon\ [0,1] \to \sigma. \tag{8.5.1}$$

We also require that $c([0,1])$ is contained in a compacted subset of σ.

The set of all curves is denoted $\mathcal{C}$. We call $b(c) = c(0)$ the beginning point of the curve and $f(c) = c(1)$ the final point of the curve. We refer collectively to the beginning and the end of a curve as endpoints of the curve. The range of a curve c is defined as $r(c) = c([0,1])$.

We want to make the set of curves into an algebraic structures. We start by defining multiplication and inversion.

Definition 8.5.2 — Operations on curves. Two curves c_1 and c_2 are composable if $f(c_1) = b(c_2)$. The composition of two composable curves c_1 and c_2 is defined as

$$(c_1 \circ c_2)(t) = \begin{cases} c_1(2t) & t \in [0, \frac{1}{2}] \\ c_2(2t-1) & t \in [\frac{1}{2}, 1] \end{cases} . \tag{8.5.2}$$

The inverse of a curve is defined as

$$c^{-1}(t) = c(1-t). \tag{8.5.3}$$

To allow for associativity, proper inverses and units we need to let the operations just defined operate on equivalence classes of curves [3].

Definition 8.5.3 — Equivalent curves. The relation $\sim$ on $\mathcal{C} \times \mathcal{C}$ is defined as follows. Let $c_1, c_2 \in$. We write $c_1 \sim c_2$ when

1. Start points are identical: $b(c_1) = b(c_2)$.
2. End points are identical: $f(c_1) = f(c_2)$.
3. The curve c_1 is identical to c_2 up to a finite number of retracings and a semianalytical reparametrization.

Lemma 8.5.4 — Equivalence relation. The relation $\sim$ on $\mathcal{C} \times \mathcal{C}$ is an equivalence relation.

Proof. For a proof see [25, page 164]. ■

The equivalence class of a curve $c \in \mathcal{C}$ is called a path, and is denoted p_c. The set of all paths p_c is denoted $\mathcal{P}$. We shall develop the algebraic perspective on $\mathcal{P}$ further in a later section. We shall then see in proposition 8.5.12 that $\mathcal{P}$ becomes a groupoid.

Before we proceed to holonomies and fluxes we define some very important terminology regarding equivalence classes of curves. Most of our discussions going forwards will be related to edges and collections of edges (graphs).

Definition 8.5.5 — Edge. An edge in an equivalence class $e \in \mathcal{P}$ of a curve $c_e \in \mathcal{C}$ where c_e is not just piecewise semianalytic but also semianalytic.

Definition 8.5.6 — Graph. A graph γ in σ is a finite collection of edges $\{e_i\}$ in σ that are disjoint, or meet only at their endpoints.

The collection of all edges is denoted Γ. The set of all endpoints of the edges in a graph γ is denoted $V(\gamma)$ (vertices). The set of all edges in a graph γ is denoted $E(\gamma)$ (edges).

Holonomies and fluxes

We have establish that we need to smear the connection variables and we now proceed with the details. In the following we denote a connection (connection field) by A, and the set of all connections by $\mathcal{A}$. A connection defines a notion of parallel transport of tangent vectors along a given path. Given a connection and a path this notion defines a map between the tangent spaces at each end of the path. Extracting this map is what we mean by smearing the connection. Smearing the connection $A \in \mathcal{A}$ along a along a curve $c \in \mathcal{C}$, going from starting point $b(c)$ to end point $f(c)$ gives us the holonomy along this curve [157]. To facilitate a definition of the holonomy in terms of a differential equation we can regard the holonomy as a function of the maximal value of the

[3] This is similar to developing the fundamental group in homotopy theory

curve parameter t. This makes the holonomy a function of t, and this makes the derivative of the holonomy with respect to t well defined.

Definition 8.5.7 — Holonomy (differential equation). Let $h[c, A, t]$ be an unknown $\mathrm{SU}(2)$-valued function[a] that we label as "the holonomy as a function of the curve parameter" $t \in [0, 1]$. We define the holonomy of the (whole) curve as $h[c, A] = h[c, A, 1]$ and we require that $h[c, A, 1] = 1_G$. (The symbol G refers to the group $\mathrm{SU}(2)$.) The function $h[c, A, t]$ is defined by the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} h[c, A, t] = h[c, A, t]\; A^i_a(c(t))\tau_i\, \dot{c}^a(t). \tag{8.5.4}$$

The notation $\dot{c}^a(t)$ denotes the tangent vector to the curve c.

[a] To be extra precise we will require this function to be valued in the fundamental representation of $\mathrm{SU}(2)$. This is required to be able to define an action of $h[c, A, t]$ on on the linear sums of τ_i.

The next definition could be given as a proposition that follows from the previous definition. We give it as an alternative more global (and less rigorous) definition, and claim without proof that properly interpreted both our definitions are equivalent.

Definition 8.5.8 — Holonomy (integral). The holonomy $h[c, A]$ of a the connection A on a manifold σ along a curve c is defined as

$$h[c, A] = \mathcal{P} \exp\left(\int_c A^i_a(c(t))(\dot{c}^a)\tau_i dt \right). \tag{8.5.5}$$

The symbol $\mathcal{P}$ represents the path ordering operator.

The holonomy components are denoted by $h[c, A]_{\alpha\beta}$, where α and β are matrix component indices. Note that all the holonomy definitions we have given so far refer to the $j = 1/2$ defining representation of $su(2)$ (since τ_i belong to the defining representation). We extend our holonomy concept to other representations by defining the objects $h^{(j)}[c, A]$ and $h^{(j)}[c, A]_{\alpha\beta}$. In this case we are replacing τ_i by $\tau_i^{(j)}$ in the definitions. The symbols $\tau_i^{(j)}$ refers to the j-representation of $\mathfrak{su}(2)$. The group $\mathrm{SU}(2)$ is connected which means that the exponential map provides a one-to-one correspondence between representations of $\mathfrak{su}(2)$ and representations of $\mathrm{SU}(2)$.

The holonomy component function is the simplest example of a cylindrical functions (a notion we will soon define). Let us make a note of an important fact about holonomies.

Proposition 8.5.9 The holonomy only depends of the path (i.e. the equivalence class) of a curve.

Any path is a collection of edges and we can therefore focus on holonomies along edges.

We can understand the smearing better by imagining the edge e to be subdivided in many short segments $\{e_i\}$ and letting the connection be constant on each such segment. The integration over the segment e_i then becomes just a multiplication of a Lie algebra element by some number. The exponential of this is just an element of the gauge group which we denote $U(e_i)$. Repeating this procedure (and observing the ordering induced by path ordering) gives us an approximate expression as a chain of group elements multiplied together

$$h[e, A] = U(e_1) \cdot \ldots \cdot U(e_n). \tag{8.5.6}$$

Before moving on to fluxes we analyze the transformation properties of the holonomies [25]. The transformation induced by the Gauss constraint is an ordinary gauge transformation. A gauge

transformation on σ is given by a map $\Lambda : \sigma \to G$. Since by the decomposition of the holonomy we made in equation 8.5.6 the intermediate transformation along the bulk of e will cancel against each other. We hence conclude that a gauge transformation only affects the holonomy at the endpoints. The effect of such a transformation on a holonomy is given by [169]

$$h_e[\mathcal{A}] \to \Lambda(b(e)) h_e[\mathcal{A}] \Lambda(f(e)). \tag{8.5.7}$$

The transformation induced by the diffeomorphism constraint is equivalent to a diffeomorphisms $\phi\colon \sigma \to \sigma$. The effect on a holonomy is given by [169]

$$h_e[\mathcal{A}] \to h_{\phi(e)}[\mathcal{A}]. \tag{8.5.8}$$

We have now found a smeared version of the connection variable and we wish to implement the same for the conjugate momenta. This is done utilizing the fact that the Hodge dual of a (densitized) vector on three-dimensional manifold is a two-form. Two-forms can be integrated over 2-dimensional surfaces without employing a metric [25].

Definition 8.5.10 — Dual of triad. The dual (i.e. the Hodge dual) of the densitized triad is defined as

$$\tilde{E} = *(E^a_i dx_a \otimes \tau^i) = \epsilon_{abc} E^a_i (dx^b \wedge dx^c) \otimes \tau^i. \tag{8.5.9}$$

Smearing E over a surface S gives the flux of S through this surface.

Definition 8.5.11 — Flux. Let S be a surface in σ and let $\tilde{E}$ be the dual of E. Let $f = f^j \tau_j$ be a function from $S \subset \sigma$ to $\mathfrak{su}(2)$. The flux of the dual of E over the surface S smeared by f is defined as

$$\begin{aligned} F_{S,f}[E] &= F[S, f, E] && (8.5.10)\\ &= \int_S \tilde{E} f && (8.5.11)\\ &= \int \epsilon_{abc} E^a_i (dx^b \wedge dx^c) \otimes \tau^i f^j \tau_j && (8.5.12)\\ &= \int \epsilon_{abc} E^a_i (dx^b \wedge dx^c) \otimes \delta^i_j f^j && (8.5.13)\\ &= \int \epsilon_{abc} f^i E^a_i dx^b \wedge dx^c. && (8.5.14) \end{aligned}$$

We do not need the transformation properties of the fluxes and we will not analyze them.

Poisson brackets of holonomies and fluxes

We have identified holonomies and fluxes as good candidates for elementary variables of the classical theory. To proceed further we must compute the Poisson brackets of the holonomy components with the fluxes. We first simplify the expression involved in the Poisson bracket by writing

$$\begin{aligned} \{F_{S,f}, h_e\} &= \int d^3x \left(\frac{\delta h}{\delta E^a_i(x)} \frac{\delta F}{\delta A^i_a(x)} - \frac{\delta h}{\delta A^i_a(x)} \frac{\delta F}{\delta E^a_i(x)} \right) && (8.5.15)\\ &= \int d^3x \left(\frac{\delta F}{\delta E^a_i(x)} \frac{\delta h}{\delta A^i_a(x)} \right). && (8.5.16) \end{aligned}$$

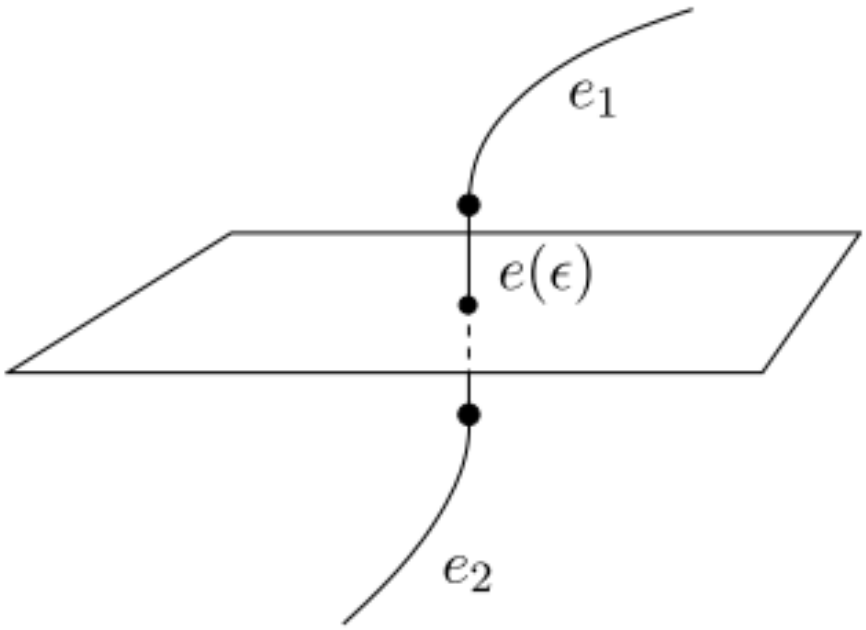


Figure 8.7: The diagram shows how we split the edge e of an holonomy h_e into three parts e_1, e_2 and $e(\epsilon)$ when we calculate the Poisson bracket between h_e and the flux F illustrated by the plane surface. The intersection point will be referred to as p (this notation is not shown in the illustration). Illustration from [170].

This result follows immediately since h in independent of E, and F is independent of A. Let us denote the (sub)holonomy of the curve c between curve-parameter values t_1 and t_2 by $h_e^{t_1,t_2}$ $(0 \leq t_1 \leq t_2 \leq 1)$. We start by calculating

$$\frac{\delta h_e\left[A\right]}{\delta A_a^i(x)} = \int_0^1 dt\ \left\{\dot{c}^a h_e^{0,t}\left[A\right] \tau_i h_e^{t,1}\left[A\right] \delta(c(t)-x)\right\}. \tag{8.5.17}$$

Next we calculate

$$\frac{\delta F_{S,f}[E]}{\delta E_i^a(x)} = \int_S dudv \left\{f^i \epsilon_{abc} dx^b \wedge dx^c \delta(X(u,v)-x)\right\}. \tag{8.5.18}$$

Let the intersection between e and S be as shown in figure 8.7, and let us denote the intersection by p. Using the Poisson bracket simplification we made earlier we get

$$\begin{aligned}
\{F_{S,f}, h_e\} &= \int d^3x \left(\frac{\delta F}{\delta E_i^a(x)} \frac{\delta h}{\delta A_a^i(x)}\right) && (8.5.19)\\
&= \int d^3x \int_0^1 dt\ \left\{\dot{c}^a h_e^{0,t}\left[A\right] \tau_i h_e^{t,1}\left[A\right] \delta(c(t)-x)\right\} \\
&\quad \int_S dudv \left\{f^i \epsilon_{abc} dx^b \wedge dx^c \delta(X(u,v)-x)\right\} && (8.5.20)\\
&= \int d^3x \int_0^1 dt \int_S dudv \\
&\quad \left\{\dot{c}^a h_e^{0,t}\left[A\right] \tau_i h_e^{t,1}\left[A\right] \left\{f^i \epsilon_{abc} dx^b \wedge dx^c \delta(c(t)-X(u,v))\right\}\right\} && (8.5.21)\\
&= \pm\beta\kappa f^i(p) h_{e_1}\left[A\right] \tau_i h_{e_2}\left[A\right] && (8.5.22)
\end{aligned}$$

In the final step the details of the flux integration simplifies to just ± 1 that depends on the orientation of the edge relative to the orientation of the surface. The details can be found in [77]. In the component version the Poisson bracket reads

$$\left\{F_{S,f}, h_e\left[A\right]_{\alpha\beta}\right\} = \pm\beta\kappa f^i(p) h_{e_1}\left[A\right]_{\alpha\rho}\tau_{i,\rho\sigma} h_{e_2}\left[A\right]_{\sigma\beta}, \tag{8.5.23}$$

where index summation is implied. Holonomies Poisson commute with holonomies, but the Poisson brackets between fluxes are non-zero and complicated, and we shall not be needing them (see [171] for the details).

Groupoids

By using holonomies we have defined a mapping from edges to $\mathrm{SU}(2)$ group elements. This has an obvious extension to a mapping from graphs to tuples of group elements (these mappings are presented in the next subsection). In this section we look at a more abstract way of defining such mappings. This will enable us to make a direct constructive definition of the configuration space $\overline{\mathcal{A}}$. By the use of projective limits we will also define another space $\overline{X}$ that will be in bijective correspondence with $\overline{\mathcal{A}}$. We will use $\overline{X}$ to assign a topology to $\overline{\mathcal{A}}$. The necessary definitions of groupoids, groupoid morphisms and projective limits can be found in the appendices.

As we have mentioned the set of equivalence classes of curves can be given an algebraic structure. Since the end points don't always match up between (equivalence classes of) curves the multiplication formed between curves defines only a partial binary operation. Specifically it defines a groupoid.

Proposition 8.5.12 — P is a groupoid. The set $\mathcal{P}$ is a groupoid under the operations of composition and inverse given by definition 8.5.2.

Proof. See [25, p166] for a proof. ■

We can also define the groupoid generated by a single graph.

Definition 8.5.13 — Groupoid of a graph. The groupoid of a graph γ is denoted by $l(\gamma)$, and is defined by adding inverses of all edges as well as adding all finite compositions of edges and inverses.

The groupoid $l(\gamma)$ is a subgrupoid of $\mathcal{P}$. A subgroupoid generated by a graph is called a tame (sub)groupoid. The set of all tame (sub)groupoids is denoted $\mathcal{L}$. We will use this set as an index set and for that role we need to equip it with a partial order. We write $l \prec l'$ when l is a subgrupoid of l'. We also note that any two elements of $\mathcal{L}$ have an upper bound given by the union of the underlying graphs. We state the relevant facts in a theorem.

Theorem 8.5.14 The set $\mathcal{L}$ with the relation $\prec$ is a directed poset.

Proof. See [25, p171] for a proof. ■

Proposition 8.5.15 Let A be an element in the set of smooth connection $\mathcal{A}$. The map $f_A\colon \mathcal{P} \to SU(2)$ defined by

$$f_A\colon p_c \mapsto h[p_c, A] \tag{8.5.24}$$

is a groupoid homomorphism.

We will use this property to give rigorous definition of the configuration space and various other spaces of the theory. We start by defining the extended configuration space.

Definition 8.5.16 — Extended configuration space. The extended configuration space $\overline{\mathcal{A}}$ is the space given by

$$\overline{\mathcal{A}} = \mathrm{Hom}(\mathcal{P}, SU(2)), \tag{8.5.25}$$

where $\mathrm{Hom}(\mathcal{P}, SU(2))$ is the set of all groupoid morphisms from $\mathcal{P}$ to $SU(2)$.

This space will be the classical configuration space of the theory we are building. The space $\mathcal{A}$ of smooth (or continuous) connections has a natural identification as a subset of $\overline{\mathcal{A}}$. The identification is given by identifying an element $A \in \mathcal{A}$ with the map $p \mapsto h[p_c, A]$ as given in proposition 8.5.15.

We need to equip $\overline{\mathcal{A}}$ with a topology. This will be done by defining compact Hausdorff space $\overline{X}$ and then constructing a bijection between $\overline{\mathcal{A}}$ and $\overline{X}$, and use this bijection to define a topology on $\overline{\mathcal{A}}$. The construction of the space $\overline{X}$ proceeds by using a projective family, and then taking its projective limit (see appendix A for some details on the procedure). We first define some spaces that we will use to construct $\overline{X}$.

Definition 8.5.17 — Hom space of a subgrupoid. Let $l = l(\gamma)$ be a subgrupoid of $\mathcal{L}$. The Hom space of the subgrupoid l is defined as

$$X_l = \mathrm{Hom}(l, SU(2)). \tag{8.5.26}$$

Let $n = |E(\gamma)|$ be the cardinality of the set of edges of γ, and let $l = l(\gamma)$. Any homomorphism in X_l is just a map from each edge element of γ to an element of $\mathrm{SU}(2)$. Thus each homomorphism X_l defines a element of $(\mathrm{SU}(2))^n$. One can show that there is a bijection between each of the space X_l and $(\mathrm{SU}(2))^n$. We use this to define transport the compact Hausdorff topology from $(\mathrm{SU}(2))^n$ to X_l.

For any subgrupoids l_1 and l_2 such that $l_2 \prec l_1$ there is a natural surjective projection $P_{l_1 l_2}$ from X_{l_1} to X_{l_2} by mapping each morphism $\mathrm{Hom}(l_1, SU(2))$ to the corresponding $\mathrm{Hom}(l_2, SU(2))$ by restriction on the domain. Since composition just gives a further restriction on the domain we automatically have

$$P_{l_1 l_2} \circ P_{l_0 l_1} = P_{l_0 l_2}. \tag{8.5.27}$$

Proposition 8.5.18 — Projective family. The family of sets $\{X_l\}$ index by $l \in \mathcal{L}$ along with the sets of projections $\{P_{l'l} \mid l \prec l' \wedge l, l' \in \mathcal{L}\}$ is a projective family. We denote the projective family (semi-casually) by $\{X_l, P_{l'l}\}$.

Proof. See [25][172][169] for a proof. ■

For each projective family one can define the projective limit. We present it as a subspace of the product Hom space.

Definition 8.5.19 — Product Hom space. The product Hom space is defined as

$$X_\infty = \prod_{l \in \mathcal{L}} X_l. \tag{8.5.28}$$

The space X_∞ is an infinite product of compact Hausdorff topological spaces and by Tychonow's theorem the product topology on X_∞ is also compact. The Hausdorff property is also preserved by products (see [104, Ch10] for a proof of these statements).

Definition 8.5.20 — Projective limit of $\{X_l, P_{l'l}\}$. Let a generic element of X_l be denoted by x_l. The projective limit of $\{X_l, P_{l'l}\}$ is the subspace of X_∞ given by

$$\overline{X} = \{\{x_l\} \in X_\infty \mid P_{l'l} x_{l'} = x_l\}\,. \tag{8.5.29}$$

The projective limit of $\{X_l, P_{l'l}\}$ is a closed subspace of X_∞ and it is therefore a compact Hausdorff space with respect to the subspace topology. This completes the construction of the space $\overline{X}$. A bijection can be made with the space $\overline{\mathcal{A}}$ such that $\overline{\mathcal{A}}$ can also be made into a compact Hausdorff topological space. To see how this works first note that

$$\overline{\mathcal{A}} = \mathrm{Hom}(\mathcal{P}, SU(2)) \tag{8.5.30}$$

so that each element of $\overline{\mathcal{A}}$ is a homomorphism. Note further that $\overline{X}$ is a tuple of the same type of homomorphisms just defined on subsets of the domain of the morphisms in $\overline{\mathcal{A}}$. We can therefore define a map ϕ between them. Letting f be an element of $\overline{\mathcal{A}}$ we define the map

$$\phi\colon \overline{\mathcal{A}} \to \{X_l, P_{l'l}\} \tag{8.5.31}$$

$$\phi\colon f \mapsto \{f|_l\}\,. \tag{8.5.32}$$

There is no perfect notation for the last line. It maps every element $f \in \overline{\mathcal{A}}$ to a potentially infinite sequence. Our intended meaning is that each element of the projective family limit defines a sequence of graphs referred to by the parameter l. The map is then such that, given an element $f \in \overline{\mathcal{A}}$, for each value of the parameter l on defines the restriction of the homomorphism to the domain l. For further details see [25][172][169].

Cylindrical functions

We have defined the holonomy, and the holonomy components, of a single edge on the spatial slice σ. We will now expand the holonomy concept to include graphs (sets of edges), and we will look at this extended holonomy concept as a map, induced by the chosen graph, from the set of all connections $\mathcal{A}$ (or $\overline{\mathcal{A}}$) to tuples of group elements. We will first sketch some concepts using ideas from lattice gauge theory [146].

Imagine the smearing along edges begin replaced by smearing along a regular lattice (which is not a meaningful concept without a metric, but this is just for pedagogical ease). In this setup a configuration of the connection field can be though of as assigning to each edge of the lattice an element of the group G. If we let the number of edges be N, we obtain the space G^N.

In LQG we use a similar construction, but we must use a "free" version of a lattice called an embedded graph. Each edge of the graph is assigned an element of the group G. This configuration space is similar to the one for lattice theory except we need to specify which graph we are referring to. The total configuration space will be all such group element assignments over the set of all "inequivalent" graphs. Note that while lattice gauge theory uses the lattice to define a discretization of the path integral, in LQG we will use graphs to define cylindrical functions which then will be quantized by canonical methods.

Each of the N edges e_i in a graph γ_N defines a corresponding holonomy map $h_{e_i}[\mathcal{A}]$, and the graph γ_N defines a map $h_{\gamma_N} : (\gamma_N, \mathcal{A}) \mapsto (h_{e_1}[\mathcal{A}], \ldots, h_{e_N}[\mathcal{A}]) \in G^N$. For a given graph γ_N (with N edges) the configuration space is $SU(2)^N$. We now define a certain type functions which in an informal notation could be describe by

$$f_{\gamma_N} : \mathcal{A} \xrightarrow{h_{\gamma_N}} \left(SU(2)^N\right) \xrightarrow{f_N} \mathbb{C}. \tag{8.5.33}$$

This is the prototype of what we will call a cylindrical functions. We make a slightly more formal definition.

Definition 8.5.21 — Cylindrical function. Let γ_N be a graph $\gamma_N = \{e_1, \ldots, e_N\}$ with corresponding holonomy map $h_{\gamma_N} : \mathcal{A} \to SU(2)^N$. A function $f_{\gamma_N} : \mathcal{A} \to \mathbb{C}$ is called a cylindrical function with respect to γ_N if there exists a function $f_N : SU(2)^N \to \mathbb{C}$ such that $f_{\gamma_N} = f_N \circ h_{\gamma_N}$.

An important aspect of this definition is that one and the same function can be cylindrical with respect to many different graphs. We denote the space of cylindrical functions with respect to γ as Cyl_γ. The space of all cylindrical functions is written as Cyl.

Definition 8.5.22 — Cyl. The space of functions Cyl is defined as

$$\mathrm{Cyl} = \bigcup_{\gamma \in \Gamma} \mathrm{Cyl}_\gamma \tag{8.5.34}$$

We write $\overline{\mathrm{Cyl}}$ for the Cauchy completion of Cyl.

The space of generalized connections

The set $\overline{\mathrm{Cyl}}$ defines a function algebra with respect to the usual point additions and multiplications in the complex numbers. There is also a natural involution given by complex conjugation. Furthermore, on a compact group, the cylindrical functions will be bounded functions, and we can define a norm by using the supremum norm on these functions. The unit element of the algebra is given by a constant map to the unit element of the complex numbers. With these stipulations $\overline{\mathrm{Cyl}}$ becomes a C^*-algebra with unit [171]. Since $\overline{\mathrm{Cyl}}$ is unital C^*-algebra it defines (by the Gelfand-Neimark functor) a compact Hausdorff topological space $\bar{\mathcal{A}}$, on which it is the set of all continuous functions $C(\bar{\mathcal{A}}, \mathbb{C})$ [171]. The space $\bar{\mathcal{A}}$ can be understood as the space which extend the space $\mathcal{A}$ of continuous everywhere supported connections, to also include (non-continuous) connections that are supported just on a given graph. Such connection are not the continuous connections one ordinarily considers in gauge theory. This space is identical to the space we defined before as $\mathrm{Hom}(\mathcal{P}, SU(2))$.

Alternative Cyl construction

In the previous subsection we have constructed the space Cyl (and $\overline{\mathrm{Cyl}}$) in a direct and straightforward manner. There is also a slightly different approach to this construction that is interesting (see [25][172][169][173] for further details and proofs). We use some facts we have already established. We repeat them here for clarity. The spaces X_l are defined as before, and likewise the projections $P_{l_1 l_2}$. Recall that $P_{l_1 l_2}$ is a projection from X_{l_1} to X_{l_2}. Given a function f on X_{l_2} we can use a pullback construction to define a (partial)function from X_{l_1}. We denote this (partial) pullback function by $P^*_{l_1 l_2} X_{l_2}$. Let $n = |E(\gamma)|$ be the cardinality of the set of edges of γ, and let $l = l(\gamma)$. As we have mentioned there is a bijection between each of the space X_l and $(\mathrm{SU}(2))^n$. On each of the spaces X_l one can define the set of continuous complex functions $C(X_l)$. Le us define two functions $f_{l_1} \in C(X_{l_1})$ and $f_{l_2} \in C(X_{l_2})$ to be equivalent if for some l_3 it is true that

$$P_{l_3 l_1} f_{l_1} = P_{l_3 l_2} f_{l_2}. \tag{8.5.35}$$

This equivalence relation will be denoted by $\sim$. We define the space

$$\mathrm{Cyl}(\overline{X}) = \left[\bigcup_{l \in \mathcal{L}} C(X_l) \right] / \sim . \tag{8.5.36}$$

We can define the operations of addition, multiplication, scalar multiplication, and complex conjugation, on the equivalences class found in $\mathrm{Cyl}(\overline{X})$ by using the point based operation in the

codomain of the functions. This makes $\mathrm{Cyl}(\overline{X})$ a *-algebra. The supremum norm is also well defined on $\mathrm{Cyl}(\overline{X})$. Taking the Cauchy completion $\overline{\mathrm{Cyl}(\overline{X})}$ produces an Abelian C*-algebra. By Gelfand-Naimark the spectrum $M(\overline{\mathrm{Cyl}(\overline{X})})$ is a compact Hausdorff space. One can show that this spectrum is homeomorphic to $\overline{\mathcal{A}}$. We established previously that the space $\overline{\mathcal{A}}$ is homeomorphic to the space $\overline{X}$ that was defined by projective methods, and thus we see that $M(\overline{\mathrm{Cyl}(\overline{X})})$ is homeomorphic to $\overline{X}$. Furthermore, by Gelfand-Naimark, $C(M(\overline{\mathrm{Cyl}(\overline{X})}))$ is isomorphic to $\overline{\mathrm{Cyl}(\overline{X})}$. Since we can write $C(M(\overline{\mathrm{Cyl}(\overline{X})})) = C(\overline{X})$ we justify the notation $\mathrm{Cyl}(\overline{X})$, which suggests that it is a set of functions on $\overline{X}$, even though, by the definition they are not functions on $\overline{X}$[4]. To connect to the definitions in the previous subsection we equate $X_{l(\gamma)}$ with Cyl_γ. We assume without further ado that this construction results in the same algebra as the one defined in the previous subsection.

Modified Poisson bracket approach

Because of the use of smearings in less that three dimensions our construction so far the Poisson bracket is not yet completely satisfactory. We will not go into much detail on how to fix this, but we give a brief outline (see [25] for the complete story). To define a proper Poisson bracket on a set of suitable functions we need to modify the standard phase space Poisson bracket setup somewhat[25]. Let us consider a phase space Ph that has the structure of a cotangent bundle over M. In the usual formulation the algebra of observables consists of functions on T^*M. One selects the functions analogous to p and q as the basic variables and develops the Poisson bracket from these. In the alternative formulation we choose a preferred set of configuration variables, i.e. a preferred set of functions $Func(M)$ just on the configuration space $C = M$. We then choose a set of preferred vector fields $V(M)$ on $C = M$ whose action on functions leaves the space $Func(M)$ invariant. The space of these functions and vector fields, is automatically endowed with a natural Lie bracket given by

$$\{\ ,\ \}\colon\ ((f_1, v_1), (f_2, v_2)) \mapsto (v_1(f_2) - v_2(f_1), [v_1, v_2])\,. \tag{8.5.37}$$

Here $v(f)$ is the action of the vector fields on functions and $[v_1, v_2]$ is the vector field Lie bracket.

This approach is the one we will use to define the classical algebra. The cylindrical function have already been defined and they will serve as the preferred functions. The cylindrical function will operate on the configuration space $\overline{\mathcal{A}}$. The vector fields on $\overline{\mathcal{A}}$ will be the Hamiltonian vector field corresponding to the fluxes introduces earlier. They will be denoted $Y(F_{S,f})$. We call the set of such smooth vector fields $V^\infty(\overline{\mathcal{A}})$. The Poisson bracket will be defined on $Cyl^\infty(\overline{\mathcal{A}}) \times V^\infty(\overline{\mathcal{A}})$. It becomes a map [25]

$$\{\ ,\ \}\colon\ \big(Cyl^\infty(\overline{\mathcal{A}}) \times V^\infty(\overline{\mathcal{A}}\big) \times \big(Cyl^\infty(\overline{\mathcal{A}}) \times V^\infty(\overline{\mathcal{A}}\big) \to \big(Cyl^\infty(\overline{\mathcal{A}}) \times V^\infty(\overline{\mathcal{A}}\big)\,. \tag{8.5.38}$$

The action of the flux vector fields on the cylindrical functions are the one defined by the action of the flux on the holonomies in equation 8.5.23. This Poisson bracket states the final form of the classical subalgebra $\mathfrak{A}$. This is essentially the holonomy-flux relations already presented put into the modified Poisson bracket framework. We will refer to this algebra as the holonomy-flux algebra $\mathfrak{A}$.

8.6 Quantization of the graph formulation

In the previous section we found the classical algebra of elementary functions $\mathfrak{A}$. This algebra will be the definition of the classical theory to be quantized. The corresponding configuration space is denoted $\overline{A}$. To proceed with quantization the abstract quantum algebra $\mathfrak{B}$ must be constructed. This

[4] This set is actually the union of functions on the various X_l's

algebra implements the Poisson Lie algebra structure in the form of commutators. The construction is described in detailed in [25][172][169]. We will refer to this algebra as the quantum holonomy-flux algebra. Once we have the algebra $\mathfrak{B}$ quantization consist of constructing a representation of this algebra on some Hilbert space. The main tool to find such representations is the GNS construction (see the appendix)(see section 8.13 - subsection GNS construction for the details of the proper use of the GNS construction). To use the GNS construction we need to define a linear functional ω on $\mathfrak{B}$. This linear functional ω will first be defined on $\overline{Cyl(\overline{X})}$ and then transported to $\mathfrak{B}$. By the representation theorem ω is equivalent to a measure on $\overline{X}$, and this is how we will define ω. To see how we transport ω to $\mathfrak{B}$ we first claim without proof that elements of $\mathfrak{B}$ are of the form

$$f \tag{8.6.1}$$
$$f \cdot Y_1 \tag{8.6.2}$$
$$f \cdot Y_1 \dots Y_k, \tag{8.6.3}$$

where $(f, 0) \in Cyl^\infty(\overline{\mathcal{A}}) \times V^\infty(\overline{\mathcal{A}})$ and $(0, Y_i) \in Cyl^\infty(\overline{\mathcal{A}}) \times V^\infty(\overline{\mathcal{A}})$. We define ω on such elements of $\mathcal{B}$ by

$$\omega(f) = \omega(f) \tag{8.6.4}$$
$$\omega(f \cdot Y_1) = 0 \tag{8.6.5}$$
$$\omega(f \cdot Y_1 \dots Y_k) = 0. \tag{8.6.6}$$

The GNS construction will in principle lead to multiple inequivalent representations (see[25]). To narrow the selection further demands are set on the representation. The conclusion for the algebra $\mathfrak{B}$ is that the representation we are about to present (or equivalently the measure we are about to present) is the unique possibility given some requirements that we state in a theorem.

Theorem 8.6.1 — Unique state theorem. There exist exactly one cyclic gauge-invariant and diffeomorphism invariant representation of the quantum holonomy-flux algebra. Moreover this representation is irreducible.

Proof. See [25] for a proof. ■

We will not present any of the details of the GNS analysis. We will instead start by defining a measure on cylindrical functions. The measure is called the Ashtekar-Lewandowski (AL) measure. We will then proceed by presenting the Hilbert space and a suitable basis. Finally we define the action of the Hilbert space operators corresponding to fluxes and holonomies.

Measure and Hilbert space

There is a natural way to define cylindrical measure for cylindrical functions over a fixed graph. We will define this cylindrical measure on Cyl_γ (or $C(X_{l(\gamma)})$) as an extension of the Haar measure on $SU(2)$. After defining this for an individual graph this will be generalized to the space Cyl.

Definition 8.6.2 — Cylindrical measure. Let $\gamma = \{e_1, \dots, e_N\}$, and let $f = f_\gamma \circ h_\gamma$ be a function in Cyl_γ. The integral of f is defined by

$$\int d\mu_{AL} f = \int_{SU(2)^N} d\mu^1_{Haar} \dots d\mu^N_{Haar} f_\gamma(g_1, \dots, g_N). \tag{8.6.7}$$

We will often refer to $d\mu_{AL}$ just as d_{mu}. Note that the integration is not defined on the domain of f but on the domain of f_γ. Now that the measure is define we can define $L^2(SU(2)^N, d\mu_{AL})$ (or $L^2(X_{l(\gamma)}, d\mu_{AL})$ and we can equip it with an inner product.

Definition 8.6.3 — Cylindrical inner product. Let $\gamma = \{e_1, \dots, e_N\}$, and let $f = f_\gamma \circ h_\gamma$ and $g = g_\gamma \circ h_\gamma$ be two functions in $L^2(SU(2)^N, d\mu_{AL})$. We define the inner product of f and g as

$$\langle f|g\rangle = \int d\mu_{AL} \overline{f} g = \int_{SU(2)^N} d\mu_{Haar} \dots d\mu_{Haar} \overline{f_\gamma(g_1, \dots, g_N)} g_\gamma(g_1, \dots, g_N). \quad (8.6.8)$$

This makes $L^2(SU(2)^N, d\mu_{AL})$ a Hilbert space. We have thus defined cylindrical functions, integration and inner product for a specific graph γ. It can be shown that the measure and the inner product is still well defined when we extend it to $\overline{\mathcal{A}}$ (or $\overline{X}$). An important part of this is to show that no matter which graphs is used (a cylindrical function is cylindrical with resepect to many graphs), we get the same value for the integral. We skip the detailed development for now (see [25][172][169][173] for further details).

Hilbert space basis

We have define the Hilbert space of LQG to be $L^2(\bar{\mathcal{A}}, d\mu)$. The next step is to find a suitable basis. The basis will constructed by defining a basis for each subspace $H_\gamma = L^2(SU(2)^N, d\mu_{AL})$. We will start with graphs with just a single edge.

According to the Peter-Weyl theorem any continuous functions on G can be expressed as a sum of functions mapping each group element to a matrix element of some irreducible representations of the group [158][174]. These functions form a basis for the set of arbitrary continuous functions[5] $F_1 \colon G \to \mathbb{R}$. We write a basis functions as

$$D^{(j_1)}_{\alpha\beta} \colon g \mapsto D^{(j_1)}_{\alpha\beta}(g). \quad (8.6.9)$$

Here $D^{(j_1)}$ is the j_1 matrix representation of $\mathrm{SU}(2)$, and α and β are the indices for the matrix element. This leads to expansions of arbitrary functions as

$$F_1(g) = \sum_{j,\alpha,\beta} f^j_{\alpha\beta} D^{(j)}_{\alpha\beta}(g). \quad (8.6.10)$$

Here $f^j_{\alpha\beta}$ is a set of complex coefficients. For functions of the type $F_N \colon G^N \to \mathbb{R}$ we can create basis functions by defining

$$D^{(j_1)}_{\alpha_1\beta_1} \dots D^{(j_N)}_{\alpha_N\beta_N} \colon (g_1, \dots, g_N) \mapsto D^{(j_1)}_{\alpha_1\beta_1}(g_1) \dots D^{(j_N)}_{\alpha_N\beta_N}(g_N). \quad (8.6.11)$$

These basis functions can be used to express arbitrary functions F_N in the form

$$F_N(g_1, \dots, g_N) = \sum_{j,\alpha,\beta} f^{j_1 \dots j_N}_{\alpha_1\beta_1 \dots \alpha_N\beta_N} D^{(j_1)}_{\alpha_1\beta_1}(g_1) \dots D^{(j_N)}_{\alpha_N\beta_N}(g_N). \quad (8.6.12)$$

Note that $f^{j_1 \dots j_N}_{\alpha_1\beta_1 \dots \alpha_N\beta_N}$ are numbers not functions. We return to the cylindrical functions which we now can write using the Peter-Weyl basis for f_N functions. We will also start using ket notation for some of these functions. Let γ be a graph with N edges $(e_1, \dots, e_N)$, and let h_{e_i} be the associated holonomies. A cylindrical basis function is written as

$$\begin{aligned} f_{\gamma_N}[\mathcal{A}] &= f_N \circ h_{\gamma_N}[\mathcal{A}] && (8.6.13)\\ &= D^{(j_1)}_{\alpha_1\beta_1}(h_{e_1}[\mathcal{A}]) \dots D^{(j_N)}_{\alpha_N\beta_N}(h_{e_N}[\mathcal{A}]) && (8.6.14)\\ &= |\gamma_N, j_1, \dots, j_N, \alpha_1, \dots, \alpha_N, \beta_1, \dots \beta_N\rangle && (8.6.15)\\ &= |\gamma_N, j_i, \alpha_i, \beta_i\rangle\,. && (8.6.16) \end{aligned}$$

[5] The subscript 1 on F_1 refers to the number of group arguments that the function takes.

The last line introduces a shorthand notation for the same object as the line above[6]. An arbitrary superposition is written as

$$F_{\gamma_N}[\mathcal{A}] = F_N \circ h_{\gamma_N}[\mathcal{A}] \tag{8.6.17}$$

$$= \sum_{j,\alpha,\beta} f^{j_1\dots j_N}_{\alpha_1\beta_1\dots\alpha_N\beta_N} D^{(j_1)}_{\alpha_1\beta_1}(h_{e_1}[\mathcal{A}]) \dots D^{(j_N)}_{\alpha_N\beta_N}(h_{e_N}[\mathcal{A}]) \tag{8.6.18}$$

$$= \sum_{j,\alpha,\beta} f^{j_1\dots j_N}_{\alpha_1\beta_1\dots\alpha_N\beta_N} |\gamma_N, j_1, \dots, j_N, \alpha_1, \dots, \alpha_N, \beta_1, \dots \beta_N\rangle \tag{8.6.19}$$

$$= \sum_{j_i,\alpha_i,\beta_i} f^{j_1\dots j_N}_{\alpha_1\beta_1\dots\alpha_N\beta_N} |\gamma_N, j_i, \alpha_i, \beta_i\rangle\,. \tag{8.6.20}$$

The last line introduces a shorthand notation for the same object as the line above. Note that the γ part of $|\gamma_N, j_i, \alpha_i, \beta_i\rangle$ specifies the map h_γ, while the (j, α, β) part specifies the function $D^{(j_N)}_{\alpha_N\beta_N}$. The function F_N is specified (implicitly) by the superposition $\sum_{j,\alpha,\beta} f^{j_1\dots j_N}_{\alpha_1\beta_1\dots\alpha_N\beta_N}$. See figure 8.8 for an example of a cylindrical basis function represented in graphical form.

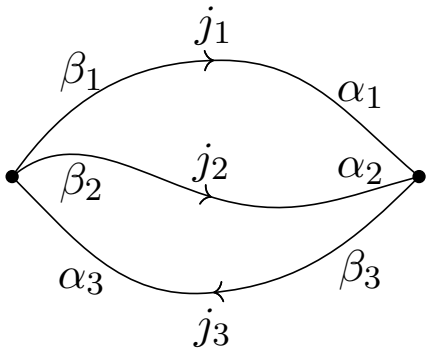


Figure 8.8: The diagram shows the quantum state $|\gamma_3, j_1, j_1, j_1, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3\rangle$. The symbol γ_3 refers to the graph shown.

Action of holonomy and flux

A central process of quantization is to express the action of the algebra elements on the Hilbert space. The holonomy components have a natural multiplicative action on the Cyl basis states [174]. We define the action of the operator $\hat{h}_e[\mathcal{A}]_{\alpha\beta}$ on a function $f \in Cyl$ as

$$\left(\hat{h}_e[\mathcal{A}]_{\alpha\beta} f\right)(g_1, \dots, g_N) = h_e[\mathcal{A}]_{\alpha\beta} D^{(j_1)}_{\alpha_1\beta_1}(h_{e_1}[\mathcal{A}]) \dots D^{(j_N)}_{\alpha_N\beta_N}(h_{e_N}[\mathcal{A}]) \tag{8.6.21}$$

This multiplicative action if of course nothing else than multiplying by a matrix element in the fundamental representations $j = 1/2$. In terms of the edges of the graph γ, this means we are adding an extra edge e to the graph γ. We can make this more clear by rewriting the action as

$$\left(\hat{h}_e[\mathcal{A}]_{\alpha\beta} f\right)(g_1, \dots, g_N) = D^{(j=1/2)}_{\alpha\beta}(h_e[\mathcal{A}]) D^{(j_1)}_{\alpha_1\beta_1}(h_{e_1}[\mathcal{A}]) \dots D^{(j_N)}_{\alpha_N\beta_N}(h_{e_N}[\mathcal{A}]). \tag{8.6.22}$$

The effect of the flux on a cylindrical state will be exemplified by the its action on a holonomy state

$$\hat{F}_{S,f} h_e[\mathcal{A}] = \pm\beta\kappa f^i(p) h_{e_1}[A]\, \tau_i h_{e_2}[A] \tag{8.6.23}$$

We have assumed that the surface S intersects the edge e once and only once, and divides e into the two edges e_1 and e_2. We see that $\hat{F}$ act as the derivate operator $\frac{\delta}{\delta A}$ at the intersection point of S and e as we would expect in a canonical quantization setup [174].

[6]Note that $D^{(j)}_{\alpha\beta}(h_e[\mathcal{A}]) = h^{(j)}_e[\mathcal{A}]_{\alpha\beta}$.

8.7 Implementing constraints

After having constructed an unconstrained Hilbert space $\mathcal{H}^{Kin}$ of square integrable cylindrical functions, we now proceed to implement the constraints [174][158]. The techniques for implementing first class constraints form a major part of refined algebraic quantization (RAQ). The technique we will be using goes by the name of group averaging. The group in "group averaging" is the Lie group generated by the Lie algebra of the first class constraints. Let us denote this group by G and refer to elements of G as g. In the classical theory the constraints are given as equations of the form

$$\phi_i = 0. \tag{8.7.1}$$

The solution of these equations specifies a submanifold of the phase space called the constraint surface. In the quantum theory we expect the constraints to be represented as

$$\hat{\phi}_i \left|\psi\right\rangle = 0. \tag{8.7.2}$$

These equations defines a subspace $\mathcal{H}^{Phys}$ of $\mathcal{H}^{Kin}$. The equation $\hat{\phi}_i \left|\psi\right\rangle = 0$ represent a given constraint directly by a corresponding operator. This direct quantum representation of the classical constraint is not always convenient. Instead of realizing the constraint ϕ_i directly as the operator $\hat{\phi}_i$ we will define the constraint by using a unitary representation of the associated group G. We define a unitary representation of G on $\mathcal{H}^{Kin}$ by the map

$$U \colon G \ni g \mapsto U(g) \in L(\mathcal{H}^{Kin}). \tag{8.7.3}$$

We then require that physical states $\left|\psi\right\rangle_{inv}$ belongs to the trivial subrepresentation given by

$$U(g) \left|\psi\right\rangle_{inv} = \left|\psi\right\rangle_{inv} \quad \forall g \in G. \tag{8.7.4}$$

The basic idea is to construct invariant states by averaging over the group. We define a set of invariant states by writing

$$\left|\psi\right\rangle_{inv} = \int \mathrm{d}\mu(g)\ U(g) \left|\psi\right\rangle. \tag{8.7.5}$$

The measure $\mathrm{d}\mu(g)$ is the Haar measure for G. All states of this form are invariant under $U(g)$. To show that this state is indeed invariant, we use the properties of the Haar measure and find that

$$U(g') \left|\psi\right\rangle_{inv} = \int \mathrm{d}\mu(g)\ U(g')U(g) \left|\psi\right\rangle = \int \mathrm{d}\mu(g)\ U(g'g) \left|\psi\right\rangle = \left|\psi\right\rangle_{inv}. \tag{8.7.6}$$

The procedure just given applies without complications if G is a compact group with 1 in the discrete part of the spectrum. An example of this can be seen in the implementation of the Gauss constraint. However, we also need to deal with non-compact groups where 1 is in the *continuous* part of the spectrum. An example of this can be seen in the implementation of the diffeomorphism constraint. The problem in the case of non-compact groups is that the solutions of equation 8.7.4 does not lie in the space $\mathcal{H}^{Kin}$. This is similar to the problem with the unbounded operators $\hat{q}$ and $\hat{p}$ in ordinary one-dimensional quantum mechanics. The eigenvectors of $\hat{q}$ and $\hat{p}$ are not in $L^2(\mathbb{R})$. The solution to this problem is to construct a Gelfand triple (also know as a rigged Hilbert space). This requires the introduction of a dense subset $\mathcal{D}$ of $\mathcal{H}^{Kin}$ and the dual space $\mathcal{D}^*$. We will not go through the construction for the non-compact case in full detail. A more complete exposition can be found in e.g. [152].

Gauss constraint

The first constraint we implement is the Gauss constraint. We start by studying how gauge transformations affect the kinematical states. Previously we saw that a gauge transformation only affects the holonomy at the endpoints. We also want to know the effect of gauge transformations on holonomy component functions. Let Λ be a map $\Lambda : \sigma \to G$ and let $U(\Lambda)$ be a unitary representation of Λ. We now consider the graph γ_1 shown in figure 8.9. For this case we can write the effect if a gauge transformation explicitly as

$$U(\Lambda)\,|\gamma_1, j, \alpha, \beta\rangle = U(\Lambda) D^{(j)}_{\alpha\beta}\left(h_e\left[\mathcal{A}\right]\right) \tag{8.7.7}$$

$$= D^{(j)}_{\alpha\beta}\left(\Lambda(b(e))\; h_e\left[\mathcal{A}\right]\; \Lambda^{-1}(f(e))\right) \tag{8.7.8}$$

$$= D^{(j)}_{\alpha\beta}\left(\Lambda(b(e))\right) D^{(j)}_{\alpha\beta}\left(h_e\left[\mathcal{A}\right]\right) D^{(j)}_{\alpha\beta}\left(\Lambda^{-1}(f(e))\right) \tag{8.7.9}$$

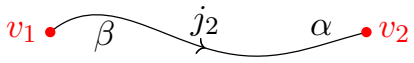


Figure 8.9: Example quantum state based on a graph with two nodes (v_1 and v_2) and a single edge (e_1). The graph itself is denoted γ_1. The symbols α and β represents matrix component indices. The half-integer j_2 indicated the $\mathrm{SU}(2)$ representation.

Let us simplify the notation by writing $\Lambda(b(e))$ as g_{v_1}, and $\Lambda^{-1}(f(e))$ as $g^{-1}_{v_2}$, where v_1 and v_2 refers to vertex 1 and vertex 2 (see figure 8.9). To find a combination of basis functions that is invariant under gauge transformation we employ the technique of group averaging.

$$|\psi\rangle = \int U(\Lambda)\,|\gamma_1, j, \alpha, \beta\rangle \tag{8.7.10}$$

$$= \int dg_{v_1}\; D^{(j)}_{\alpha\beta}\left(g_{v_1}\; h_e\left[\mathcal{A}\right]\; g^{-1}_{v_1}\right) \tag{8.7.11}$$

$$= \int dg_{v_1}\; D^{(j)}_{\alpha m}(g_{v_1}) D^{(j)}_{mn}\left(\; h_e\left[\mathcal{A}\right]\right) D^{(j)}_{n\beta}(g^{-1}_{v_1}) \tag{8.7.12}$$

$$= D^{(j)}_{mn}\left(\; h_e\left[\mathcal{A}\right]\right) \int dg_{v_1}\; D^{(j)}_{\alpha m}(g_{v_1}) D^{(j)}_{n\beta}(g^{-1}_{v_1}) \tag{8.7.13}$$

The gauge transformation only acts at the nodes of a graph. We can extend the one-edge example we just presented to graphs with more edges and nodes. If there are N-nodes there will be N different group elements involved in the gauge transformation. We average over these effect by integrating over all nodes. The function given by

$$f_0(h_1, \ldots, h_L) = \int dg_1 \ldots dg_N f(g_1 h_1 g_2^{-1}, \ldots, g_{N-1} h_L g_N^{-1}), \tag{8.7.14}$$

is invariant under gauge transformations at the nodes [174]. The integration goes over all distinct nodes. Note that two or more edges can share nodes at one or both ends. This integral can be used to find an explicit formula for the coefficient of the expansion in basis functions. For the example in figure 8.10 we can make all this explicit. The state is given as

$$|\psi\rangle = \sum f^{j_1 j_2 j_3}_{\alpha_1\beta_1\alpha_2\beta_2\alpha_3\beta_3}\,|\gamma_3, j_1, j_2, j_3, \alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3\rangle \tag{8.7.15}$$

$$= \sum f^{j_1 j_2 j_3}_{\alpha_1\beta_1\alpha_2\beta_2\alpha_3\beta_3} D^{(j_1)}_{\alpha_1\beta_1}(h_{e_1}\left[A\right])\; D^{(j_2)}_{\alpha_2\beta_2}(h_{e_2}\left[A\right])\; D^{(j_3)}_{\alpha_3\beta_3}(h_{e_3}\left[A\right]). \tag{8.7.16}$$

Writing h_{e_1} for $h_{e_1}\left[A\right]$ and suppressing all indices we introduce the shorthand notation

$$|\xi\rangle = f(h_{e_1}, h_{e_2}, h_{e_3}) = D^{(j_1)}_{\alpha_1\beta_1}(h_{e_1}\left[A\right])\; D^{(j_2)}_{\alpha_2\beta_2}(h_{e_2}\left[A\right])\; D^{(j_3)}_{\alpha_3\beta_3}(h_{e_3}\left[A\right]). \tag{8.7.17}$$

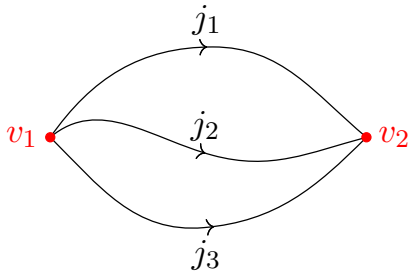


Figure 8.10: Example quantum state with two nodes and three edges.

Using this shorthand notation we can express the effect of a gauge transformation as

$$U(g)\,|\xi\rangle = f(g_{v_1} h_{e_1} g_{v_2}^{-1}, g_{v_1} h_{e_2} g_{v_2}^{-1}, g_{v_1} h_{e_3} g_{v_2}^{-1}). \tag{8.7.18}$$

We construct the invariant state by the group averaging method as

$$|\xi\rangle_{inv} = \int \mathrm{d}\mu(\Lambda)\; U(\Lambda)\,|\gamma_3, j_1, j_2, j_3\rangle \tag{8.7.19}$$

$$= \int \mathrm{d}g_{v_1}\,\mathrm{d}g_{v_1}\; f(g_{v_1} h_{e_1} g_{v_2}^{-1}, g_{v_1} h_{e_2} g_{v_2}^{-1}, g_{v_1} h_{e_3} g_{v_2}^{-1}) \tag{8.7.20}$$

$$\begin{aligned} = \int \mathrm{d}g_{v_1}\,\mathrm{d}g_{v_1}\; & D^{(j_1)}_{\alpha_1\beta_1}(g_{v_1}\, h_{e_1}[A]\, g_{v_2}^{-1}) \\ & D^{(j_2)}_{\alpha_2\beta_2}(g_{v_1} h_{e_2}[A]\, g_{v_2}^{-1}) \\ & D^{(j_3)}_{\alpha_3\beta_3}(g_{v_1} h_{e_3}[A]\, g_{v_2}^{-1}). \end{aligned} \tag{8.7.21}$$

On the full state we get

$$\begin{aligned} |\psi\rangle_{inv} = \int \mathrm{d}g_{v_1}\,\mathrm{d}g_{v_2} \sum & f^{j_1 j_2 j_3}_{\alpha_1\beta_1\alpha_2\beta_2\alpha_3\beta_3} \\ & D^{(j_1)}_{\alpha_1\beta_1}(g_{v_1}\, h_{e_1}[A]\, g_{v_2}^{-1}) \\ & D^{(j_2)}_{\alpha_2\beta_2}(g_{v_1} h_{e_2}[A]\, g_{v_2}^{-1}) \\ & D^{(j_3)}_{\alpha_3\beta_3}(g_{v_1} h_{e_3}[A]\, g_{v_2}^{-1}). \end{aligned} \tag{8.7.22}$$

Since D is a group homomorphism we can write this as

$$\begin{aligned} = \int \mathrm{d}g_{v_1}\,\mathrm{d}g_{v_2} \sum & f^{j_1 j_2 j_3}_{\alpha_1\beta_1\alpha_2\beta_2\alpha_3\beta_3} \\ & D^{(j_1)}_{\alpha_1 m_1}(g_{v_1}) D^{(j_1)}_{m_1 n_1}(h_{e_1}[A]) D^{(j_1)}_{n_1\beta_1}(g_{v_2}^{-1}) \\ & D^{(j_2)}_{\alpha_2 m_2}(g_{v_1}) D^{(j_2)}_{m_2 n_2}(h_{e_2}[A]) D^{(j_2)}_{n_2\beta_2}(g_{v_2}^{-1}) \\ & D^{(j_3)}_{\alpha_3 m_3}(g_{v_1}) D^{(j_3)}_{m_3 n_3}(h_{e_3}[A]) D^{(j_3)}_{n_3\beta_3}(g_{v_2}^{-1}) \end{aligned} \tag{8.7.23}$$

$$\begin{aligned} = \sum & f^{j_1 j_2 j_3}_{\alpha_1\beta_1\alpha_2\beta_2\alpha_3\beta_3} \\ & D^{(j_1)}_{m_1 n_1}(h_{e_1}[A]) D^{(j_2)}_{m_2 n_2}(h_{e_2}[A]) D^{(j_3)}_{m_3 n_3}(h_{e_3}[A]) \\ & \int \mathrm{d}g_{v_1}\; D^{(j_1)}_{\alpha_1 m_1}(g_{v_1}) D^{(j_2)}_{\alpha_2 m_2}(g_{v_1}) D^{(j_3)}_{\alpha_3 m_3}(g_{v_1}) \\ & \int \mathrm{d}g_{v_2}\; D^{(j_1)}_{n_1\beta_1}(g_{v_2}^{-1}) D^{(j_2)}_{n_2\beta_2}(g_{v_2}^{-1}) D^{(j_3)}_{n_3\beta_3}(g_{v_2}^{-1}) \end{aligned} \tag{8.7.24}$$

$$= \sum f^{j_1 j_2 j_3}_{\alpha_1\beta_1\alpha_2\beta_2\alpha_3\beta_3} P^{\alpha_1\alpha_2\alpha_3}_{m_1 m_2 m_3} P^{n_1 n_2 n_3}_{\beta_1\beta_2\beta_3} D^{(j_1)}_{m_1 n_1}(h_{e_1}[A]) D^{(j_2)}_{m_2 n_2}(h_{e_2}[A]) D^{(j_3)}_{m_3 n_3}(h_{e_3}[A]). \tag{8.7.25}$$

In the last equation we have lifted half of all the indices to confirm to conventional matrix notation, which would result from writing the matrices we have used as $D^{(j_1)\ \alpha_1}_{\ \ m_1}(g_{v_1})$. We have defined a new projector symbol given by

$$P^{\alpha_1\alpha_2\alpha_3}_{m_1 m_2 m_3} = \int \mathrm{d}g_{v_1}\, D^{(j_1)}_{\alpha_1 m_1}(g_{v_1}) D^{(j_2)}_{\alpha_2 m_2}(g_{v_1}) D^{(j_3)}_{\alpha_3 m_3}(g_{v_1}). \tag{8.7.26}$$

In turns out that this projector is related to the $3j$-symbols

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ j_1 & j_2 & j_3 \end{pmatrix} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \tag{8.7.27}$$

by

$$\begin{aligned} P^{\alpha_1\alpha_2\alpha_3}_{m_1 m_2 m_3} &= \iota^{\alpha_1\alpha_2\alpha_3} \iota_{m_1 m_2 m_3} \\ &= \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ j_1 & j_2 & j_3 \end{pmatrix} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \end{aligned} \tag{8.7.28}$$

We can now write the complete invariant states as (see figure 8.11)

$$|\psi\rangle_{inv} = \sum f^{j_1 j_2 j_3}_{\alpha_1\beta_1\alpha_2\beta_2\alpha_3\beta_3} \tag{8.7.29}$$

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ j_1 & j_2 & j_3 \end{pmatrix} \begin{pmatrix} j_1 & j_2 & j_3 \\ \beta_1 & \beta_2 & \beta_3 \end{pmatrix} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} n_1 & n_2 & n_3 \\ j_1 & j_2 & j_3 \end{pmatrix} \tag{8.7.30}$$

$$D^{(j_1)}_{m_1 n_1}(h_{e_1}[A]) D^{(j_2)}_{m_2 n_2}(h_{e_2}[A]) D^{(j_3)}_{m_3 n_3}(h_{e_3}[A]) \tag{8.7.31}$$

$$= \sum f^{j_1 j_2 j_3}_{\text{contracted}} \iota_{v_1} \iota_{v_2} D^{(j_1)}(h_{e_1}[A]) D^{(j_2)}(h_{e_2}[A]) D^{(j_3)}(h_{e_3}[A]). \tag{8.7.32}$$

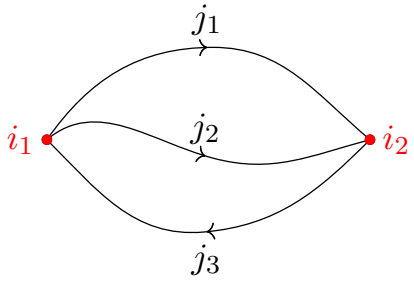


Figure 8.11: The diagram shows spin graph with intertwiners.

Where $\mathrm{contracted}$ represents contracted indices. We thus find that the invariant states are made by inserting intertwiners at each vertex and contracting them with the representation matrices. Such states are called spin-network states. We write a general spin-network states as

$$|\gamma, j_k, i_l\rangle = \sum f \prod D^{j_k} \prod \iota_l. \tag{8.7.33}$$

Area and volume operator

Before moving on to the next constraints we present the area and the volume operators [174][77]. These operators provide a method to interpret the spin-network states as quantum geometric states. The operators are diagonal in the spin-network basis and have discrete eigenvalues as we now will see. Let as surface S in σ with coordinates (x, y) be given, and let q be the pullback metric. The classical area is given as

$$A(S) = \int_S dxdy\sqrt{\det(q)}. \tag{8.7.34}$$

This can be expressed using triads E^a_i and surface normal form n_a as

$$A(S) = \int_S dxdy\sqrt{E^a_i E^b_j \delta^{ij} n_a n_b}. \tag{8.7.35}$$

If we use triads smeared on a small surfaces S_I we can express the area operator in terms of densitized triad operators (we can disregard the smearing function) $E(S_I)$ as

$$\hat{A}(S) = lim_{N\to\infty} \sum_I \sqrt{E(S_I)E(S_I)}. \tag{8.7.36}$$

We assume that we apply this operator to a state that pierces each surface S_I at most one time. Let us consider the action when there is only one such piercing site. The action of the flux operator on a holonomy state (in the j-th representation) is

$$\hat{E}h^{(j)}_e[\mathcal{A}] = \beta\kappa h^{(j)}_{e_1}[\mathcal{A}]\ \tau^{(j)}\ h^{(j)}_{e_2}[\mathcal{A}] \tag{8.7.37}$$

If we analyze the action of two such operator on a holonomy state we find

$$\begin{aligned}
\hat{E}\hat{E}h^{(j)}_e[\mathcal{A}] &= (\beta\kappa)^2 h^{(j)}_{e_1}[\mathcal{A}]\ \tau^{(j)}\tau^{(j)}\ h^{(j)}_{e_2}[\mathcal{A}] && (8.7.38)\\
&= (\beta\kappa)^2 h_{e_1}[\mathcal{A}]\ j(j+1)\, h_{e_2}[\mathcal{A}] && (8.7.39)\\
&= (\beta\kappa)^2 j(j+1) h_{e_1}[\mathcal{A}]\ \ h_{e_2}[\mathcal{A}] && (8.7.40)\\
&= (\beta\kappa)^2 j(j+1) h_e[\mathcal{A}] && (8.7.41)
\end{aligned}$$

Hence we conclude that the action of the area operator is given by

$$\hat{A}(S)\left|\psi\right\rangle = \beta\kappa\sqrt{j(j+1)}\left|\psi\right\rangle. \tag{8.7.42}$$

To facilitate the physical interpretation we can express this in more familiar terms as

$$\hat{A}(S)\left|\psi\right\rangle = 8\pi l_p^2\beta\sqrt{j(j+1)}\left|\psi\right\rangle, \tag{8.7.43}$$

where l_p is the Planck length[7]. Thus we see that areas are quantized in units of the Planck length squared (very roughly speaking). The result can be generalized to situations where there are many edges that crosses the surfaces S with very similar results (see e.g. [77]).

The volume operator can be set up in a similar manner using the product of three flux operators. The result are much more complicated, but it turns out that volume is a feature of the nodes of spin-networks, while as we saw, area is associated to the edges of the graph. We write the form of the volume operator for a given region R as

$$V(R) = \int \mathrm{d}^3x\sqrt{\frac{1}{3!}\epsilon_{abc}\epsilon^{ijk}|E^a_i E^b_j E^c_k|}, \tag{8.7.44}$$

We mention the volume operator here as it is a natural compacnion to the area operator. The volume operator is also essential for the implementation of the Hamiltonian constraint. We will not analysis the volume operator any further. A detailed explanation of the volume operator can be found in [25].

[7] The Planck length is defined as $l_p = \sqrt{\frac{G\hbar}{c^3}}$.

Diffeomorphism constraint — S-knots

We will now apply the diffeomorphism constraints to the gauge invariant states constructed above [174][174]. The technique is in principle the same as for the Gauss constraint but in this case the group is not compact. This makes it necessary to use a Gelfand triple. We will not enter into the technical aspects of this at this time. Instead we explain how the group averaging idea works for the diffeomorphism constraint in intuitive terms.

Let us assume that we have a unitary representation of the diffeomorphism group. We now divide the group into subgroups based on the effect on spin-network states. We define Diff_γ to be the set of diffeomorphisms that map γ to itself. We define TDiff_γ to be the normal subgroup of Diff_γ that preserves every edge of γ. The group TDiff_γ is the group of trivial diffeomorphism of γ. The quotient group

$$\mathrm{GS}_\gamma = \frac{\mathrm{Diff}_\gamma}{\mathrm{TDiff}_\gamma}, \tag{8.7.45}$$

is the finite group of graph symmetries of γ. We first average over this group by defining

$$|\psi\rangle_{inv} = \frac{1}{|\mathrm{GS}_\gamma|} \sum_{\phi \in \mathrm{GS}_\gamma} U(\phi) \, |\psi\rangle \, . \tag{8.7.46}$$

Next we consider diffeomorphism that move the graph γ. We define the quotient group

$$\mathrm{MDiff}_\gamma = \frac{\mathrm{Diff}_\gamma}{\mathrm{TDiff}_\gamma}. \tag{8.7.47}$$

The next step requires more techniques to properly define but we go ahead and write down the idea. In the same manner as before we define (in a non-rigorous manner)

$$|\psi\rangle_{inv} = \frac{1}{|\mathrm{MDiff}_\gamma|} \sum_{\phi \in \mathrm{MDiff}_\gamma} U(\phi) \, |\psi\rangle_{inv} \, . \tag{8.7.48}$$

This completes our informal walk through of the handling of the diffeomorphism constraint.

Hamiltonian constraint

The space constructed above represent a rigorously defined kinematical Hilbert space for quantum geometry. The final step, necessary to include dynamics in the theory, is to implement the Hamiltonian constraint. The Hamiltonian constraint is a projection onto allowed kinematical states but it also serves a dual role. In simplified terms the states we get after implementing this constraint are states containing information about correlations between several observables. There is no time parameter, and a relational perspective has to be adopted to extract dynamics. The main thing to understand is that implementing this constraint will in principle bring out the dynamics of the theory, but that dynamics in a background independent setting is something relational which is more subtle than "time-evolution". See section 8.8 for more on relational dynamics.

The smeared Hamiltonian constraint can be written as

$$H(N) = \int d^3x \epsilon^{ij}_k \frac{E^a_i E^b_j}{\sqrt{det(E)}} \left(F^k_{ab} - 2(1+\gamma^2) K^i_a K^j_b \right). \tag{8.7.49}$$

This is a complicated non-polynomial expression which is difficult to translate into an operator with a well defined action on the states. We will handle this by several classical rewrites where we express this constraint using Poisson brackets. Afterward we will rewrite the resulting brackets

using holonomies and fluxes, as well as implement a spatial regularizing procedure to achieve a well defined action on the states.

The first difficulty of the expression is the presence of a denominator in the first term. In addition the denominator involves a square root. We will solve both these issues by the rewrite [175]

$$\frac{E_i^a E_j^b}{\sqrt{det(E)}}\epsilon^{ijk}\epsilon_{abc} = \left\{A_a^k, V\right\}, \tag{8.7.50}$$

where $V = \int \sqrt{det(E)}$ is the volume operator from the previous section. We can rewrite the terms involving the K's like so

$$\bar{K} = \left\{\mathcal{H}^E(1), V\right\} \tag{8.7.51}$$

$$K_a^i = \left\{A_a^i, \bar{K}\right\} \tag{8.7.52}$$

We can now rewrite the two parts of the Hamiltonian constraint as [175]

$$H^E(N) = \int d^3x N\epsilon^{abc}\delta_{ij}F_{ab}^k \left\{A_a^k, V\right\} \tag{8.7.53}$$

$$T(N) = \int d^3x N\epsilon^{ijk}\epsilon_{abc} \left\{A_a^i, \left\{\mathcal{H}^E(1), V\right\}\right\} \left\{A_a^i, \left\{\mathcal{H}^E(1), V\right\}\right\} \left\{A_a^k, V\right\}. \tag{8.7.54}$$

The complete Hamiltonian constraint is written as

$$H(N) = H^E(N) - 2(1+\gamma^2)T(N). \tag{8.7.55}$$

These two parts are referred to as the Euclidean part and the Lorentzian part. We will focus only on the Euclidean part. The Lorentzian part is more complex but follows the same pattern. The Euclidean part is expressed using A, F and V. The next step is to express A, F and V using holonomies. We can expand the holonomy along a line segment along the x^α coordinate. This gives us

$$h_{e_\alpha}[\mathcal{A}] = 1 + \epsilon A_a^i \tau_i + \mathcal{O}(\epsilon^2). \tag{8.7.56}$$

By symmetry we also get

$$h_{e_\alpha}^{-1}[\mathcal{A}] = 1 - \epsilon A_a^i \tau_i + \mathcal{O}(\epsilon^2). \tag{8.7.57}$$

By plain calculation we show that we can write

$$\left\{A_a^k, V\right\} = h_{e_\alpha}^{-1}\left\{h_{e_\alpha}, V\right\} + \mathcal{O}(\epsilon^2). \tag{8.7.58}$$

This takes care of the last part of Euclidean expression. Next we will deal with the curvature part. Let α_{ab} be triangular loop (i.e. a loop in the form of three straight lines forming a triangle). The curvature can be expressed using the holonomy around this loop. We write

$$h_{\alpha_{ab}} = 1 + \frac{1}{2}\epsilon^2 F_{ab}^i \tau^i + \mathcal{O}(\epsilon^4), \tag{8.7.59}$$

and also

$$h_{\alpha_{ab}}^{-1} = 1 - \frac{1}{2}\epsilon^2 F_{ab}^i \tau^i + \mathcal{O}(\epsilon^4). \tag{8.7.60}$$

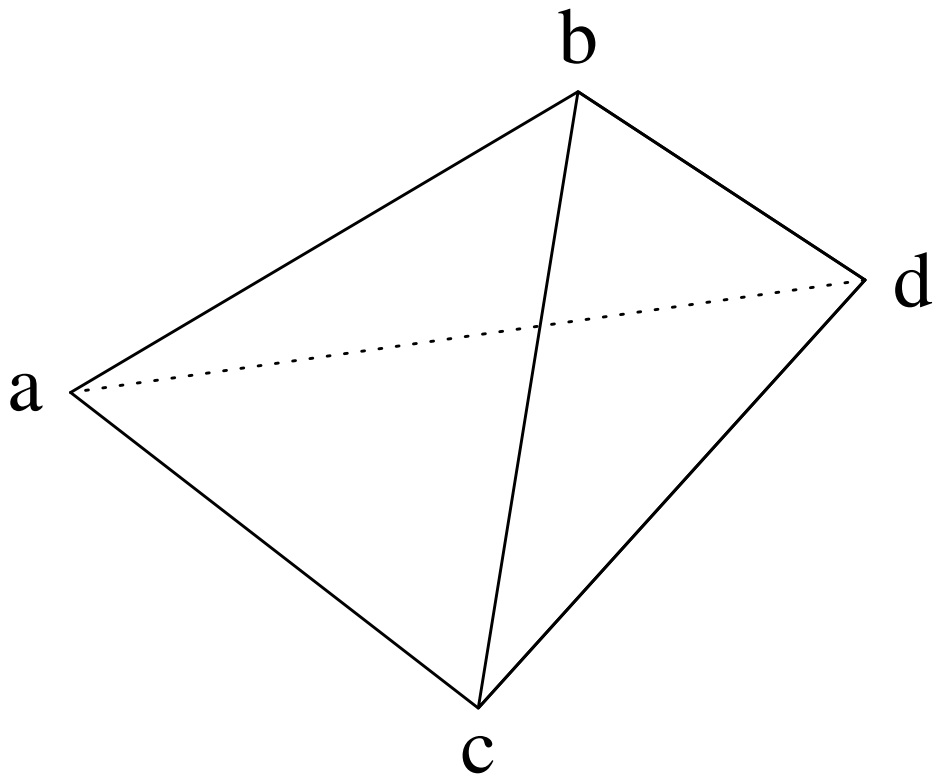


Figure 8.12: Triangulation used for the definition of the Hamiltonian constraint.

This leads to the following expression

$$\frac{1}{2}\epsilon^2 F^i_{ab}\tau^i = (h_{\alpha_{ab}} - h^{-1}_{\alpha_{ab}}) + \mathcal{O}(\epsilon^4). \tag{8.7.61}$$

We now imagine a triangulation of Σ adapted to the holonomy paths we have just used. In figure 8.12 the triangular loop we have been using is around the triangle abc, and the line we used to calculate the holonomy is along the line ad. The (Euclidean) Hamiltonian constraint will then be expressed as these expression define on this triangulation in the limit where the "size" of the triangulation goes to zero.

$$H^E = \lim_{\epsilon\to 0}\sum N_I \epsilon^{abc} Tr\left(\left(h_{\alpha_{ab}} - h^{-1}_{\alpha_{ab}}\right) h^{-1}_{e_\alpha}\left\{h_{e_\alpha}, V\right\}\right) \tag{8.7.62}$$

The purpose of all the previous manipulations are now clear. The above expression can be converted to an operator with a well defined actions on the states.

$$\hat{H}^E = \lim_{\epsilon\to 0}\sum N_I \epsilon^{abc} Tr\left(\left(\hat{h}_{\alpha_{ab}} - \hat{h}^{-1}_{\alpha_{ab}}\right) \hat{h}^{-1}_{e_\alpha}\left\{\hat{h}_{e_\alpha}, \hat{V}\right\}\right) \tag{8.7.63}$$

By these methods we have been able to express (part of) the Hamiltonian constraint as an explicit and computable operator on the Hilbert space of Gauss and diffeomorphism invariant states. Solutions of the Hamiltonian constraint have been found. This represent promising progress but this problem is still not satisfactory dealt with. There are still to many unexplained ambiguities in the construction of the operator, and the space of solutions are not yet described in a manner that is fully satisfactory. In principle a least, one proposed solution is at hand, and again in principle this would solve the dynamics of the system. However even if we considered the present solutions of

the Hamiltonian constraints to be a proper definition of the physical Hilbert space, we would still only have an implicit understanding of the dynamics. The constraints only define a Hilbert space. From this Hilbert space the correlated changes of observables have to be extracted by relational means. This is the problem we turn to next.

8.8 Dynamics in LQG

Any Dirac observable must preserve the physical Hilbert space and must therefore commute with all constraints. Since the Hamiltonian is a linear sum of constraints, these Dirac observables must also commute with the Hamiltonian. This means that the Dirac observables are constant with respect to the evolution induced by the Hamiltonian. This is often referred to as the "problem of time", but is actually a natural and expected consequence of creating a background free theory. In such theories time is not a fundamental entity. This means that we cannot define dynamics as one-parameter algebra automorphisms. Instead the dynamics of LQG has to be implemented by relational means [77][25][165]. The proper method for doing dynamics is not completely settled, but the basic idea is pretty clear. We will explain the approach in three steps. First we will review Rovelli's idea of evolving constants of motion. This idea is capable of describing dynamics in the classical case [93]. We then move on to a technique based on conditional probabilities suggested by Page and Wooters [176]. We finally combine the essence of both of these ideas in the approach based on conditional probabilities of evolving constants suggested by Gambini and Pullin [177]. We base our presentation on [165][178][179].

An evolving constant of motion is a one-parameter family of Dirac observables $O(t)$ [93]. By convention the parameter of such a family is denoted t. This t is just a parameter and does not represent time. For each (allowed) value of t the function $O(t)$ is a Dirac observable (with vanishing Poisson bracket with all constraints). Let the system be described by canonical variables $\{q_1, q_2, p_1, p_2\}$. The functions corresponding to these variables $\{Q_1, Q_2, P_1, P_2\}$ are not Dirac observables. We require that $O(t)$ be such that when $t = q_1$ then $O(t) = q_2$. This brings back the notion of change in an observable. Every Dirac observable in the one-parameter family is a good observable and we have now parameterized its relation to a canonical variable. However, the parameter t is not a Dirac observable. The evolving constants do describe parameterized change, but in terms of a non-observable parameter. This might not be entirely satisfactory even at the classical level. When we consider the quantum extension of the evolving constants idea we face a problem. Canonical variables (or rather the functions corresponding to canonical variables) are mapped by the representation $\mathcal{Q}$ to operators on $\mathcal{H}$. It is difficult to make sense of an operator as a parameter for another operator. One could make some more or less arbitrary decision to exclude the parameter variable q_1 (Q_1) from quantization, but this is hard to motivate on general grounds.

To better deal with describing parametrized change in the quantum theory conditional probabilities was introduced[176]. In this approach one first selects two quantum observables represented by operators X and T. One then tries to formulate the question "What is the probability of measuring X to be x_0 when T is measured to be t_0?". This probability can be expressed in terms of projection operators. We define the following formula

$$P(X = x_0 | T = t_0) = \frac{\langle \Psi | \, P_{T_{t_0}} P_{X_{x_0}} \, | \Psi \rangle}{\langle \Psi | \, P_{T_{t_0}} \, | \Psi \rangle}. \tag{8.8.1}$$

Now, the question is which observables X and T should we choose, and on which space should they operate. If one chooses X and T to be Dirac observables then they will represent non-changing properties, and we cannot use them to describe parameterized change. If instead one choose observables X and T, which are not Dirac observables, then they will potentially represent changing properties. If the relevant space of states is chosen as the space of physical states, then

the projection of non-Dirac observables do not necessarily have a closed action on this space. This makes the equation 8.8.1 meaningless as the inner product is no longer defined. If the relevant space is chosen to be something other than the space of physical states, then the formula can make mathematical sense but its interpretation is unclear. The predictions of LQG is encoded in the space of physical states. Using any other space calls into question what kind of dynamical behavior one is actually extracting by this method.

Gambini and Pullin have suggested a modified version of the conditional probabilities approach that makes different suggestion for which operators should be used in the conditional expression [177]. It suggests that one should construct two parameterized families of Dirac operators $X(t)$ and $T(t)$, and define the conditional probability as

$$P(X = x_0|T = t_0) = \lim_{\tau\to\infty} \frac{\int_{-\tau}^{\tau} \mathrm{d}t \, \langle\Psi| \, P_{T_{t_0}}(t) P_{X_{x_0}}(t) \, |\Psi\rangle}{\int_{-\tau}^{\tau} \mathrm{d}t \, \langle\Psi| \, P_{T_{t_0}}(t) \, |\Psi\rangle}. \tag{8.8.2}$$

In this case one is able to use the space of physical states since the projection operators are Dirac observables. The parameter t is not observable, but it is also not a part of the final expressions as it is integrated over. This version seems to address several of the weakness of the approaches of the earlier approaches.

8.9 The classical limit

Any acceptable theory of quantum gravity has to reproduce what we already know about space and time at lower energies. This is usually though to occur in some sort of classical and/or macroscopic limit. The classical limit is the limit of the Planck constant going to zero, while the macroscopic limit is intended here to mean the limit of many "quanta". (We put quanta in quotes since the construction of larger structures from quanta is only loosely defined in some quantum gravity approaches.) Specifically for LQG, this means that the theory has to have a limit where it reproduces (most of) the predictions of GR. This limit has proved to be difficult to uncover. The discrete to continuum transition that is necessary is a key part of the difficulty. When moving from an ordinary quantum mechanical description of matter to a classical one, one does not ordinarily derive the appearance of continuum matter. Instead one derives how quantum particles move in the classical limit. When doing the same for a quantum geometric theory one must confront the discrete to continuum transition in a more head on fashion. In essence one sets out to make the continuum background stage of physics reappear from the discrete quantum building blocks. Of course such a continuum background stage will not truly appear, much like actual continuum matter does not truly appear from the atomic world, but it must appear as a very precise approximation at low energies. There are many special techniques and questions related to obtaining the classical limit. We will not pursue this question in any further detail.

8.10 Black hole entropy and entanglement entropy

Due to limitations of time and space we will not give a separate treatment of black hole entropy and entanglement entropy within LQG. These topic will be treated in chapter 9 mostly in the context of string theory. In the course of the treatment of those topic under the string theory heading will include a few sidebars on how the concepts are used in LQG. It will also form a part of the summary of chapter 10 and in the postscriptum. We put this short note in this chapter so as to signal that these topics, even though we have not given them a separate treatment, will form an important part of the last parts of chapters 9 and 10.

8.11 Cosmology

One consequence of GR was the extension of the theory of celestial mechanics to a theory of the dynamics of the universe. The dynamics of the universe has since been known as the field of cosmology. It is probably correct to say that cosmology did not exist as a physical science before GR. The invention of cosmology was a major step forward, independent of whether one sees this as the result of the general relativistic view that the laws of gravity are laws governing the structure of spacetime, or as due to the perhaps more philosophical realization that the universe itself might be a dynamical physical entity. Regardless of what was the most central aspect of this realization, it placed the history of the universe within the domain of gravitational theory.

A key task for cosmology is finding the proper gravitational equations (and solutions) for the whole universe. We shall briefly review the simplest classical cosmological model. Afterwards we move on to look at a quantum version of the same model. The results of this section are well known and will not be individually references. We base our presentation on [180][147],[148] and [158].

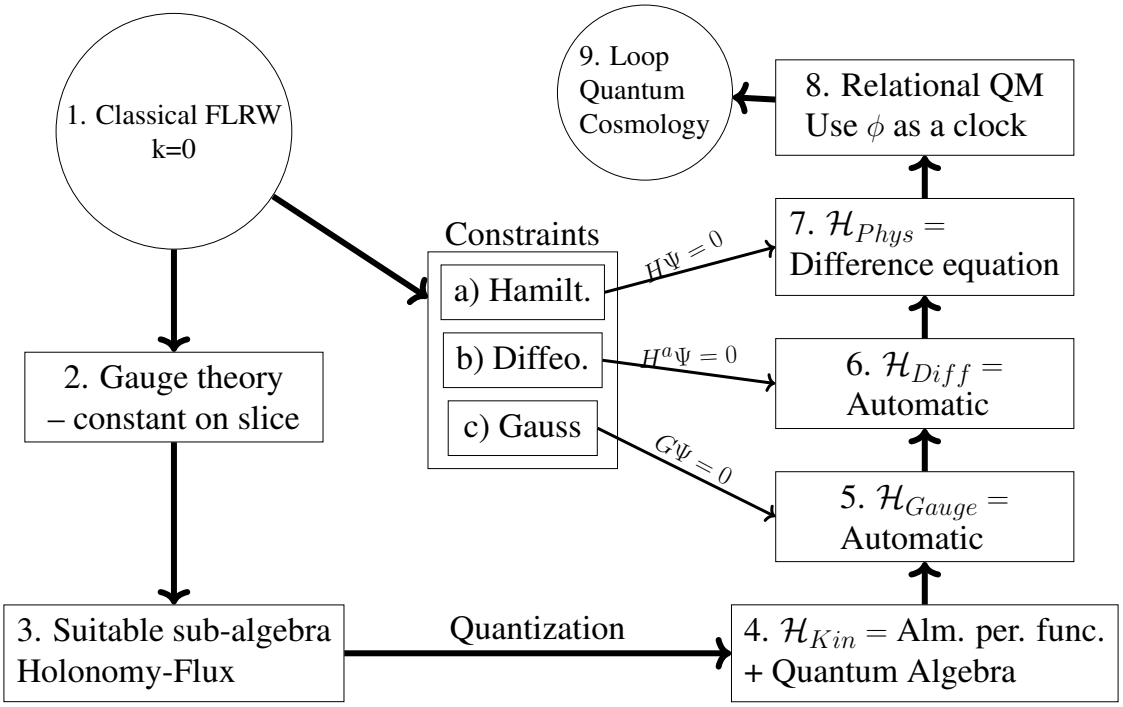


Figure 8.13: The diagram shows how one LQC model is constructed from the classical FLRW $k = 0$ model. Abbreviations: alm. per. func. = almost periodic functions, Hamilt. = Hamiltonian constraint, diffeo. = diffeomorphism constraint

Classical Homogeneous isotropic flat cosmological models

Cosmology aims to make a model of the whole universe. To make such a model we must start by making some assumptions and simplifications. A common and useful assumpti ons is that the universe is simple and uniform. This translates to the assumption that the universe is homogeneous and isotropic. Homogeneity and isotropy are obviously not what we observe in the universe today, but on very large distance scales, and at low angular resolutions, the averages found in the various subdivisions are near identical. Nevertheless, the universe is very non-uniform at shorter scales, and it is an fortuitous and significant finding that homogeneous and isotropic models seem to capture many aspects of the history of the universe quite well. The Friedmann-Lemaitre-Robertson-Walker (FLRW) model with $k = 0$ and $\Lambda = 0$ is one important simple model of the universe. In the FLRW

model there is only one dynamical variable, namely the scale factor $a(t)$. We will define this model by giving the the form of the metric in co-moving coordinates. The FLRW ($k = 0$, $\Lambda = 0$) metric is given by [148]

$$ds^2 = -dt^2 + a(t)\left(dx^2 + dy^2 + dz^2\right). \tag{8.11.1}$$

Intuitively this can be though of as a flat three-dimensional geometry that expands with increasing values of t from a singular beginning at $t = 0$. We will now proceed to treat this system quantum mechanically.

Loop quantum cosmology

Quantum cosmology is the study of cosmology within a quantum mechanical framework. Usually we look at quantum cosmology as being specific models within a general theory of quantum gravity. For LQG constructing a cosmological model within the complete theory has not yet succeeded. Instead one tries to find quantum mechanical models inspired by the full theory but in simpler settings. This direction of research goes by the name loop quantum cosmology (LQC).

LQC is mostly based on quantizing well-known classical cosmological solutions by LQG derived methods. In the case we are presenting, we start with a classical phase-space, as described by the FLRW model, and proceed to quantize this along the lines of LQG. This must at present be considered a separate theory from LQG, that is inspired and heuristically based on LQG. The major point that connects LQC with LQG is using a similar classical phase space and a similar set of quantization techniques. The hope is that LQC models will give a picture of the universe that is closely related to the results a full cosmological solution of LQG would give.

LQC is also interesting from a slightly different point of view. LQG is a complex theory where extracting concrete physical predictions is quite difficult. Any concrete examples of solutions where actual predictions are possible would be of interest. This could tell us more about how to go about deriving physics from LQG. The problem of time and the problem of implementing the Hamiltonian constraint are important LQG issues that can be effectively addressed in the simpler LQC setting. By extension such concrete models, even though they are not actually LQG models, can also serve as an indicator of potential problems within LQG, as we shall see to some extent in the "Critique of the model" subsection.

Quantum FLRW model

The classical system that we will quantize is the FLRW model $k = 0$ and $\Lambda = 0$. We will consider the topology of σ to be $\mathbb{R}^3$, which is a non-compact space. This means that any integral of homogeneous objects will diverge (or be zero). We therefore define a cubical fiducial cell $\mathcal{V}$ of σ with finite volume V. All calculations will initially be done just referring to this fiducial cell. In the end we must take the limit $\mathcal{V} \to \mathbb{R}^3$. We define Cartesian coordinates x^a on $\mathcal{V}$ with associated triads and cotriads $\mathring{e}$ and $\mathring{\omega}$. The edges of the cubical cell are aligned with the coordinate axes.

The Ashtekar-Barbero connection and the densitized triads become

$$A_a^i = cV^{-\frac{1}{3}}\mathring{\omega}_a^i = cV^{-\frac{1}{3}}\delta_a^i \tag{8.11.2}$$

$$E_i^a = pV^{-\frac{2}{3}}\sqrt{\mathring{q}}\mathring{e}_i^a = pV^{-\frac{2}{3}}\delta_i^a. \tag{8.11.3}$$

The volume factors compensate for side lengths and the side areas of the fiducial cell $\mathcal{V}$. Writing out the expression for the underlying geometric objects we get

$$A = A_a^i dx^a \otimes \tau_i = cV^{-\frac{1}{3}}\mathring{\omega}_a^i dx^a \otimes \tau_i = cV^{-\frac{1}{3}}\delta_a^i dx^a \otimes \tau_i \tag{8.11.4}$$

$$E = E_i^a \partial_a \otimes \tau^i \ = pV^{-\frac{2}{3}}\sqrt{\mathring{q}}\mathring{e}_i^a \partial_a \otimes \tau^i = pV^{-\frac{2}{3}}\delta_i^a \partial_a \otimes \tau^i. \tag{8.11.5}$$

In these equations c and p are constants fields on σ. They can of course be depend on t. In the cosmology they will appear as the basic dynamics variables. The Poisson brackets between c and p are determined to be

$$\{c, p\} = 8\pi G/3. \tag{8.11.6}$$

Holonomies and fluxes

Let $e_{(b)}$ be a line segment of length $l^{(b)}$, parallel to the (b)-th coordinate axis of the aligned Cartesian coordinate system of the cubical volume V. We calculate the holonomy along the edge[8] $e_{(b)}$ to be

$$h[e_{(b)}, \mathcal{A}] = \mathcal{P} \exp\left\{ \int_{e_{(b)}} cV^{-\frac{1}{3}} \mathring{\omega}_a^i dx^a \otimes \tau_i(\partial_{(b)}) \right\} \tag{8.11.7}$$

$$= \mathcal{P} \exp\left\{ \int_{e_{(b)}} cV^{-\frac{1}{3}} \mathring{\omega}_a^i dx^a (\partial_{(b)}) \tau_i \right\} \tag{8.11.8}$$

$$= \mathcal{P} \exp\left\{ cV^{-\frac{1}{3}} l^{(b)} \delta_a^i \delta^a_{(b)} \tau_i \right\} \tag{8.11.9}$$

$$= \mathcal{P} \exp\left\{ cV^{-\frac{1}{3}} l^{(b)} \tau_{(b)} \right\} \tag{8.11.10}$$

Since we chose to specify the edge by ∂_b the index (b) on the τ matrix is unusual, but this is as expected since we are using δ_a^i, and we will simply rename (b) to (k). We define $\mu = V^{-\frac{1}{3}} l^{(k)}$ and introduce the shorthand notations $h^\mu_{(k)} = h[e_{(k)}, \mathcal{A}]$. Note that μ is not an index but refers to the quantity μ (the area) that we just defined. When we complete the calculation we get

$$h^\mu_{(k)} = h[e_{(k)}, \mathcal{A}] = \exp\left\{ cV^{-\frac{1}{3}} \tau_{(k)} l^{(b)} \right\} = \exp\left\{ \mu c \tau_{(k)} \right\} = \cos\Big(\frac{\mu c}{2}\Big)\mathbb{I} + 2\sin\Big(\frac{\mu c}{2}\Big)\tau_{(k)} \tag{8.11.11}$$

The factors of 2 and $\frac{1}{2}$ in the last equation derive from using the definition $\tau_i = -\frac{i}{2}\sigma_i$. The holonomies are matrix valued functions and we therefore want to build an algebra from functions that are holonomy matrix component valued. The set of all holonomy component functions is equivalent to complex linear combinations of the basis functions $\mathcal{N}_\mu(c)$, which are defined as

$$\mathcal{N}_\mu(c) = \exp\Big(\frac{i\mu c}{2}\Big) = \cos\Big(\frac{\mu c}{2}\Big) + i\sin\Big(\frac{\mu c}{2}\Big). \tag{8.11.12}$$

The function space spanned by functions of this form is called the space of almost periodic function of the real line[9]. This space of functions can also be formulated as all continuous on a compactification of the real line $\mathbb{R}_B$ called the Bohr compactification. This is the space of functions (observables) on the configuration space. The subspace of square integrable functions will become the Hilbert space. We write this gravitational kinematic Hilbert space as

$$\mathcal{H}^{grav}_{kin} = L^2\left(\mathbb{R}_B, d\mu_B\right). \tag{8.11.13}$$

The flux can be calculated to be

$$F_S[f] = \int_S E^c_i f^i \epsilon_{abc} dx^a \wedge dx^b = pV_0^{-\frac{2}{3}} A_{S,f}. \tag{8.11.14}$$

[8] The parenthesis (b) is meant to indicate a fixed index. There is no summation on this index.

[9] See appendix J for more on almost periodic functions and the Bohr compactification.

Here $A_{S,f}$ is an area-test-function factor. The flux is therefore essentially the same as p and we will use p as our flux variable. We can now express the Poisson bracket between $\mathcal{N}_\mu(c)$ and p as

$$\{\mathcal{N}_\mu(c), p\} = i\frac{4\pi G\gamma}{3}\mu\mathcal{N}_\mu(c). \tag{8.11.15}$$

The action of the quantum operators $\hat{\mathcal{N}}_\mu$ and $\hat{p}$ corresponding to $\mathcal{N}_\mu(c)$ and p can now be derived. The action of the operator $\hat{\mathcal{N}}_{\mu'}(c)$ corresponds to the multiplicative action of the function

$$\mathcal{N}_{\mu'}(c) = \exp\left\{\frac{i\mu' c}{2}\right\} \tag{8.11.16}$$

on a function

$$\mathcal{N}_\mu(c) = \exp\left\{\frac{i\mu c}{2}\right\} = |\mu\rangle\,. \tag{8.11.17}$$

We write the action as

$$\hat{\mathcal{N}}_{\mu'}(c)\,|\mu\rangle = \exp\left\{\frac{i\mu' c}{2}\right\}\exp\left\{\frac{i\mu c}{2}\right\} \tag{8.11.18}$$

$$= \exp\left\{\frac{i(\mu+\mu')c}{2}\right\} \tag{8.11.19}$$

$$= \left|\mu+\mu'\right\rangle. \tag{8.11.20}$$

The operator $\hat{p}$ is defined by a derivative as

$$\hat{p} = -i\frac{8\pi l^2\beta}{3}\frac{\partial}{\partial c} \tag{8.11.21}$$

When $\hat{p}$ acts on a function

$$\mathcal{N}_\mu(c) = \exp\left\{\frac{i\mu c}{2}\right\}, \tag{8.11.22}$$

we get the action

$$\hat{p}\,|\mu\rangle = -i\frac{8\pi l^2\beta}{3}\frac{\partial}{\partial c}\exp\left\{\frac{i\mu c}{2}\right\} \tag{8.11.23}$$

$$= -i\frac{8\pi l^2\beta}{3}(\frac{i\mu}{2})\exp\left\{\frac{i\mu c}{2}\right\} \tag{8.11.24}$$

$$= \frac{4\pi l^2\beta}{3}\mu\exp\left\{\frac{i\mu c}{2}\right\} \tag{8.11.25}$$

$$= p(\mu)\,|\mu\rangle \quad \text{with} \quad p(\mu) = \frac{4\pi l^2\beta}{3}\mu. \tag{8.11.26}$$

This completes the setup of the unconstrained Hilbert space.

Constraints

It is now time to implement the constraints. The Gauss constraints and the diffeomorphism constraints are automatically solved by the homogeneity of the space. The only non-trivial constraint is the Hamiltonian constraint. We recall that the general Hamiltonian constraint is given by

$$H(N) = \int d^3x\epsilon^{ij}_k\frac{E^a_iE^b_j}{\sqrt{det(E)}}\left(F^k_{ab} - 2(1+\beta^2)K^i_aK^j_b\right) \tag{8.11.27}$$

The last part is zero for the current setup so we need only focus on the first part (the Euclidean part). With this simplification the Hamiltonian constraint becomes

$$H(N) = \int d^3x \epsilon^{ij}_k \frac{E^a_i E^b_j}{\sqrt{det(E)}} F^k_{ab}. \tag{8.11.28}$$

This can also be expressed as

$$C_{grav} = -\frac{6}{\gamma^2} c^2 \sqrt{|p|} \tag{8.11.29}$$

Like in the full theory, we use Thiemann's rewrite trick to tackle the Hamiltonian constraint, and we get

$$\frac{E^a_i E^b_j}{\sqrt{det(E)}} \epsilon^{ijk} \epsilon_{abc} = \left\{ A^k_a, V \right\}. \tag{8.11.30}$$

The connection and the curvature are not define as quantum operator, and therefore we cannot apply this formula directly. The basic idea is to rewrite the connection and the curvature in terms of holonomies. We get the expression

$$\left\{ A^k_a, V \right\} = \mathrm{Tr}\Big(h^\mu_k \left\{ \left(h^\mu_k\right)^{-1}, V \right\} \Big). \tag{8.11.31}$$

This is closely parallel to the derivations in the full theory. The curvature can be expressed using holonomies around a standard plaque denoted $\square_{ab}$ (the plaque follows the coordinate lines of coordinates a and b). We denote such a holonomy by $h_{\square_{bc}}$. For the curvature we then get

$$F^k_{ab} = \lim_{\mu \to 0} \frac{h_{\square_{de}} - \mathbb{I}}{\mu^2} \tau^k \delta^d_a \delta^e_b. \tag{8.11.32}$$

We rewrite this completely in terms of holonomies as

$$F^k_{ab} \propto \mathrm{Tr}\left(h^\mu_i h^\mu_j (h^\mu_i)^{-1} (h^\mu_j)^{-1} \right). \tag{8.11.33}$$

In the full theory we could take the limit as the area goes to zero, but this is not possible in the homogeneous case. The question then arises of what to do instead. The leading answer is that the discrete spectrum of the full theory area operator shows that there exists a minimal area. This further suggests that instead of taking the limit of the area going to zero, one should take the limit of the area going to the minimal value Δ. The length value that gives this minimal area is denote by $\overline{\mu}$ where

$$\overline{\mu} = \sqrt{\frac{|p|}{\Delta}}. \tag{8.11.34}$$

When we apply this suggestion we get

$$C_{\text{gravity}} = \mathrm{Tr}\left(h^\mu_i h^\mu_j (h^\mu_i)^{-1} (h^\mu_j)^{-1} h^\mu_k \left\{ \left(h^\mu_k\right)^{-1}, V \right\} \right) \tag{8.11.35}$$

$$= \sin^2(\bar{\mu}c) \left[\sin(\bar{\mu}c) V \cos(\bar{\mu}c) - \cos(\bar{\mu}c) V \sin(\bar{\mu}c) \right]. \tag{8.11.36}$$

The action of the Hamiltonian constraint on the states is then given as

$$C_{\text{gravity}} |\mu\rangle = C^-(v) |\mu - 4\bar{\mu}\rangle + C^0(v) |\mu\rangle + C^+ |\mu + 4\bar{\mu}\rangle. \tag{8.11.37}$$

To achieve the goal of some sense of time evolution we need to add some way of measuring "time" to the system. One suggestion would be to use the scale parameter as a time parameter, however this is only monotonic for a limited set of possible evolutions. A more successful approach is to add some suitable matter field. The matter content we will add is a homogeneous isotropic scalar field ϕ. The presence of a scalar field allow us to define a measure of time that is independent of the scale parameter. The only non-trivial Poisson bracket involving ϕ is given by

$$\{\phi, p_\phi\} = 1. \tag{8.11.38}$$

The (kinematic) Hilbert space used for the scalar field is

$$\mathcal{H}_{kin}^{matt} := L^2\left(\mathbb{R}, d\mu\right). \tag{8.11.39}$$

The presence of the scalar fields extends the definition of of the Hamiltonian constraint.

$$C = C_{matter} + C_{gravity} \tag{8.11.40}$$

$$C_{matter} := 8\pi G \frac{p_\phi^2}{|p|^{3/2}} \tag{8.11.41}$$

$$\tag{8.11.42}$$

We can now express the full Hamiltonian constraint with matter. Since the curvature is determined by a non-local expression of holonomies around a finite plaques this will lead to a difference equation instead of a differential equation. The operator p_ϕ is implemented as ∂_ϕ. This means the equation

$$C_{\text{Hamilt.}} \left|\mu, \phi\right\rangle \ = \left(C_{\text{gravity}} + C_{\text{matter}}\right) \left|\mu, \phi\right\rangle, \tag{8.11.43}$$

can be written as

$$C^-(v) \left|\mu - 4\bar{\mu}, \phi\right\rangle + C^0(v) \left|\mu, \phi\right\rangle + C^+ \left|\mu + 4\bar{\mu}, \phi\right\rangle - \frac{\partial^2}{\partial \phi^2} \left|\mu, \phi\right\rangle = 0. \tag{8.11.44}$$

The symbols C^-, C^0, C^+ represents numerical prefactors that we will not define. Equation 8.11.44 can be solved by numerical methods. Starting from large volume states that are peaked about the classical solution leads to a evolution without a Big Bang singularity. An example of the kind of solutions that results from this equation is shown in figure 8.14.

Critique of the model

While the bouncing universe results based on minisuperspace LQC models are very appealing, one must be cautious with regards to their merits. One could hope that the quantum FLRW-models are the analogs of the Bohr hydrogen model and the classical FLRW-models, where very simple models capture a lot of the relevant physics. However, there is no guarantee that such simple models are relevant. The LQC approach described above involves several assumptions and approximations. The high degree of symmetry lead to a natural suspicion as to whether the results achieved are a good indicator of what happens in the high density domain. An analysis of the hypersurface deformation algebra (see [182] for a definition), leads to the conclusion that the beta function of the Lie algebroid (see [183][184] for definitions) becomes negative at Planck density leading to a space without time. This raises serious concerns about the possibilities of causal evolution in the high density domain. Certainly it casts doubt over the "bounce" picture that LQC presents [185][186]. Example of signature change have also been found in other LQG related systems[10] [187][188].

[10] We have not had time to properly study these papers but we include them because of the great significance of any formt of signature change.

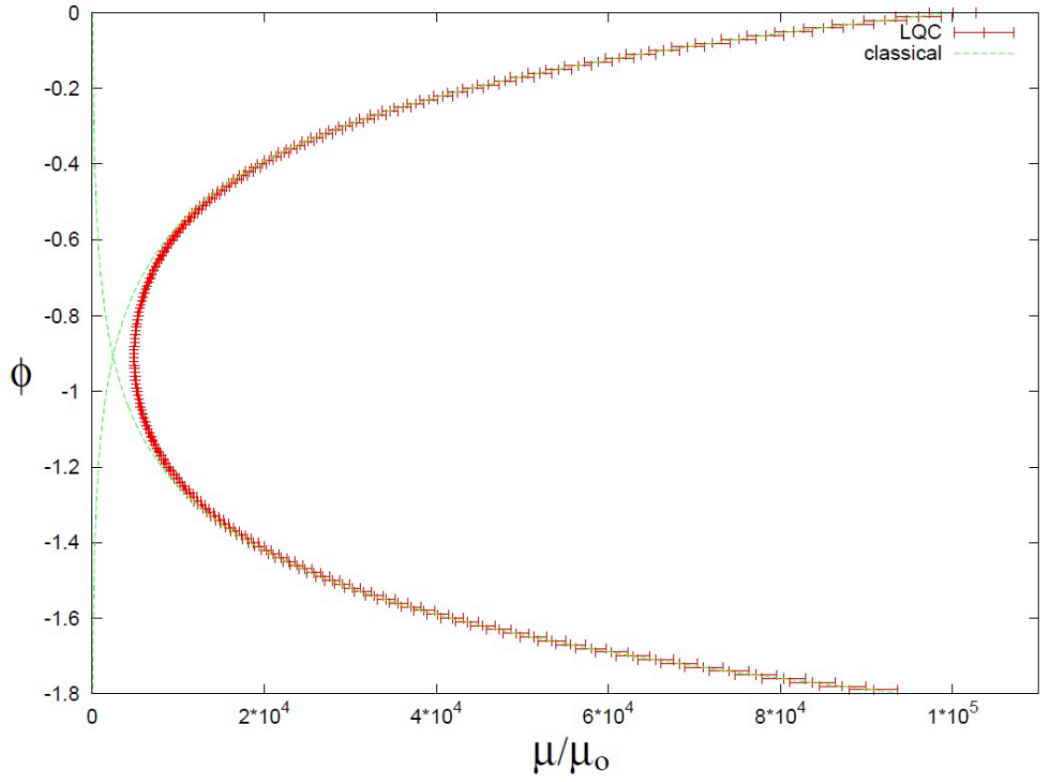


Figure 8.14: The diagram shows how the universe evolves in a quantum FLRW model. The ϕ values on the y-axis represents a relational measure of time. The x-axis represents a measure of the size of the universe in units of the distance μ_0 (corresponding to the minimal area). The red bars represent the results of a numerical simulation of the LQC model. There is a smooth transition from a collapsing phase to an expanding phase at low volume. There are no singular points where the volum vanishes. The faint dashed green lines shows the corresponding classical solutions. This plot shows how the singularity of the classical model is replaced by a quantum induced bounce a low volume. Diagram from [181].

8.12 LQG in higher dimensions

The above formulation of LQG builds crucially on a specific connection based reformulation of GR that is only possible in $D+1=4$. However, Bodendorfer et al. have created a different connection based reformulation that applies for any $D+1 \geq 3$ [189][190][191][192][193][194]. This approach produces a gauge theory with gauge group $\mathrm{SO}((1,D))$ or $\mathrm{SO}(D+1)$ with an extra constraint called the simplicity constraint. The theory is more complicated but the structure is similar to what we have developed in this chapter.

Let us first examine what special features make traditional $(3+1)$-dimensional LQG possible. An essential fact is that the (A,E) pair of connection and densitized triad variables present with a standard Poisson bracket between themselves where specifically A commutes with itself [189]. It is also crucial that the group of the gauge connection is the compact group $\mathrm{SU}(2)$ [189]. The theory depends on having a compact group to do harmonic analysis on. Finally, it is important that the classical setup has only first class constraints.

Let us look at the degrees of freedom for the most central variables (see table 8.2). To be able to use triads as the conjugate variables of the connection, the connection and triads must have the same number of degrees of freedom. From table 8.2 we see that this only happens when $D=3$.

Variable	DOF or Rank	For D=3
Metric	$D(D+1)/2$	6
Connection	$D^2(D-1)/2$	9
Triad	D^2	9
SO(D)	$D(D-1)/2$	3

Table 8.2: Degrees of freedom in LQG variables.

The formula

$$\Gamma_a^i \to A_a^i = \Gamma_a^i + \gamma K_a^i \tag{8.12.1}$$

also contains an essential fact which is unique to $D = 3$. The Γ_a^i transforms in the adjoint representation of $\mathrm{SO}(3)$ while K_a^i transforms in the fundamental representation of $\mathrm{SO}(3)$. Only for $D = 3$ are these two representations the same (as they must be for this to work).

We will not review the details of this approach. We state only that it is possible in $(D+1)$ dimensions to find a new connection A_{aIJ} and a new set of triad related conjugate momenta π^{aIJ} with canonical Poisson brackets [189]

$$\left\{A_{aIJ}(x), \pi^{bKL}(y)\right\} = 2\beta\delta_a^b\delta_{[I}^K\delta_{J]}^L\delta^D(x,y) \tag{8.12.2}$$

$$\{A_{aIJ}(x), A_{bKL}(y)\} = 0 \tag{8.12.3}$$

$$\left\{\pi^{aIJ}(x), \pi^{bKL}(y)\right\} = 0 \tag{8.12.4}$$

where we have used units with $\kappa = 1$.

We do want to emphasize how important this new achievement is for LQG. First of all it sets LQG free of its roots in the quantization of GR. It is now clear that LQG is better seen as a set of methods for quantizing any connection based theory in a background free manner. Secondly, many interesting gravitational theories are formulated outside of $(3+1)$ dimensions. LQG is now able to apply its methods to those theories. This involves theories of supergravity, which are of interest as the low-energy limit of string theory. One potential application of this is to extend the set of theories that can be calculated on the gravity side of the gauge-gravity duality [144].

8.13 Notes

A brief history

Quantum gravity research has a long history (see [25] or [77]). The possibility of quantum effects modifying GR was noted by Einstein himself at a very early stage [195]. Skipping several decennia where GR took a backseat on the physics stage we move to the late 1950s. The Hamiltonian formulation of GR and the study of quantization of constrained Hamiltonian systems was made by Dirac [196, 197] and Bergman [198, 199]. The Hamiltonian analysis was completed by Arnowitt, Deser and Misner ADM (1962 publication, reference information difficult to find). Wheeler and DeWitt wrote down a quantum gravitational Schrodinger equation called the Wheeler-DeWitt (WDW) equation in 1967 [200, 201, 202]. The WDW equation is basically the expression of the quantum Hamiltonian constraint in the ADM approach. The WDW Hamiltonian constraint is a non-polynomial expression and this approach remained mostly formal since one was unable to define precise mathematical content to correspond to the terms of the WDW-equation.

Twenty years later, in 1986, Ashtekar found a new Hamiltonian formulation of GR which made the Hamiltonian constraint polynomial [153][154]. This made GR much more simliar to gauge theories. However, direct quantization of the Ashtekar variables did not succeed. Further progress required the ideas of holonomies and Wilson loops. In 1988 Jacobson and Smolin found the loop representation [155] and these ideas where furhter developed by Rovelli and Smolin [156][203]. Wilson loops form an over-complete basis and the better suited spin-network basis was introdued in [204].

The initial formulation of LQG used a complex Immirzi parameter and produced complex GR instead of real GR. Implementing reality conditions was very difficult and basically canceled the advantages of the polynomial version of the Hamiltonian constraint. In addition this approach used the non-compact gauge group SL(2C). This prevented the use of the more powerful tools available for compact groups. Finally, the approach used a Hamiltonian that was twice densitized, and it was later shown that only a singly densitized expressions can be quantized without divergences. The theory was therefore reformulated by Thiemann in 1998 using Ashtekars original variables but with the compact gauge group $SU(2)$ and a real Immirzi parameter [175][205][206]. This meant the Hamiltonian constraint went back to being a complex non-polynomial expression. Thiemann was still able to implement a version of this operator as concrete action on the states.

A reformulation of LQG methods to arbitrary dimensions ($D + 1 \geq 3$), which can also be applied to supergravity theories, was made by Bodendorfer, Thiemann and Turn in 2013 [189][190][191][192][193][194].

GNS construction

The GNS construction produces a *-algebra representation (or if you will a *-algebra homomorphism). In the context of quantization the algebra we want to represent by this construction is not the algebra given by the associative (and commutative) part of the classical Poisson algebra. Instead what we want to represent is the abstract algebra (which we have called $\mathfrak{B}$) that is Lie algebra isomorphic to the Lie algebra portion of the classical algebra (which we have called $\mathfrak{A}$) where the Lie algebra product is given by a commutator. We have not constructed such abstract quantum algebras for the algebras we discuss in this chapter and we therefore want to make sure that the proper application of the GNS construction is perfectly clear. It can be confusing that the the linear functional used in the GNS construction for LQG is defined on the classical algebra $\mathfrak{A}$, but not that we have explained how it is transported to $\mathfrak{B}$.

In figure 8.5 we have made a presentation of quantization without reference to the GNS construction. In this case we have just a Lie algebra homomorphism (Lie algebra representation). This skips the middle step of an abstract quantum algebra and also skips the GNS step. We can view a quantization in either of these way. However, in this simplified case we do not have the GNS technology available and we cannot use the GNS construction to analyze the representation theory.

Let us clarify a bit by using the example of 1-D QM. The Poisson algebra of the the functions Q and P and 1 is Lie algebra isomorphic to the Heisenberg algebra. The Heisenberg algebra is abstract quantum algebra. We can then use GNS construction on this algebra to find the representations. The Stone - von Neumann theorem guarantees that there is a unique (up to scale) non-trivial central strongly continuous unitary representation. One can also view the quantization in the simplified manner by focusing on the Lie algebra homomorphism directly from the classical Poisson algebra to the Hilbert space operator algebra.

We have mentioned the Weyl countermeasure in the text that deals with the presence of unbounded functions in the classical (sub)algebra. The Weyl countermeasure can be applied to the classical (sub)algebra before constructing (or in the process of constructing) the abstract quantum algebra (see e.g. [25, Ch3, Ch7] and [207] for the application of the Weyl technique to the holonomy-flux algebra).

Various details

When we are calling supergravity perturbative we mean that it contains a spin-2 based formulation of gravity. In traditional quantization the spin-2 graviton propagate on a background spacetime. This makes the standard definition of gravity within supergravity a perturbative definition. The theory itself is not perturbative beyond this limitation. It is likely that supergravity can be quantized by loop methods in a background independent manner.

V Extended theories III - String theory

9. String theory

In the years immediately after its birth, no one knew that string theory was a theory of strings. It seems very possible that we're currently in a similar situation. When the theory is better understood, it may have little to do with strings.

Tong

Summary

Perturbative string theory can be seen as a generalization from one-dimensional world-line QFT to two-dimensional world-sheet QFT. Except for spacetime, the string is the only fundamental entity in the theory. Different particles with different spins are all realized as states of the string. The string spectrum contains an infinite set of states including a massless graviton spin-2 state. The graviton state gives rise to a UV finite spin-2 theory of gravity. When supersymmetry is added, string theory can be consistently quantized in a 10-dimensional spacetime (subject to some consistency requirements on the spacetime). String theory incorporates both gauge bosons and chiral fermions. This can happen in several different ways. Compactifications and intersecting branes are the two leading approaches for making contact with four-dimensional physics. The dynamics of the theory is defined by on-shell perturbative S-matrix elements. The definition of the "free theory" is sufficient for obtaining this perturbative expansion. The theory is believed to be finite order-by-order to all orders. The AdS/CFT conjecture states that string theory on $AdS_5 \times S^5$ is dual to a four-dimensional CFT. This is known as gauge/gravity duality. Entanglement entropy is a central aspect of how gauge theory states and spacetime geometries are related by such dualities.

9.1 Introduction

In this chapter we will review string theory. At the outset we forego any ambitions of being comprehensive or self-contained. Instead we want to make a targeted review centered around three topics. Our first topic is the foundation of string theory. We try to write down a compact definition of what (perturbative) string theory is (section 9.12). This is challenging because as of yet there is no well defined set of principles upon which string theory is founded. Our major tool in this respect will be conformal field theory (CFT) and the world-sheet formalism. The second topic is how string theory can lead to a four-dimensional theory resembling the standard model. While we will not go into complete detail about the low-energy results, we study the major principles involved in both traditional and brane based compactifications of string theory (sections 9.15, 9.16 and 9.17). Our study of phenomenology includes moduli stabilization and the idea of the string landscape (sections 9.18 and 9.19). Our third topic is how string theory attempts to build a theory of quantum gravity. This third topic will be approached in several ways. We study the low-energy limit of string theory in the form of supergravity, and also analyze the requirements on the geometric backgrounds that strings propagate in (section 9.20). The subject of black hole thermodynamics is important for the quantum gravity aspect of string theory, and we present it in section 9.21. From there we move on to look at AdS/CFT (gauge/gravity duality) in some detail (section 9.22). We end our study of gravity in string theory by looking at the idea of entanglement entropy in holographic theories (section 9.23).

Due to enormity of the field coupled with our time and space limitations, we will further limit our review in several ways. We will skip derivations of many of the details involved. We refer to the many textbooks on the subject for more complete derivations (see [208][142][209][210]). Some of the necessary components to construct string theory, like CFTs, supersymmetry, supergravity and various other subjects will be used but we will not review this subjects separately. A basic familiarity with these subjects is assumed on the part of the reader. M-theory and F-theory will not be considered. The construction of the heterotic string will not be reviewed. CFT will be important but only as a pathway to conceptual insights, we will not use it for its computational power.

While we aim to get to our three main subjects fairly quickly, we still afford time for a short introduction to string theory in preparation for the later developments. The chapter starts with an informal introduction to some of the key ideas of string theory (section 9.2). Next we present a historical note. Some of the peculiarities of string theory are easier understood if one knows their historical context, and we therefore spend some time on string history (sections 9.20 and 9.24). The canonical material on bosonic and fermonic strings and their interactions, is presented in sections 9.4, 9.5, 9.6, 9.7, 9.8, 9.9 and 9.10. The material related to non-perturbative aspects and dualities is presented in sections 9.13 and 9.14, and also in subsections of section 9.22. A few details on spinors, supersymmetry and supergravity are quickly summarized in a subsection of section 9.24.

9.2 Generalities

In this section we present an informal overview of string theory. The statements of this section are not individually referenced because of their more casual wording. The proper references for the information of this section can be found in the rest of this chapter (see also the introductory chapter of [211] and [208]).

String theory starts from the idea that the fundamental entities of matter are not zero-dimensional points, but are instead one-dimensional objects, called strings. A string traces out a (1+1)-dimensional world-sheet in spacetime. Strings can move by translation and rotation, as well as by transversely vibrating in various different modes. A string is usually assumed to be very small by common length scales, its length often being taken to be of the order of the Planck length.

(However, more accurately, the string scale is set by the free dimensionful parameter α', and this scale is not necessarily very much higher than the upper range of currently available experiments. Still, we will continue with the common assumption that strings are "small".)

At energy scales well below the Planck scale, strings will appear as pointlike elementary particles. The different quantized vibrational modes will correspond to different elementary particles. Thus, in the string picture, every kind of fundamental particle is just a string in a certain vibrational mode. From the particle point of view, string theory contains an infinite number of fundamental particles, since there is an infinite number of vibrational modes. However, only the very lightest of these modes play any role as normal on-shell particles. The rest of them live hidden lives as off-shell intermediate states in scattering processes.

Strings come in several varieties. There can be open strings, closed strings, oriented strings and unoriented strings. There are six basic string theory setups, consisting of the bosonic string theory and the five superstring theories. Bosonic degrees of freedom are found in all string theories. Superstring theories are string theories with extra fermionic degrees of freedom that are related by supersymmetry to the bosonic degrees of freedom. We will occasionally use the name fermionic strings as an alternative name for supersymmetric string theories. The five supersymmetric string theories are called type I, type IIA, type IIB, heterotic $\mathrm{SO}(32)$ and heterotic $\mathrm{E}_8 \times \mathrm{E}_8$. There are also two additional theories, called M-theory and F-theory, which are 11 and 12-dimensional string like theories. Later we will see that this "one plus five plus two" type classification of string theories is useful, but quite limited. There are also many other "string theories" that can be defined, that bear little resemblance to any of these canonical models.

To set the stage a little bit we introduce some further facts about string theories in an informal manner. Superstring theories live in (9+1)-dimensional space, and the bosonic string lives in (25+1)-dimensional space. Again, we should be aware that this statement about dimensionality is only strictly true when we stick to the canonical "one plus five" string theories. In addition to M-theory and F-theory, it has been discovered that there are e.g. four-dimensional string theories that are not compactified version of a ten-dimensional theory in any geometric sense of the word. The perturbative sector of IIA, IIB and the heterotic theories consists of only closed strings, while in the type I theory there are both closed and open strings. The low-energy limit of the ten-dimensional superstring theories are ten-dimensional supergravity theories. The $\mathcal{N} = 1$ supergravity theories resulting from type I and heterotic string theories include an additional gauge theory in the form of a super Yang-Mills theory. For type I the gauge group must be $\mathrm{SO}(32)$, and for the heterotic theories the gauge group must be $\mathrm{SO}(32)$ or $\mathrm{E}_8 \times \mathrm{E}_8$. The low-energy $\mathcal{N} = 2$ supergravity theories resulting from the IIA and IIB superstring theories, do not include gauge fields. (Note that gauge fields can be introduced by adding non-perturbative objects in the form of D-branes, see section 9.17.)

While we are on the subject of dimensions and low-energy limits we get to an important point. We can assume that even the first humans knew that there were only four macroscopic dimensions in the world. This fact makes it necessary to explain how a ten-dimensional theory is meant to describe a four-dimensional world. The basic model for making contact with (3+1)-dimensional physics is to assume that the extra six dimensions are invisible because they are very small. The process of "shrinking" the extra dimensions, thereby making them unobservably small, is called compactification. As we will see in section 9.20 compactification was first studied by Kaluza and Klein. The compactified dimensions can be considered as a space in itself, and it is often referred to as the compactified space.

An important category of such compactified spaces are six-dimensional manifolds called Calabi-Yau manifolds. It was compactification on Calabi-Yau spaces that first produced semi-realistic models of strings in four dimensions. The study of compactifications is the basis for the field of string phenomenology. The compactification manifolds have free parameters which go by the

name of moduli. There are many such moduli and they can each take many different values. The consequence of this is that there is a very big collection of possible low-energy models of the ten-dimensional superstring. This big collection is called the string landscape. The presences of the landscape makes it difficult to imagine how string theory could be able to produce a unique four-dimensional theory.

String theories are not just theories of strings, they also include higher-dimensional objects called branes. The most important type of branes are known as D-branes. Recall that string theory at the outset is defined by a perturbation series, and is therefore a purely perturbative theory. D-branes allow string theory to also include some non-perturbative effects. D-branes are not part of the perturbative spectrum of string theories, and it is helpful to consider them to be a kind of background field, analogous to background fields in ordinary electrodynamics. This explains why D-branes can be put in and removed by the theorist according to the background he or she wants to create. D-branes are very important for creating models of low-energy physics. At the outset, only open string theory and the heterotic string theories have gauge groups in ten-dimensions. D-branes provide an additional method of introducing gauge groups in string theories. We shall see already in the next section why this is so crucial.

9.3 A historical prelude

On many occasions a historical perspective on developments within theoretical physics can be informative. We will present the canonical event-by-event history in a later section (section 9.24), and that account presents most of the historical context that we need. Still, there is a ten-year period leading up to the anomaly cancellation results of 1984, and the first heterotic string Calabi-Yau compactifications in 1985, that are important for understanding string theory. Without understanding this historical context, both the explosion of interest in string theory in the following years, as well as the direction of research in the following decades, is harder to comprehend. By this we specifically mean that the preceding ten-year period can partly explain how the heterotic string theory compactified on Calabi-Yau manifolds came to be *the* prime candidate for a unified theory for the next decade and a half. It can also set the stage for the later developments concerning M-theory and D-branes. (See [212] for an insider look at this period. For more on Kaluza-Klein history see [213], and for more on supergravity history see [214][215].)

As these developments more properly belong to supegravity research, and are also of a more technical nature, they did not find their proper place within the history in section 9.24, and we therefore want to present them here. This also presents an opportunity to introduce some important ideas which we shall be needing in later sections. So, without further ado, let us try and capture the mood developing in theoretical physics in the period from around 1974 up to 1984. A small but important part of the story, pertaining to Kaluza-Klein theories, actually occurred between 1919 and 1926, and we will recount that part first.

We begin our journey just after the first world war, when fundamental physics mostly meant the theory of GR and the theory of electromagnetism. In 1919, Kaluza wrote a letter to Einstein presenting a model of five-dimensional GR which he then interpreted in terms of a four-dimensional world [216]. (This work was published in 1921.) Kaluza's aim was to created a unified theory of gravitation and electromagnetism by considering these two phenomena as resulting from a theory of pure gravity in five dimensions. Let us see how this works.

As we know, the only dynamical field in GR, in the second-order formulation, is the metric. To write down Kaluza's proposal, let us represent five-dimensional indices by $M, N \in \{0, \ldots, 4\}$, and four-dimensional indices by $\mu, \nu \in \{0, \ldots, 3\}$. Let us also set $z = (x_0, \ldots, x_4)$, $x = (x_0, \ldots, x_3)$

and $y = x_4$. Kaluza subsectioned his five-dimensional metric g_{MN} by writing it as

$$g_{MN} = \left(\begin{array}{c|c} g_{\mu\nu} & g_{\mu 4} \\ \hline g_{4\nu} & g_{44} \end{array}\right). \tag{9.3.1}$$

Here $g_{\mu\nu}$ is the four-dimensional metric, and $g_{4\mu} = g_{\mu 4}$ is a four-dimensional vector. The four-dimensional will be denoted by A_μ. The component g_{44} is a four-dimensional scalar, denoted ϕ, which is taken to be constant. By using the five-dimensional Einstein-Hilbert action, and by using his cylindrical assumption (i.e. no field depends on y) as well as well assuming weak fields, Kaluza was able to derive the vacuum equations for four-dimensional GR plus electromagnetism.

In 1926, Klein reworked Kaluza's ideas by introducing a more explicit notion of the fifth dimension and a more careful analysis of its consequences [217]. Kaluza assumed that spacetime had the structure $M_5 = M_4 \times S^1$. The compact dimension is S^1 which means that any continuous function of this coordinate can be expressed by harmonic analysis in a countably infinite basis of Fourier functions. This Fourier expansion means that we can rewrite the five-dimensional fields as an infinite sum of four-dimensional fields. We use the same decomposition as we used in the Kaluza model to write

$$g_{\mu\nu}(z) = \sum_n g_{\mu\nu,n}(x) e^{iny} \tag{9.3.2}$$

$$A_{\mu\nu}(z) = \sum_n A_{\mu\nu,n}(x) e^{iny} \tag{9.3.3}$$

$$\phi_{\mu\nu}(z) = \sum_n \phi_{\mu\nu,n}(x) e^{iny}. \tag{9.3.4}$$

If we assume the circle to have a radius of the order of the Planck length, we find that all modes except the zero modes will be fields with mass terms of the order of positive integer multiples of the Planck mass. For low-energy physics such very massive fields should be irrelevant, and we are left with the zero modes. This is a more satisfying account of why Kaluza's cylinder assumption should be true. Starting from the five-dimensional Einstein-Hilbert action, and focusing only on the zero-modes, we arrive at

$$S = r \int \sqrt{-g_0} \left(\frac{1}{\kappa^2} R(g_0) - \frac{1}{4} e^{\phi_0} F_{\mu\nu}(A_0) F^{\mu\nu}(A_0) - \frac{1}{6\kappa^2} (\nabla\phi)^2 \right). \tag{9.3.5}$$

The symbol $R(g_0)$ denotes the four-dimensional Ricci scalar calculated from the zero mode of $g_{\mu\nu}$. The symbol $F_{\mu\nu}(A_0)$ denotes the four-dimensional field strength tensor calculated from the zero mode of A_μ. One can show that A_μ is a connection with gauge group equal to the isometry group of the compactified dimension, which is $\mathrm{U}(1)$ in this case.

With the exception of Einstein and Klein few physicist continued to work on Kaluza-Klein theories in the following 40 years. We therefore skip ahead to the 1960s. In parallel with the realization of the importance of non-Abelian Yang-Mills theories it was realized that the Kaluza-Klein construction could be extended to include spaces with more than one extra dimension [218]. The results are basically the same, in addition to four-dimensional GR one gets a four-dimensional gauge theory based on the isometry group of the compactified space, as well as scalar fields for each "radius" of the compactified space.

In the mid-to-late 1970s the four-dimensional standard model with the gauge group $\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(3)$ was becoming firmly established [219]. It was then natural to ask whether there is a compactified space that can produce the standard model gauge group. We have seen that S^1 produces $\mathrm{U}(1)$, which takes care of the first factor. One can show that the lowest dimension manifold that

can produce $\mathrm{SU}(2)$ is S^2, and that the lowest dimension manifold that can produce $\mathrm{SU}(3)$ is $\mathbb{C}P^2$ [220]. Thus the compactified space $S^1 \times S^2 \times \mathbb{C}P^2$ gives the gauge group $\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(3)$.

The space $S^1 \times S^2 \times \mathbb{C}P^2$ is seven-dimensional, which means that this is the minimum dimension compactified space with an isometry group that includes the standard model gauge group as a subgroup [220]. The exact dimensionality of the compact space becomes especially important when we want to consider Kaluza-Klein compactifications of higher-dimensional theories containing fermion fields, which is what we will do next.

The above examples of compactifications only considered higher-dimensional pure gravity theory, where the only dynamical content was the metric field. A proper description of matter must include a description of fermions. This requires using the vielbein based first-order formulation of gravity. When using the first-order formulation it is possible to generate fermions just by compactification of pure gravity, at least in the supergravity case. We have already presented one reason why seven-dimensional compactified spaces might be singled out, and we will present another one when we get to supergravity. For now, let us just assume that seven dimensions are somehow singled out, and let us analyze the consequences this has for fermionic fields undergoing compactification.

Fermions are represented by spinors which are elements of the representation space of Clifford algebras (or equivalently Spin algebras) [221][222]. The properties of such representations depends on the dimensionality of the Clifford algebra. A Weyl representation supports a $\mathbb{Z}_2$ grading of the spinor space by chirality. Such a structure is only possible in an even number of dimensions [222]. Furthermore we know from gauge theory that when we introduce massless spin-1 bosons with charges the fermions of the theory must transform in a representation of the gauge group to ensure Poincaré invariance [44] (see also chapter 3). If the spinor in questions support grading by chirality the various spinors belonging to each grading can transform in different representations of the gauge group. If the right-handed and left-handed fermions transform in different representations, a Dirac mass term is not possible, and hence such chiral fermions must be massless (in the Dirac mass term sense) [44]. A seven-dimensional space does not support chirality, and therefore mass terms are allowed, and this spills over into the four-dimensional theory. This means that eleven-dimensional theories with seven compactified dimensions cannot lead to a four-dimensional theory with chiral fermions [223].

Let us now turn to the parallel development in supersymmetry and supergravity theory during the late 1970s and early 1980s. Supersymmetry was invented in the early 70s [224][225] (see [212] for further details). The first theories with local supersymmetry, the so called supergravity theories, appeared in 1975 and 1976 [226]. Supergravity theories were the first proper extension of GR ever invented, and promised a theory of gravity integrated with all other forces. Any kind of theory with local Poincaré invariance (and by extensions, local supersymmetry invariance) must be a theory of gravity. In the $\mathcal{N} = 1$ supersymmetry case, this involves the spin-2 graviton, and its supersymmetric partner, the spin-3/2 gravitino. In addition to the $\mathcal{N} = 1$ case, there are also supergravity theories with local extended supersymmetry. In this context we recall the general conclusion that, given some natural assumptions, a QFT with a finite number of fields cannot contain interacting fields of spin higher than two. Since supersymmetry results in multiplets with particles of several different spins this means that there are limits on the amount of supersymmetry any supergravity theory can have if it is to avoid fields of spin higher than two. In $D = 4$ this upper limit is $\mathcal{N} = 8$ and in $D = 11$ the upper limit is $\mathcal{N} = 1$. Thus, in 1978, Nham concluded that $D = 11$ is the largest possible dimensionality of a supergravity theory respecting this bound [227]. We see that $D = 4 + 7$ is again singled out. We also note that $D = 11, \mathcal{N} = 1$ is a unique theory where no extra fields can be added. With the Kaluza-Klein promise of deriving gauge symmetries from the isometry group of the compactified manifold, combinded with the uniqueness of $D = 11, \mathcal{N} = 1$ supergravity [227] one seemed to be on-track for a unified theory. This was

clearly a leading approach for a unified theory [213][228].

Theories in $D = 10$ dimensions were also considered, and here the situation is a bit different. In $\mathcal{N} = 1, D = 10$ supergravities we have already mentioned that extra vector and chiral supermultiplets can be added [141][229]. This theory is not unique since many gauge groups are a priori possible [213][229]. The $D = 10$ theory does however support chiral fermions. The chiral fermions in this case (at least those belonging to the vector multiplet) must now be in the same representation as the gauge bosons, which means they always have to be in the adjoint representation. Chiral theories always have the potential for anomalies because of parity violating interactions, and this possibility must be carefully checked. In $\mathcal{N} = 1, D = 10$ there is an additional anomaly stemming from the gravitational sector. Initially there seemed to be no way to achieve a theory of $D = 10, \mathcal{N} = N$ supergravity that did not contain anomalies. As we soon shall see the story undergoes an unexpected twist in 1984.

Before we get to 1984, we need to know a bit more about supergravity. Supergravity theories are non-renormalizable just like GR. However, renomalizability is only a relevant criteria for theories with UV-divergences. A theory which is finite does not need renormalization. The early days of supergravity research brought hope that supergravity theories could be finite. The question is not easy to answer, but it was shown that pure supergravity theories are indeed finite at both the one-loop and the two-loops level. Further work did not uphold the initial promise. Pure $\mathcal{N} = 1$ supergravity is divergent at the 3-loop level [230][230], and supergravity coupled to matter is divergent at one-loop level [231]. This was another major setback for the supergravity program. Several extended supergravity theories have yet to show any divergences, and may be finite [232], but although theories with extended supersymmetry are interesting for theoretical purposes, only theories with $\mathcal{N} = 1$ and $\mathcal{N} = 0$ can give chiral fermions which are necessary for realistic theories.

We have now sketched developments in Kaluza-Klein theory and supergravity theory in the years leading up to 1984, and we are now just about ready to make our historical point. The status in early 1984 is as follows. For several years, supergravity combined with Kaluza-Klein theory have been the leading approach for a unified theory combining gravity with all the other interactions [213]. However, serious difficulties are becoming apparent. The standard model gauge group can be generated from scratch in $D = 11, \mathcal{N} = 1$ supergravity, but this model does not permit chiral fermions. The standard model gauge group is compatible with $D = 10, \mathcal{N} = 1$ supergravity, and chiral fermions are also possible. In this case the gauge group is not generated by compactifications. For $D = 10$ supergravity the gauge group must be present already in ten dimensions, making the theory non-unique. In addition the theory always seemed to contain anomalies. Finally, all supergravity theories were non-renormalizable, and all of the realistic ones contained divergences. The problems seemed insurmountable, and the stage was now set for string theory.

In 1984 Green and Schwartz showed that $D = 10, \mathcal{N} = 1$ supergravity, with some extra terms (the "Chaplain-Manton terms"), was anomaly free for the gauge group $\mathrm{SO}(32)$ as well as for $\mathrm{E}_8 \times \mathrm{E}_8$ [233]. They also showed that the gauge group $\mathrm{SO}(32)$, and the necessary extra terms, could be uniquely derived from the theory of open superstrings [233]. Soon after, Gross, Harvey, Martinec and Rohm presented two new superstring theories, called heterotic string theories, which had $D = 10, \mathcal{N} = 1$ supergravity with gauge groups $\mathrm{SO}(32)$ or $\mathrm{E}_8 \times \mathrm{E}_8$ as their low-energy limit [234, 235, 236]. In addition, it was already strongly indicated that string theory could be finite order-by-order to all orders, and with these developments a new powerful paradigm took center stage.

The supergravity approach was in a sense dead, being both divergent and unable to predict the gauge group without eliminating the possibility of chiral fermions at the same time. Still, supergravity survived as the low-energy limit of superstring theory. The original Kaluza-Klein approach was also dead, but a gist of it survived in the string compactification program. The preferred compactification spaces were now three-dimensional complex manifolds (equivalent to

six real dimensions) called Calabi-Yau manifolds [237]. Calabi-Yau compactifications of heterotic string theories became the number one candidate for a unified theory. In the heterotic theory the gauge group was already present in the ten-dimensional theory, and not derived from the isometries of the compactified space. Even so, the compactified space still dictated important aspects of the low-energy theories. Just to name one such property, perhaps the most important one, the choice of Calabi-Yau manifolds dictated that all low-energy models had $\mathcal{N} = 1$ global supersymmetry [237]. This seemed to be a perfect match since at this time, and for many years later, there was a widespread and strong conviction that the UV-sensitivities of the Higgs mass must be cured by global $\mathcal{N} = 1$ supersymmetry broken just above the electroweak scale (see section 10.3 for more on this).

The decade after 1984 became a decade for the intense study of the properties of Calabi-Yau manifolds, as well as their relationship to four-dimensional physics. Many of the connections between string theory and mathematical research, emerged in this period. However, the influence of the supergravity and the Kaluza-Klein backstory did not end there. The research into supergravity and Kaluza-Klein theory had clearly established that gauge groups were very difficult to construct using the internal isometry approach [142, ch6]. It was also clear that producing chiral fermions was a major constraint on any unified model. When D-branes were introduced in string theory [238] one of its chief accomplishments was that it presented a novel way to introduce gauge groups and chiral fermions. In addition, the braneworld idea, which is the idea that low energy physics is limited to some hypersurface of the full ten-dimensional space, became a completely new way of reducing the effective dimensions of a ten-dimensional theory. Finally, with the introduction of M-theory in 1995 [239], which has $D = 11, \mathcal{N} = 1$ supergravity as its low-energy limit, the connection with the previous supergravity research came full circle, and the hope of a new unique eleven-dimensional theory was partially reawakened (see e.g. [240]).

9.4 Introduction to perturbative string theory

We are now ready to start studying string theory proper. We shall make a few introductory comments before we start, since string theory can be quite complex and often the view of the path is lost in all of the details.

For the bosonic and fermionic string we will proceed as follows. The first step will be postulating the fundamental action. Then we analyze the symmetries of this action. From the action we derive the equations of motions of all the involved degrees of freedom. Next we look at possible ways to simplify the action and the equations of motion by using the symmetries of the action. The global symmetries of the action will lead to conserved current which will lead to conserved charges. Some of the equation of motions will turn out to be constraint equations that can be expressed using conserved charges. The conserved charges are generators of global symmetries and they form closed Lie algebras. From this we move on to quantization.

We quantize the classical free string theories by canonical methods. As we are mostly seeking to establish the spectrum of each theory we will usually prefer to solve the constraint on the classical side, so as not to have to deal with unphysical states. The quantization process will yield specific restrictions on the dimensions of the string theory. The Lie algebra of constraint will be represented as an operator algebra on the quantum side. Following [241] we can summarize this approach as:

$$\text{Action} \rightarrow \text{Symmetries} \rightarrow \text{Current} \rightarrow \text{Algebra} \rightarrow \text{Constraints} \tag{9.4.1}$$

After having dealt with free theories and the spectrum using canonical methods we will switch to path integral methods for the interactions. It will turn out that the action of the free theory also determines all the interactions of the theory in the form of a perturbative S-matrix expansion. The

specific details of this perturbative expansion are responsible for string theory being finite loop by loop at all orders.

In conformal gauge (superconformal gauge) the world-sheet theory of string theories are conformal (superconformal) theories. A CFT description is a useful alternative and complement to the ordinary string description and we go through a few details to make this connection. Finally we use the CFT picture to make a sketch of the foundation of abstract perturbative string theory.

We will use natural units with $\hbar = c = 1$. Which means that $[E] = [M] = [L]^{-1} = [T]^{-1}$. This means that coordinates have dimension $[L]$, and derivatives w.r.t. coordinates have dimension $[L]^{-1}$ and the action is dimensionless. The metric and its determinant is dimensionless. For Newton's constant we set $\kappa = 8\pi G$. Our metrics will be mostly plus for Lorentzian spacetimes, and we will use Euclidean metrics for CFTs and path integrals.

9.5 Classical bosonic string theory

The string action

We will start out by analyzing string physics in (D+1)-dimensional Minkowski space, which we denote $M = (\mathbb{R}^{D+1}, \eta)$, with the metric $\eta = (-, +, \ldots, +)$. Let Q be some rectangle $Q \subset \mathbb{R}^2$. String theory starts with a map

$$X \colon Q \to M \tag{9.5.1}$$

$$X \colon (\tau, \sigma) \mapsto (X^0(\tau, \sigma), \ldots, X^D(\tau, \sigma)). \tag{9.5.2}$$

The map X is an embedding of the two-dimensional rectangle Q into M. When we need to be specific we will take $Q = [-t, t] \times [0, l]$. We call the image $X(Q)$ the world-sheet of the string. We let indices μ, ν take values $\mu, \nu \in \{0, \ldots, D\}$ and we let α, β take values $\alpha, \beta \in \{0, 1\}$. We set $\sigma^0 = \tau$ and $\sigma^1 = \sigma$. The movement of the classical version of the string is given by demanding that the action, given by the world-sheet area, is stationary. The Nambu-Goto action is the literal expression of this principle, but it is difficult to use for quantization. The Nambu-Goto action is given by [211]

$$S = -T \int d^2\sigma \sqrt{-\det(g)} = -T \int d^2\sigma \sqrt{-\det\left[\eta_{\mu\nu} \partial_\alpha X^\mu(\tau, \sigma) \partial_\beta X^\nu(\tau, \sigma)\right]} \tag{9.5.3}$$

Where g is the pullback of η by X given by

$$g = X^* \eta. \tag{9.5.4}$$

The symbol T is a dimensional constant called the string tension, usually given as $T = \frac{1}{2\alpha'}$. As the integral has dimensions $[L]^2$, T must have dimensions $[L]^{-2}$ for the action to be dimensionless.

The α' is the Regge slope parameter and has dimensions $[L]^2$. The α' parameter is the only free parameter in string theory. Sometimes it is convenient to also consider the string coupling g_s to be a free parameter, but this parameter is in principle determined by the expectation value of the dilaton. We will introduce the string coupling g_s and the dilaton in a later section. From the α' parameter we can define the string length $l_s = 2\pi\sqrt{\alpha'}$, and the string mass scale $M_s = 1/\sqrt{\alpha'}$. In principle α' is only restricted to be

$$10^{-35} m \leq \sqrt{\alpha'} \leq 10^{-19} m. \tag{9.5.5}$$

However, getting the correct low-energy gravity limit imposes that $\kappa^2 \sim l_s^{24} g_s^2$.

The Polyakov action is classically equivalent to the Nambu-Goto action, but it is better suited for use in quantization. The Polyakov action is given by [211]

$$S = -\frac{T}{2}\int d^2\sigma\sqrt{-h}h^{\alpha\beta}g_{\alpha\beta} \tag{9.5.6}$$

$$= -\frac{T}{2}\int d^2\sigma\sqrt{-h}h^{\alpha\beta}\partial_\alpha X^\mu(\tau,\sigma)\partial_\beta X^\nu(\tau,\sigma)\eta_{\mu\nu}. \tag{9.5.7}$$

Here h is a new metric variable called the world-sheet metric. This metric is independent of the pullback world-sheet metric we defined for the Nambu-Goto action. The Polyakov action is used for both open and closed strings. A closed string traces out a tube like world-sheet in spacetime, while an open string traces out a sheetlike world-sheet in the form of a surface with boundaries.

It is of great importance to understand the symmetries of the Polyakov action and their consequences. The classical Polyakov action is invariant under [211]:

- Global Poincaré transformations of X^μ
- Two-dimensional world-sheet diffeomorphisms
- Two-dimensional world-sheet Weyl transformations of h

Later we will use these symmetries for simplifying the action and its equations of motions.

The equations of motions

The equations of motions follow from this action by considering variation of the dynamical variables- The dynamic variables are the metric h, and the first order derivatives of the scalar fields X^μ. The Polyakov action doesn't depend on X^μ directly so the change in the action is given by [211]

$$\delta S = \int d^2\sigma\left\{\frac{\partial L}{\partial h^{\alpha\beta}}\delta h^{\alpha\beta} + \frac{\partial L}{\partial(\partial_\alpha X^\mu)}\delta(\partial_\alpha X^\mu)\right\} \tag{9.5.8}$$

$$= \int d^2\sigma\left\{\frac{\partial L}{\partial h^{\alpha\beta}}\delta h^{\alpha\beta} + \frac{\partial L}{\partial(\partial_\tau X^\mu)}\delta(\partial_\tau X^\mu) + \frac{\partial L}{\partial(\partial_\sigma X^\mu)}\delta(\partial_\sigma X^\mu)\right\}. \tag{9.5.9}$$

Setting $\delta S = 0$, and setting boundary terms to zero, we get the equations of motion [211]

$$\partial_\alpha\left(\sqrt{-h}h^{\alpha\beta}\partial_\beta X_\mu\right) = 0 \tag{9.5.10}$$

$$\partial_\alpha X^\mu\partial_\beta X^\nu\eta_{\mu\nu} - \frac{1}{2}h_{\alpha\beta}h^{\gamma\delta}\partial_\gamma X^\mu\partial_\delta X^\nu\eta_{\mu\nu} = 0. \tag{9.5.11}$$

There are three sets of boundary conditions that will set the boundary terms to zero (see table 9.1 for definitions). Open strings allow for two different types of boundary conditions called Neumann or Dirichlet boundary conditions, while closed strings must have periodic boundary conditions.

Boundary condition	**Expression**
Closed string - Periodic condition	$X^\mu(\tau,\sigma) = X^\mu(\tau,\sigma+l)$
Open string – Neumann condition	$\frac{\partial X^\mu}{\partial\sigma}\big\vert_{\sigma=0,l} = 0$
Open string – Dirichlet condition	$\frac{\partial X^\mu}{\partial\tau}\big\vert_{\sigma=0,l} = 0$

Table 9.1: Definitions of the three possible boundary conditions.

The Dirichlet type boundary condition was ignored in early string history as it breaks Poincaré invariance and therefore seemed unphysical, but in the current view the Dirichlet type condition is

very important. The acceptance of Dirichlet conditions stems from the fact that open string with Dirichlet boundary condition can be interpreted as strings with end points fixed to actual physical higher-dimensional non-perturbative objects called D-branes. We will discuss D-branes in section 9.13.

Solving the equations of motion

The equation of motions are complicated to solve in the form given in equations 9.5.10 and 9.5.11. We want to use the gauge freedom to simplify both the action and the equations of motion. This will be done by gauge fixing, which means choosing some representative of the gauge equivalence class in place of using the most general expression. Note that we should be careful not to rederive the equations of motion from a gauge fixed action. This could lead to eliminating proper dynamical degrees of freedom, which would change the theory into a different one.

We start by using diffeomorphism invariance to set

$$h^{\alpha\beta} = e^{\phi}\eta^{\alpha\beta}, \tag{9.5.12}$$

and by using the Weyl transformation we can set $\phi = 0$. This gauge is called the conformal gauge. The metric is now given as

$$h^{\alpha\beta} = \eta^{\alpha\beta}. \tag{9.5.13}$$

The Polyakov action in conformal gauge is [211]

$$S = -\frac{T}{2}\int d^2\sigma\eta^{\alpha\beta}\partial_\alpha X^\mu(\tau,\sigma)\partial_\beta X^\nu(\tau,\sigma)\eta_{\mu\nu}. \tag{9.5.14}$$

This action is invariant under conformal transformations. We will discuss this further in a later section. The partial set of equations of motions that can be derived from this action are [211]

$$\left(\frac{\partial^2}{\partial\sigma^2} - \frac{\partial^2}{\partial\tau^2}\right) X^\mu = 0. \tag{9.5.15}$$

We see that we are missing the degrees of freedom concerning the world-sheet metric $h_{\alpha\beta}$. The proper full set of equations of motion, derived by fixing the gauge in the original equations, are [211]

$$\left(\frac{\partial^2}{\partial\sigma^2} - \frac{\partial^2}{\partial\tau^2}\right) X^\mu = 0 \tag{9.5.16}$$

$$\partial_\alpha X^\mu \partial_\beta X^\nu \eta_{\mu\nu} - \frac{1}{2}h_{\alpha\beta}\partial_\gamma X^\mu \partial^\gamma X^\nu \eta_{\mu\nu} = 0 \tag{9.5.17}$$

We write this in short form as

$$\nabla^2 X^\mu = 0 \tag{9.5.18}$$

$$T_{\alpha\beta} = 0, \tag{9.5.19}$$

where we have defined the energy-momentum tensor $T_{\alpha\beta}$ by setting

$$T_{\alpha\beta} = \partial_\alpha X^\mu \partial_\beta X^\nu \eta_{\mu\nu} - \frac{1}{2}h_{\alpha\beta}\partial_\gamma X^\mu \partial^\gamma X^\nu \eta_{\mu\nu}. \tag{9.5.20}$$

For $T_{\alpha\beta}$ we can also deduce that it is traceless $T^\alpha_\alpha = 0$ and divergence free $\nabla^\alpha T_{\alpha\beta} = 0$. The equation $\nabla^2 X^\mu = 0$ is just the wave equation. The general solution of this equation can be written as

$$X^\mu(\tau,\sigma) = \sum_{n\in\mathbb{Z}} a_n^\mu(\tau) e^{in\frac{2\pi}{l}\sigma}. \tag{9.5.21}$$

We can set $a_0^\mu(\tau) = a_0^\mu + b_0^\mu \tau$, to get

$$X^\mu(\tau,\sigma) = a_0^\mu + b_0^\mu \tau + \sum_{n\in\mathbb{Z}\setminus\{0\}} \left(a_n^\mu(\tau) e^{-in\frac{2\pi}{l}(\tau-\sigma)} + \overline{a}_n^\mu(\tau) e^{in\frac{2\pi}{l}(\tau+\sigma)} \right). \tag{9.5.22}$$

A general solution can be written as

$$X^\mu(\tau,\sigma) = X_R^\mu(\tau-\sigma) + X_L^\mu(\tau+\sigma). \tag{9.5.23}$$

We define world-sheet light-cone coordinates by setting

$$\sigma^- = \tau - \sigma. \tag{9.5.24}$$
$$\sigma^+ = \tau + \sigma \tag{9.5.25}$$

We also set

$$a_0^\mu = x^\mu, \quad b_0^\mu = \alpha' p^\mu, \quad a_n^\mu = i\sqrt{\frac{\alpha'}{2}}\frac{\alpha_n^\mu}{n}, \quad \overline{a}_n^\mu = i\sqrt{\frac{\alpha'}{2}}\frac{\alpha_{-n}^\mu}{-n}. \tag{9.5.26}$$

We can now write the explicit form of the closed string solutions as [211]

$$X_R^\mu(\sigma^-) = \frac{1}{2}x^\mu + \frac{1}{2}\frac{2\pi}{l}\alpha' p^\mu \sigma^- + i\sqrt{\frac{\alpha'}{2}}\sum_{n\neq 0}\frac{1}{n}\alpha_n^\mu e^{-in\frac{2\pi}{l}\sigma^-} \tag{9.5.27}$$

$$X_L^\mu(\sigma^+) = \frac{1}{2}x^\mu + \frac{1}{2}\frac{2\pi}{l}\alpha' p^\mu \sigma^+ + i\sqrt{\frac{\alpha'}{2}}\sum_{n\neq 0}\frac{1}{n}\overline{\alpha}_n^\mu e^{-in\frac{2\pi}{l}\sigma^+}. \tag{9.5.28}$$

Sometimes an extra constant c^μ is introduce replacing the first terms x^μ by $x^\mu \pm c^\mu$. For open strings with von Neumann conditions at both ends we get

$$X^\mu(\tau,\sigma) = x^\mu + \frac{2\pi}{l}\alpha' p^\mu \tau + i\sqrt{2\alpha'}\sum_{n\neq 0}\frac{1}{n}\alpha_n^\mu e^{-in\frac{\pi}{l}\tau}\cos\left(\frac{n\pi\sigma}{l}\right). \tag{9.5.29}$$

For open strings with Dirichlet conditions at both ends we get

$$X^\mu(\tau,\sigma) = x_0^\mu + \frac{1}{l}(x_1^\mu - x_0^\mu)\sigma + i\sqrt{2\alpha'}\sum_{n\neq 0}\frac{1}{n}\alpha_n^\mu e^{-in\frac{\pi}{l}\tau}\sin\left(\frac{n\pi\sigma}{l}\right). \tag{9.5.30}$$

The constraint algebra

The conservation equation for the energy-momentum tensor using world-sheet light-cone coordinates are

$$\partial_- T_{++} = 0 \tag{9.5.31}$$
$$\partial_+ T_{--} = 0. \tag{9.5.32}$$

These equations make it possible to define an infinite set of conserved currents leading to an infinite set of conserved charges. This is done by integrating T_{--} or T_{++} over σ with arbitrary functions of σ. Selecting a set of exponential functions as basis functions, these charges can be expressed for closed strings as

$$L_n = -\frac{l}{4\pi^2}\int_0^l d\sigma e^{-\frac{2\pi i}{l}n\sigma} T_{--} = \frac{1}{2}\sum_m \alpha_{n-m}\alpha_m \tag{9.5.33}$$

$$\overline{L}_n = -\frac{l}{4\pi^2}\int_0^l d\sigma e^{+\frac{2\pi i}{l}n\sigma} T_{++} = \frac{1}{2}\sum_m \overline{\alpha}_{n-m}\overline{\alpha}_m. \tag{9.5.34}$$

The charges form an algebra called the Witt algebra (or the conformal algebra) [242] given by

$$\{L_m, L_n\} = -i(m-n)L_{m+n} \tag{9.5.35}$$
$$\{\overline{L}_m, \overline{L}_n\} = -i(m-n)\overline{L}_{m+n} \tag{9.5.36}$$
$$\{\overline{L}_m, L_n\} = 0 \tag{9.5.37}$$

To be precise, equation 9.5.35 and equation 9.5.36 both define the Witt algebra. This means that for closed strings there are two copies of the Witt algebra and the two copies commute with each other. Open strings only have one copy of the Witt algebra. For open strings we write

$$L_n = -\frac{l}{2\pi^2}\int_0^l d\sigma \left\{e^{-\frac{\pi i}{l}n\sigma}T_{--} + e^{+\frac{\pi i}{l}n\sigma}T_{++}\right\} \tag{9.5.38}$$
$$= \frac{1}{2}\sum_m \alpha_{n-m}\alpha_m \tag{9.5.39}$$

and

$$\{L_m, L_n\} = -i(m-n)L_{m+n}. \tag{9.5.40}$$

9.6 Quantized bosonic string theory

We quantize the bosonic string by using canonical methods. This involves the usual Lie algebra morphism from the classical Poisson bracket to the quantum Poisson bracket (see section 8.3 of chapter 8 for more on quantization). If the constraints are not solved before quantization they must be imposed as restrictions on the Hilbert space afterwards to properly define the physical states (see sections 8.2 and 8.3 for more on constraints and their quantization). It is essential that non-physical states decouple from the theory when the constraints are imposed. This requirement leads to an important restriction on the theory. Non-physical states only decouple when the so called normal ordering constant is equal to one and the total dimensions of the string spacetime is $D+1 = 26$ (we will define the normal ordering constant in a moment).

Our primary aim in this section is to establish the spectrum of states of the theory. We therefore want to utilize light-cone gauge quantization. In light-cone gauge the constraints can be explicitly solved. Thus, the physical spectrum of the bosonic string is most easily analyzed in light-cone gauge.

The quantized constraint algebra

String theory is a constrained theory. The Witt algebra of classical constraints must be implemented on the Hilbert space to define the proper physical Hilbert space. The generators of the algebra are expressed as products of oscillators. Classically these oscillators commute, but in the quantum theory these oscillators are in general non-commuting operators. For those generators that are expressed as products of commuting operators we can define normal ordered quantum version without any changes. However, for L_0 the operators of the product does not commute. We must therefore replace L_0 by $L_0 + a$ where the extra constant factor a is called the normal ordering constant [142][211]. It accounts for the noncommutativity of its factors. The algebra in the quantum case is therefore not the Witt algebra but the Virasoro algebra [242]. The Virasoro algebra is a central extension of Witt algebra. The definition of the Virasoro algebra is given by

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{c}{12}m(m^2-1)\delta_{m+n}$$
$$[L_m, c] \quad = 0,$$

where c is an additional algebra basis element (the central charge), that commutes with all other basis elements. For the closed string there are two copies of the Virasoro algebra, corresponding to left-movers and right-movers, and the two copies commute with each other. For the open string there is only one copy of the Virasoro algebra.

Closed string states

To uncover the spectrum of states we build the Fock space of physical states by acting on the ground state with creation operators. Note that we are now using quantum oscillators as defined for the light-cone gauge. These oscillators are denoted by α_m^j where $j \in \{2, \ldots, d-1\}$. The integer $d = D + 1$ is the total dimension of spacetime. This means that there are $(d-2)$ values of the index j. We start by defining the ground state by

$$\alpha_m^j \left|0, p^\mu\right\rangle = \overline{\alpha}_m^j \left|0, p^\mu\right\rangle = 0 \ (m > 0), \quad \hat{p}^\mu \left|0, p^\mu\right\rangle = p^\mu \left|0, p^\mu\right\rangle . \tag{9.6.1}$$

The excited states of the closed string are generated from the ground state by acting with one or more $\alpha_{n_k}^i$ and one or more $\overline{\alpha}_{m_l}^j$ such that

$$n_k < 0 \wedge m_l < 0 \wedge \sum_k n_k = \sum_l m_l. \tag{9.6.2}$$

(Note that each $\alpha_{n_k}^i$ will have an independent index of same type as i. These will be denoted by $i, j, k \ldots$ etc.) Using this logic we see that the first excited state is given by

$$\alpha_{-1}^i \overline{\alpha}_{-1}^j \left|0, p^\mu\right\rangle . \tag{9.6.3}$$

The second level excited states are given by

$$\alpha_{-2}^i \overline{\alpha}_{-2}^j \left|0, p^\mu\right\rangle , \quad \alpha_{-1}^i \alpha_{-1}^k \overline{\alpha}_{-2}^l \left|0, p^\mu\right\rangle , \quad \alpha_{-1}^i \alpha_{-1}^j \overline{\alpha}_{-1}^k \overline{\alpha}_{-1}^l \left|0, p^\mu\right\rangle , \quad \alpha_{-2}^i \overline{\alpha}_{-1}^j \overline{\alpha}_{-1}^k \left|0, p^\mu\right\rangle . \tag{9.6.4}$$

The mass of a closed string state is determined by the relation

$$M^2 = \frac{4}{\alpha'}(N-1), \tag{9.6.5}$$

where N is a non-negative integer denoting the excitation level of the state. The number N is defined by

$$N = -\sum_k n_k = -\sum_l m_l. \tag{9.6.6}$$

We see from this mass relation that the lowest lying state of the spectrum ($N = 0$) is a scalar tachyon. The first excited state ($N = 1$) is massless, and can be decomposed as a symmetric tensor, an antisymmetric tensor and a scalar. In addition there is an infinite set of massive excitations ($N > 1$) with masses proportional to positive integer multiples of m_s. The lowest lying states for closed bosonic string theory are listed in table 9.2.

Open string states

The excited states of the open string are generated from the ground state by acting with one or more $\alpha_{n_k}^i$ which are such that

$$n_k < 0. \tag{9.6.7}$$

State	Description	Expression	$\alpha' M^2$
$N = 0$	Scalar tachyon	$\|0; p^\mu\rangle$	-4
$N = 1$	Graviton, antisymmetric tensor, dilaton	$\alpha^i_{-1}\alpha^j_{-1}\|0; p^\mu\rangle$	0
$N = 2$	Various massive states	See the text	4

Table 9.2: Lowest lying states of bosonic closed string theory

Using this logic we see that the first excited state is given by

$$\alpha^i_{-1}\,|0, p^\mu\rangle\,. \tag{9.6.8}$$

The second level of excited states are given by

$$\alpha^i_{-2}\,|0, p^\mu\rangle\,, \quad \alpha^i_{-1}\alpha^j_{-1}\,|0, p^\mu\rangle\,. \tag{9.6.9}$$

The mass of the open string states is determined by the relation

$$M^2 = \frac{1}{\alpha'}(N-1), \tag{9.6.10}$$

where N is a non-negative integer denoting the excitation level. The number N is defined by

$$N = -\sum_k n_k. \tag{9.6.11}$$

We see from the mass relation the the lowest states of the open string is a scalar tachyon just like we found for the closed string. The first excited states is a massless vector boson. In addition, as for the closed string, there is an infinite set of massive excitations with masses proportional to a positive integer multiple of m_s. The lowest lying states of open string theory are listed in table 9.3.

State	Description	Expression	$\alpha' M^2$
$N = 0$	Tachyon	$\|0; p^\mu\rangle$	-1
$N = 1$	Massless vector boson	$\alpha^i_{-1}\,\|0; p^\mu\rangle$	0
$N = 2$	Various states	See the text	1

Table 9.3: Lowest lying states of bosonic open string theory

Because of the presence of the tachyon in the spectrum the vacuum is considered to be the wrong vacuum. There are also no fermions in bosonic string theory, and therefore the bosonic string is considered more of a useful toy model. The bosonic string also plays a role in defining the heterotic string theories.

9.7 Classical superstring theories

We have seen that the bosonic string theory contains tachyons and is also unable to represent fermions, both of which makes it unsuitable as a candidate theory for actual physics. To solve the problems of the bosonic string extra fermionic degrees of freedom are added to the string in the

form of (1+1)-dimensional spinor-fields propagating on the string. To achieve the intended problem fixing, the bosonic and fermionic degrees of freedom must be related by supersymmetry. There are two common formalisms to introduce supersymmetry on the string. In the Ramond-Neveu-Schwartz (RNS) formulation one introduces a model with manifest world-sheet supersymmetry, while in the Green-Schwartz (GS) formulation one introduces a model with manifest spacetime supersymmetry. It can be shown that they both have world-sheet and spacetime supersymmetry (after GSO projection, see later sections). We focus only on the RNS formulation.

The string action

In the bosonic string model we used an action that could be seen as scalar fields coupled to two-dimensional gravity. In the fermionic case we will have the action of a supersymmetric matter multiplet coupled to two-dimensional supergravity [243][244]. In detail the action that generalizes the bosonic Polyakov action is the action of a chiral supermultiplet (X^μ, ψ^μ, F^μ) coupled to the $\mathcal{N} = 1$ supergravity multiplet $(e^a_\alpha, \chi_\alpha, A)$ in $D = 2$. Here F and A are an auxiliary real scalar field required for off-shell closure. The ψ is a Majorana fermion, and χ is a spin-3/2 gravitino. The gravitino is the superpartner of the graviton represented by the zwei-bein e.

To give a presentation of the complete action requires a detour into supergravity theory and its notation that we want to avoid. We will soon give a simpler gauge fixed action that can derived from the full supergravity action. In symbolic form the complete action is given by

$$S = \int d^2\sigma \; L(X^\mu, \psi^\mu, e^a_\alpha, \chi_\alpha). \tag{9.7.1}$$

The symmetries of this action are as follows [211][243]:

- Global spacetime Poincaré invariance
- Two-dimensional world-sheet diffeomorphism invariance
- Local two-dimensional world-sheet supersymmetry
- Weyl invariance
- Super Weyl invariance
- Local world-sheet Lorentz invariance

The equations of motion

From the full action we derive the equations of motions for all physical fields. Since there are four different field, given by $(X^\mu, \psi^\mu, e^a_\alpha, \chi_\alpha)$, there will be four sets of equations. Again the we only give the symbolic version of the equations by writing

$$X : S_1(X^\mu, \psi^\mu, e^a_\alpha, \chi_\alpha) = 0 \tag{9.7.2}$$
$$\psi : S_3(X^\mu, \psi^\mu, e^a_\alpha, \chi_\alpha) = 0 \tag{9.7.3}$$
$$e : S_3(X^\mu, \psi^\mu, e^a_\alpha, \chi_\alpha) = 0 \tag{9.7.4}$$
$$\chi : S_4(X^\mu, \psi^\mu, e^a_\alpha, \chi_\alpha) = 0. \tag{9.7.5}$$

For our actual computations we will be using the superconformal gauge. In this gauge the zwei-bein and the gravitino degrees of freedom decouple and we are left with the action given by [211][243]

$$S = -\frac{1}{8\pi} \int d^2\sigma \; \left\{ \frac{2}{\alpha'} \partial_a X^\mu \partial^a X^\nu \eta_{\mu\nu} + 2i\overline{\psi}^\mu \gamma^a \partial_a \psi_\mu \right\}. \tag{9.7.6}$$

This action is identical to the action of a free chiral superfield in two dimensions. We note that $[X] = -1$,$[\psi] = 1/2$,$\left[d^2\sigma\right] = -2$ and $[\alpha'] = 2$ which makes $[S] = 0$. The γ^a are 2×2 gamma

matrices, and we are using two-dimensional Dirac-Majorana spinors which we can write using two Weyl-Majorana spinors ψ_+^μ, ψ_-^μ as

$$\psi^\mu = \begin{pmatrix} \psi_+^\mu \\ \psi_-^\mu \end{pmatrix}. \tag{9.7.7}$$

The ψ_+^μ is referred to as left-movers and ψ_-^μ as right-movers.

To make sure that all proper dynamic degrees of freedom are included the equations of motions have to be derived by gauge fixing the original equations of motion instead of deriving them from this gauge fixed action. We get the following result [211]:

$$\begin{aligned} X &: \partial_\alpha \partial^\alpha X^\mu = 0 \\ \psi &: \gamma^\alpha \partial_\alpha \psi \quad = 0 \\ \chi &: T_{F\alpha} \qquad = 0 \\ e &: T_{\alpha\beta} \qquad = 0. \end{aligned}$$

In these equation we have defined the supercurrent

$$T_{F\alpha} = -\frac{1}{4}\sqrt{\frac{2}{\alpha'}}\gamma^\beta \gamma_\alpha \psi^\mu \partial_\beta X_\mu, \tag{9.7.8}$$

and the stress-energy tensor

$$T_{\alpha\beta} = -\frac{1}{\alpha'}\left(\partial_\alpha X^\mu \partial_\alpha X_\mu - \frac{1}{2}\eta_{\alpha\beta}\partial^\alpha X^\mu \partial_\alpha X^\mu\right) - \frac{i}{4}\left(\overline{\psi}^\mu \gamma_\alpha \partial_\beta \psi^\mu + \overline{\psi}^\mu \gamma_\beta \partial_\alpha \psi^\mu\right). \tag{9.7.9}$$

We consider the equations derived for the gravitino and the metric to represent constraint equations as the action does not contain derivatives for these fields. We see that the part of $T_{\alpha\beta}$ involving just the bosonic fields is the same as for the bosonic string. Both $T_{F\alpha}$ and $T_{\alpha\beta}$ satisfy current conservation laws:

$$\partial^\alpha T_{F\alpha} = 0 \tag{9.7.10}$$

$$\partial^\alpha T_{\alpha\beta} = 0. \tag{9.7.11}$$

These conserved currents will lead to an infinite set of conserved charges as we will see in a minute.

Solving the equations of motion

To simplify the equations of motion we use world-sheet light-cone coordinates and we get the equations [211]

$$\begin{aligned} X: &\quad \partial_+ \partial^- X^\mu = 0 \\ \psi: &\quad \partial_- \psi_+^\mu = 0 \\ &\quad \partial_+ \psi_-^\mu = 0 \\ \chi: &\quad T_{F+} = -\frac{1}{2}\sqrt{\frac{2}{\alpha'}}\psi_+^\mu \partial_+ X_\mu = 0 \\ &\quad T_{F-} = -\frac{1}{2}\sqrt{\frac{2}{\alpha'}}\psi_-^\mu \partial_- X_\mu = 0 \\ e: &\quad T_{++} = -\frac{1}{\alpha'}\partial_+ X^\mu \partial_+ X_\mu - \frac{i}{2}\psi_+^\mu \cdot (\partial_+ \psi^\mu) = 0 \\ &\quad T_{--} = -\frac{1}{\alpha'}\partial_- X^\mu \partial_- X_\mu - \frac{i}{2}\psi_-^\mu \cdot (\partial_- \psi_-^\mu) = 0 \\ &\quad T_{+-} = T_{-+} = 0. \end{aligned}$$

The $\cdot$ signifies summation over the μ index. We now need to specify what boundary conditions apply. For the bosonic parts the boundary conditions are as for the bosonic string. For the fermionic part of closed strings we can have

$$\psi_+^\mu(\tau,\sigma) = \pm\psi_+^\mu(\tau,\sigma+l) \tag{9.7.12}$$
$$\psi_-^\mu(\tau,\sigma) = \pm\psi_-^\mu(\tau,\sigma+l). \tag{9.7.13}$$

The $\pm$ options are naturally referred to as periodic (for $+$) and anti-periodic (for $-$) boundary conditions. By using all combinations we get a total of four different boundary conditions which we will denote as a two-tuple with elements from the set $\{NS, R\}$. The first element of the pair denotes the option chosen for the right-movers (ψ_-), and the second elements of the pair denotes the option chosen for the left-movers (ψ_-). The letters NS denote the equal sign option (the plus sign in the above equations), and the letter R denotes the opposite sign option (the minus sign in the above equations). The four possibilities are written as (NS, NS), (NS, R), (R,NS) and (R,R).

For the fermionic part of open strings we can have

$$\psi_+^\mu(\tau,0) = \pm\psi_-^\mu(\tau,0) \tag{9.7.14}$$
$$\psi_+^\mu(\tau,l) = \pm\psi_-^\mu(\tau,l). \tag{9.7.15}$$

Note that these equations relate right-movers and left-movers (unlike in the closed string case). The overall sign between right-movers and left-movers can be fixed by convention. We therefore decide to set the sign convention at the $(\sigma = 0)$-end as

$$\psi_+^\mu(\tau,0) = \psi_-^\mu(\tau,0). \tag{9.7.16}$$

The possibilities at the $(\sigma = l)$-end, denoted as (NS) and (R), are then

$$\psi_+^\mu(\tau,l) = -\psi_-^\mu(\tau,l) \quad \text{(NS)} \tag{9.7.17}$$
$$\psi_+^\mu(\tau,l) = +\psi_-^\mu(\tau,l) \quad \text{(R)}. \tag{9.7.18}$$

We are now ready to write down the mode expansion for the fermionic fields. For the closed string we get [211]

$$\psi_+^\mu(\tau,\sigma) = \sqrt{\frac{2\pi}{l}} \sum_{r\in\mathbb{Z}+\frac{1}{2}} \bar{b}_r^\mu e^{-\frac{2\pi}{l}ir(\tau+\sigma)} \quad \text{(NS+)} \tag{9.7.19}$$

$$\psi_-^\mu(\tau,\sigma) = \sqrt{\frac{2\pi}{l}} \sum_{r\in\mathbb{Z}+\frac{1}{2}} b_r^\mu e^{-\frac{2\pi}{l}ir(\tau-\sigma)} \quad \text{(NS-)} \tag{9.7.20}$$

$$\psi_+^\mu(\tau,\sigma) = \sqrt{\frac{2\pi}{l}} \sum_{r\in\mathbb{Z}} \bar{b}_r^\mu e^{-\frac{2\pi}{l}ir(\tau+\sigma)} \quad \text{(R+)} \tag{9.7.21}$$

$$\psi_-^\mu(\tau,\sigma) = \sqrt{\frac{2\pi}{l}} \sum_{r\in\mathbb{Z}} b_r^\mu e^{-\frac{2\pi}{l}ir(\tau-\sigma)} \quad \text{(R-)} \tag{9.7.22}$$

where the right-movers and the left-movers can be picked independently from the (NS) or (R)

variety. For the open string we get [211]

$$\psi_{\pm}^{\mu}(\tau,\sigma) = \sqrt{\frac{\pi}{l}} \sum_{r\in\mathbb{Z}+\phi} b_r^{\mu} e^{-\frac{\pi}{l} ir(\tau\pm\sigma)} \quad \text{(NN)} \tag{9.7.23}$$

$$\psi_{\pm}^{\mu}(\tau,\sigma) = \pm\sqrt{\frac{\pi}{l}} \sum_{r\in\mathbb{Z}+\phi} b_r^{\mu} e^{-\frac{\pi}{l} ir(\tau\pm\sigma)} \quad \text{(DD)} \tag{9.7.24}$$

$$\psi_{\pm}^{\mu}(\tau,\sigma) = \sqrt{\frac{\pi}{l}} \sum_{r\in\mathbb{Z}+\phi} b_r^{\mu} e^{-\frac{\pi}{l} ir(\tau\pm\sigma)} \quad \text{(ND)} \tag{9.7.25}$$

$$\psi_{\pm}^{\mu}(\tau,\sigma) = \pm\sqrt{\frac{\pi}{l}} \sum_{r\in\mathbb{Z}+\phi} b_r^{\mu} e^{-\frac{\pi}{l} ir(\tau\pm\sigma)} \quad \text{(DN)}, \tag{9.7.26}$$

where the right-movers and the left-movers must be from the same (NS) or (R) variety. The (NS) conditions are implemented by $\phi = 1/2$ and the (R) condition by $\phi = 0$. The labels (NN), (DD), (ND), and (DN) indicate the selected combination 0f Neumann and Dirichlet boundary conditions. Except for the signs these four equations are identical. Some authors separate them by placing different upper indices on the fields $\psi_{\pm}^{\mu}$ the oscillators b_r^{μ} (see e.g. [211]).

The constraint algebra

The equations for the conservation of the supercurrent and the energy-momentum tensor in world-sheet light-cone coordinates are

$$\partial_- T_{++} = 0 \tag{9.7.27}$$

$$\partial_+ T_{--} = 0 \tag{9.7.28}$$

$$\partial_- T_{F+} = 0 \tag{9.7.29}$$

$$\partial_+ T_{F-} = 0 \tag{9.7.30}$$

These equations make it possible to define an infinite set of conserved currents leading to an infinite set of conserved charges. This is done by integrating T_{--}, T_{++}, T_{F+} and T_{F-} over σ with arbitrary functions of σ. Selecting a set of exponential functions as basis functions, the conserved charges dervied from the supercurrent and energy-momentum tensor are defined for closed string as [211]

$$L_m = -\frac{l}{4\pi^2} \int_0^l d\sigma e^{-\frac{2\pi}{l} im\sigma} T_{--}(\sigma) \tag{9.7.31}$$

$$= \frac{1}{2} \sum_{m\in\mathbb{Z}} \alpha_{n-m}\alpha_m + \frac{1}{2} \sum_{r\in\mathbb{Z}+\phi} \left(r + \frac{n}{2}\right) b_{-r} b_{n+r} \tag{9.7.32}$$

$$G_r = -\frac{1}{\pi}\sqrt{\frac{l}{2\pi}} \int_0^l d\sigma e^{-\frac{2\pi}{l} ir\sigma} T_{F-}(\sigma) \tag{9.7.33}$$

$$= \sum_m \alpha_{-m}\alpha_{r+m} \quad \text{with} \quad r \in \mathbb{Z}+\phi. \tag{9.7.34}$$

Corresponding definition apply for $\overline{L}$ and $\overline{G}$ based on T_{++} and T_{F+}. The charges satisfy the super-Witt algebra [245]

$$\begin{aligned}
\{L_m, L_n\} &= -i(m-n)L_{m+n} \quad \text{with} \quad m,n \in \mathbb{Z} \\
\{L_m, G_r\} &= -i\left(\frac{1}{2}m - r\right) G_{m+r} \\
\{G_r, G_s\}_+ &= -2i L_{r+s} \quad \text{with} \quad r,s \in \mathbb{Z}+\phi.
\end{aligned}$$

There is a corresponding algebra involving $\overline{L}$ and $\overline{G}$ for the right-movers. For the open string there is just one such algebra. Note that the super-Witt algebra is a Lie superalgebra [246][247] where the L_m are in the even part, and the G_r are in the odd part.

9.8 Heterotic string theories

At the time of the Green and Schwartz anomaly cancellation result in 1984 there were only three known consistent superstring theories, type I, type IIA and type IIB. The Green and Schwartz results suggested that there might exists an additional string theory with gauge group $E_8 \times E_8$. This presumption was confirmed the very next year, when the heterotic string theories were discovered [234, 235, 236]. We will not treat the construction of heterotic string theories in any detail (see [211] or [208] for textbook treatments). We will simply declare that heterotic strings a hybrid of right moving bosonic degrees of freedom and left-moving fermionic degrees, two additional theories called the $\mathrm{E}_8 \times \mathrm{E}_8$ heterotic and the SO(32) heterotic string theory are possible. Since superstrings live in $D = 10$ and bosonic strings lives in $D = 26$ this may sound peculiar. The explanation is that the bosonic string is compactified in a non-geometric way on a particular torus defined by a lattice. This is most easily understood in the CFT framework but we will not present this.

The introduction of the heterotic string theories was especially important since they realized the gauge group $E_8 \times E_8$. Several interesting gauge groups can be embedded as subgroups of E_8. Among them are the groups E_6, $\mathrm{SO}(10)$ and $\mathrm{SU}(5)$ which are all attractive from a grand unified theory perspective. Another important aspect of heterotic theories were they way they non-geometrically "compactified" the 16-dimensional "bosonic" dimensions. This non-geometric compactification scheme was generalized to create new string theories defined directly in four dimensions [211].

9.9 Quantized superstring theories

We proceed with quantization of superstring theory in the same manner as we did for the bosonic string. Again, we want to use light cone gauge to avoid having to deal with constraints. Our main interest is to establish the spectrum of the theories. To decouple ghost as ensure consistency the fermionic string theory require the (NS) normal ordering constant a_{NS} to be $-1/2$, and the total spacetime dimension of the string must be $d = D + 1 = 10$ [211]. For the (R) sector the normal ordering constant a_R is zero [211].

The quantized constraint algebra

Analogous to the bosonic string theory, in superstring theory the super-Witt algebra of classical constraints must be implemented on the Hilbert space of the quantum theory to define the proper physical Hilbert space. Again there are issues of ordering leading the necessity for a normal ordering constant. The algebra in the quantum case is therefore not the super-Witt algebra but the super-Virasoro algebra. The super-Virasoro algebra is a central extension of super-Witt algebra (it can also be considered to be a Lie superalgebra extensions of the Virasoro algebra). In the classical case we have two distinct super-Witt algebras, the Ramond (R) and the Neveu-Schwartz (NS) version. Corresponding to this there are two distinct super-Virasoro algebras. The definition

of the super-Virasoro algebra is given by

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{d}{8}m(m^2 - 2\phi)\delta_{m+n} \tag{9.9.1}$$

$$[L_m, G_r] = \left(\frac{1}{2}m - r\right) G_{m+r} \tag{9.9.2}$$

$$\{G_r, G_s\} = 2L_{r+s} + \frac{d}{2}\left(r^2 - \frac{\phi}{2}\right)\delta_{r+s} \tag{9.9.3}$$

where $\phi = 1/2$ corresponds to the NS version and $\phi = 0$ corresponds to the R version.

Open strings states

We will begin be constructing open string states. Note that the construction of the type I open string theory is a separate question, that will be dealt with later. The open string states we will define now are not identical with the type I open string states.

Before we start to build up the Fock space, we must define the ground state (or states). The definition of the ground states is the same for both open and closed strings. We first define the (NS) ground state by

$$\alpha_m^j \left|0, p^\mu\right\rangle_{NS} = b_r^j \left|0, p^\mu\right\rangle_{NS} = 0 \quad m \in \{1, 2, \ldots\}\, r \in \{1/2, 3/2, \ldots\} \tag{9.9.4}$$

The (R) ground state is defined by

$$\alpha_m^j \left|0, p^\mu\right\rangle_{R} = b_r^j \left|0, p^\mu\right\rangle_{R} = 0 \quad m \in \{1, 2, \ldots\}\, r \in \{1, 2, \ldots\} \tag{9.9.5}$$

From hereon we will suppress the p^μ information from the ground states. The (R) sector ground state will turn out to be a ten-dimensional spinor. It can therefore be decomposed into two Weyl spinors with opposite chirality. We write them as $\left|a\right\rangle_R$ and $\left|\dot{a}\right\rangle_R$.

In the fermionic string there are several sectors. We consider the (NS) sector for open strings. The first excited state in the open string NS sector is given by

$$b_{-1/2}^i \left|0\right\rangle_{NS} \tag{9.9.6}$$

The second level of exited states are given by

$$b_{-1/2}^i b_{-1/2}^j \left|0\right\rangle_{NS} \quad \alpha_{-1}^i \left|0\right\rangle_{NS}. \tag{9.9.7}$$

The NS sector mass relation is

$$\alpha' M^2 = N_B + N_{NS} - \frac{1}{2}. \tag{9.9.8}$$

We see that the lowest lying state is a tachyon, and that first exited state is a massless vector boson. The higher states are massive.

The first excited states of the (R) sector are given by

$$b_{-1}^i \left|0\right\rangle_R \quad \alpha_{-1}^i \left|0\right\rangle_R. \tag{9.9.9}$$

The R sector mass relation is

$$\alpha' M^2 = N_B + N_R. \tag{9.9.10}$$

We see that the lowest lying states are massless, and that the first excited states are massive.

All in all the massless and tachyonic states define a set of four different states (see table 9.4) that we label by class and parity as (NS+)(NS-)(R+)(R-). The pluses and minus refer to parity and not to right-movers and left-movers. Note that (NS-) is a higher mass-level state than all the others.

Class	**Description**	**Expression**	$\alpha' M^2$	$(-1)^F$
(NS)				
	scalar	$\|0\rangle$	$-\frac{1}{2}$	-1
	vector	$b^i_{-1/2}\|0\rangle$	0	$+1$
(R)				
	spinor	$\|a\rangle$	0	$+1$
	spinor	$\|\dot{a}\rangle$	0	-1

Table 9.4: Lowest lying states of fermionic open string theory.

Closed string states

We now proceed to analyze the states of the the closed string. The closed string states can be build by tensoring two open string states. The first element of the tensor product represents the left-mover and the second element represents the right-mover. There are four different open string states denoted by class and parity as (NS+)(NS-)(R+)(R-). This results in a possible set of $4 \times 4 = 16$ different combinations. However, we must exclude the six combinations of the (NS-) state with the three others states since this would involve combining different mass-level states (which is not permitted because of the mass-level matching condition). This results in $16 - 6 = 10$ different combinations (see table 9.5 for a listing of the lower lying states).

Class	**Description**	**Expression**	$\alpha' M^2$	$(-1)^{\overline{F}}$	$(-1)^F$
(NS, NS)					
	scalar tachyon	$\|0\rangle \otimes \|0\rangle$	-2	-1	-1
	graviton, ast-2, dilaton	$b^i_{-1/2}\|0\rangle \otimes b^i_{-1/2}\|0\rangle$	0	$+1$	$+1$
(R, R)					
	scalar, ast-2, ast-4	$\|a\rangle \otimes \|c\rangle$	0	$+1$	$+1$
	scalar, ast-2, ast-4	$\|\dot{a}\rangle \otimes \|\dot{c}\rangle$	0	-1	-1
	vector, ast-3	$\|\dot{a}\rangle \otimes \|c\rangle$	0	-1	$+1$
	vector, ast-3	$\|a\rangle \otimes \|\dot{c}\rangle$	0	$+1$	-1
(R, NS)					
	dilatino, gravitino	$\|a\rangle \otimes b^i_{-1/2}\|0\rangle$	0	$+1$	$+1$
	dilatino, gravitino	$\|\dot{a}\rangle \otimes b^i_{-1/2}\|0\rangle$	0	-1	$+1$
(NS, R)					
	dilatino, gravitino	$b^i_{-1/2}\|0\rangle \otimes \|a\rangle$	0	$+1$	$+1$
	dilatino, gravitino	$b^i_{-1/2}\|0\rangle \otimes \|\dot{a}\rangle$	0	$+1$	-1

Table 9.5: Lowest lying states of fermionic closed string theory. Abbrevations: ast-n = antisymmetric tensor field of rank n.

We will not do a separate analysis for the heterotic string theories.

GSO projection

The spectrum of the supersymmetric string theories we have presented still contain a tachyon state and the spectrum is not supersymmetric. This can be fixed by a projection invented by Gliozzi, Scherk and Olive (GSO) in 1976 [248]. The GSO projection originated as a technique to fix the spectrum and eliminate the tachyon, but it can be derived from consistency requirements of the quantum theory [249]. The consistency requirement that leads to the GSO projections is the requirements of modular invariance [211, Ch9]. Supersymmetry of the spectrum can then be seen as a consequence of modular invariance. The GSO projection is based on selecting out states based on the value of the fermion number operators. The operators $(-1)^F$ and $(-1)^{\overline{F}}$ counts number of fermions for right-movers and left-movers of the closed string. In the open string sector $(-1)^F$ does the same thing.

There are two different GSO projections for the closed string. The GSO projection defined by requiring $(-1)^F = 1$ and $(-1)^{\overline{F}} = 1$ defined the type IIB theory. The GSO projection defined by requiring that $(-1)^F = 1$ for states from the (NS,NS) sector, $(-1)^F = -1$ and $(-1)^{\overline{F}} = -1$ for the (R,R),$(-1)^{\overline{F}} = -1$ for the (R,NS) sector, and $(-1)^F = -1$ for the (NS,R) sector defines the type IIA theory.

We summarize the properties of the spectrum of the various theories in the following tables.

Sector	**IIA**	**IIB**	**I**	**Heterotic**
NS-NS	$G_{\mu\nu}, B_{\mu\nu}, \phi$	$G_{\mu\nu}, B_{\mu\nu}, \phi$	$G_{\mu\nu}, \phi$, A^a_μ	$G_{\mu\nu}, B_{\mu\nu}, \phi, A^a_\mu$
NS-R	$\chi^1, \chi^2, \lambda^1, \lambda^2$	$\chi^1, \chi^2, \lambda^1, \lambda^1$	χ^1, λ^1, ψ	χ^1, λ^1, ψ
R-R	C_1, C_3	C_0, C_2, C_4	C_2	-

Table 9.6: Brief listing of the massless spectrum for the various supersymmetric string theories.

Type	**Open /closed**	**Oriented**	**Chiral**	**LE-limit**	**Gauge group**	**Branes**
I	both	no	yes	N=1	SO(32)	(1, 5)
IIa	closed	yes	no	N=2A	–	(0,6), (2,4), (8)
IIb	closed	yes	yes	N=2B	–	(-1,7), (1,5), (3,3), (9)
Het SO(32)	closed	yes	yes	N=1	SO(32)	none
Het $E_8 \times E_8$	closed	yes	yes	N=1	$E_8 \times E_8$	none

Table 9.7: Properties of string theories

Type I string theory

Beyond the GSO projection which we used to define type IIA and IIB string theories, further projections are also possible. Starting with IIB string theory, we can use a projection based on the operator Ω which exchanges the right-moving sector with the left-moving sector. We can then define the spectrum by

$$\text{Type I closed string spectrum} = {}^{\text{Type IIB spectrum}}\!/_{\Omega}. \tag{9.9.11}$$

The theory based on this spectrum is not consistent by itself, but if we add states from the original open string spectrum we can create a consistent theory. This theory of open and closed string has $\mathrm{SO}(32)$ gauge symmetry and $\mathcal{N} = 1$ supersymmetry, and is called type I string theory [211].

Heterotic string theory

We will not present the details of the quantization of the heterotic string theories (see [211] or [208] for textbook treatments). The procedure result in a theory of closed string with $\mathrm{SO}(32)$ or $E_8 \times E_8$ gauge symmetry and $\mathcal{N} = 1$ supersymmetry. The theories are called $\mathrm{SO}(32)$ or $E_8 \times E_8$ heterotic string theory.

9.10 String interactions

In the following sections we will mostly refer to bosonic string theory to avoid the extra complications of superstring theory. For the purposes of doing actual physics we are of course mostly interested in the superstring theories, but for the purposes of explaining certain conceptual and technical aspects in the simplest available setting, we prefer bosonic string theory. To some extent we try to comment on any significant departures from the bosonic model that are required to deal with the superstring.

So far we have treated strings without interactions. The free theory is derived from a quantization of the Polyakov action combined with a proper handling of the constraints that occur. The preferred way of introducing interactions in string theory is by using the Feynman path integral idea together with the Polyakov action. The path integral integrates over all possible intermediate propagation paths, and we will discover that this can be seen as a topological expansion. This topological expansion is then taken as the definition of the string theory S-matrix [250].

We will focus on interactions of closed bosonic strings. The basic idea is illustrated in figures 9.1 and 9.2. One or several closed strings are initially vibrating in a specific mode, they then propagates through spacetime and ends up vibrating some other specific modes. The probability amplitude is expressed as the path integral over all possible inequivalent embeddings of the world-sheet into spacetime, as well as all possible geometries for such embeddings, where each embedding and geometry is weighted by the exponential of the Polyakov action. For now, we treat the specification of in-going and out-going states as being accommodated by supplying $\tau_{initial}$ and τ_{final} boundary conditions for the path-integral.

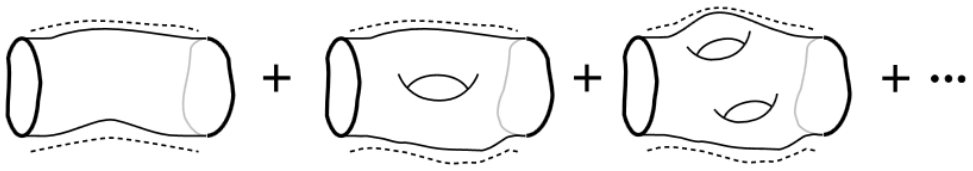


Figure 9.1: The direction of time is from left to right. Amplitude of a closed string state (left-side) turning into another closed string state (right-side) expressed as a topological expansion of path integrals. Illustration from [249]

The action we will employ is a slight generalization of the Polyakov action. We add a topological term in the form of a dilaton field [142][250]. The dilaton can be considered a scalar background field on par with the background Minkowski metric. The dilaton action reads

$$S_\phi = \frac{\lambda}{4\pi} \int d^2\sigma \sqrt{h} R. \tag{9.10.1}$$

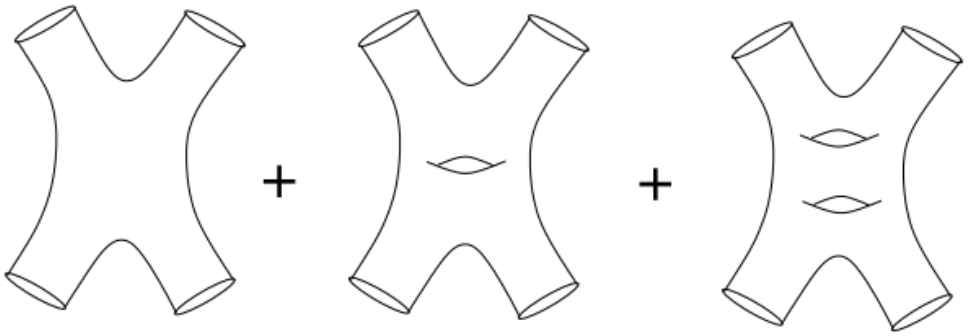


Figure 9.2: The direction of time is from bottom to top. Amplitude of two closed string states (bottom) turning into two other closed string states (top), expressed as a topological expansion of path integrals. Illustration from [250]

Here R is the two-dimensional world-sheet Ricci scalar. We let S_p be the Polyakov action and we let the initial and final states and momenta be labeled by (j^i, p_i) The path integral is then given by

$$\mathcal{A}(j_i, p_i) = C \int \mathcal{D}X\mathcal{D}h \exp(-S_p - S_\phi). \tag{9.10.2}$$

In this expression $\mathcal{D}X$ represents integration of various embeddings of the world-sheet, $\mathcal{D}h$ represents integration over metrics and C is a normalization constant. To make actual sense of this path integral, the in-coming and and out-going states have to be taken to infinity [143]. This means we will be calculating amplitudes for asymptotic states. This is natural in a diffeomorphism invariant theory because local operators are not diffeomorphisms invariant. In a pictorial depiction this corresponds to making the tubes connecting the initial state to the interaction infinitely long (see figure 9.3 a). Because of conformal invariance such asymptotic states can be mapped to small insertion points (infinitesimal holes) on a compact connected surface (see figure 9.3 b) [251][143]. This remapping to insertion points restricts the states to be on-shell states [251]. For each world-sheet history such a remapping leaves a compact closed connected surface. Each of the various histories of the world-sheet corresponds to one such compact surface. Such surfaces have been topologically classified and the topological equivalence class can is specified by a single non-negative integer. The surfaces that represents a world-sheet history will also have a geometry. This division into equivalence classes of topology, and a set of associated geometries on them, means that we have reduced the Feynman path integral to an integer indexed sum over topologies each of which is integrated over with respect to possible configurations for fields and geometries. With this setup the path integral for an n-point amplitude is given by

$$\mathcal{A}^n(j_i, p_i) = \sum_{\text{topologies}} C \int \mathcal{D}X\mathcal{D}h \exp(-S_p + S_\phi). \tag{9.10.3}$$

The specific state (j^i, k_i) found at such an insertion point can always be represented by an operator $\mathcal{V}_{j_i}(k_i)$ because of the state-operator correspondence (we will return to this when discussing CFTs). Such operators corresponding to physical states are called vertex operators. This means we can write the path integral as

$$\mathcal{A}_n(j_i, k_i) = \sum_{\text{topologies}} C \int \mathcal{D}X\mathcal{D}h \exp(-S_p + S_\phi) \prod_{i=1}^{n} \mathcal{V}_{j_i}(k_i). \tag{9.10.4}$$

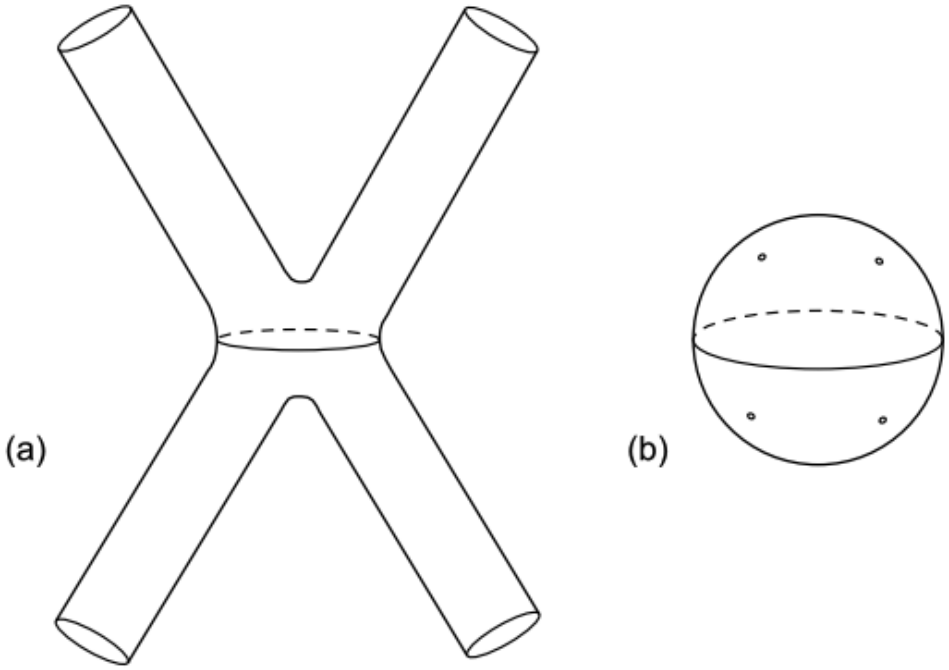


Figure 9.3: Asymptotic states of closed string theory can be pulled into insertion point on a compact surface. Illustration from [143].

The operator $\mathcal{V}_{j_i}(k_i)$ is given as the integral of a local operator by

$$\mathcal{V}_{j_i}(k_i) = \int d^2\sigma\sqrt{h}\mathcal{V}_{j_i}(k_i, \sigma_i). \tag{9.10.5}$$

To deal with the dilaton part of the action we need to make a slight digression. Let Σ be a two-dimensional closed manifold. The integral over Σ of the two-dimensional Ricci scalar gives us the Euler characteristic χ of Σ. We write this as

$$\chi(\Sigma) = \frac{1}{4\pi}\int_\Sigma \sqrt{g}R. \tag{9.10.6}$$

We note for later reference that if Σ is triangulable we can express χ as

$$\chi(\Sigma) = v - e + f, \tag{9.10.7}$$

where v is the number of vertices, e is the number of edges, and f is the number of faces in the triangulation. The Euler characteristic can also be expressed using the concept of genus. Formally the genus is defined in terms of the Euler characteristic, but informally it can be defined as the number of handles of a surface. If we let g be the genus we can express the relation between them as

$$\chi(\Sigma) = 2 - 2g. \tag{9.10.8}$$

We can use integral expression of the Euler characteristic to perform the part of the path integral

involving the dilaton action. We then get that

$$\mathcal{A}_n(j_i, p_i) = \sum_{\text{topologies}} e^{-\lambda\chi} C \int \mathcal{D}X\mathcal{D}h \exp(-S_p) \prod_{i=1}^{n} \mathcal{V}_{j_i}(p_i) \tag{9.10.9}$$

$$= \sum_{\text{topologies}} e^{-2\lambda(1-g)} C \int \mathcal{D}X\mathcal{D}h \exp(-S_p) \prod_{i=1}^{n} \mathcal{V}_{j_i}(p_i) \tag{9.10.10}$$

$$= \sum_{\text{topologies}} \left(g_s^2\right)^{(g-1)} C \int \mathcal{D}X\mathcal{D}h \exp(-S_p) \prod_{i=1}^{n} \mathcal{V}_{j_i}(p_i). \tag{9.10.11}$$

Here g is the genus of the topology in question, and we have defined the string coupling g_S as $g_S = e^{\lambda}$.

The next step is dealing with the proportionality constant C. This constant must accommodate the overcounting implicit in the path integral of a gauge invariant action. The symmetries that we need to deal with are the diffeomorphism invariance and the Weyl invariance. We write this as

$$C = \frac{1}{\text{Vol}_{\text{Diff}\times\text{Weyl}}}. \tag{9.10.12}$$

When we let the insertion points of the vertex operators be parametrize by complex coordinates z we can express the n-point amplitude as

$$\mathcal{A}_n(j_i, p_i) = \sum_{\text{topologies}} \left(g_s^2\right)^{(g-1)} \int \frac{\mathcal{D}X\mathcal{D}h}{\text{Vol}_{\text{Diff}\times\text{Weyl}}} \exp(-S_p) \prod_{i=1}^{n} \int d^2 z_i \sqrt{h}\, \mathcal{V}_{j_i}(p_i, z_i). \tag{9.10.13}$$

The complete analysis of the proper way to deal with the overcounting of the measure $\mathcal{D}X\mathcal{D}h$ is beyond the scope of our presentation (see [211][251] for an analysis). It involves using Riemann surfaces, modular groups, moduli spaces for Riemann surfaces and Teichmuller spaces.

9.11 Conformal field theory

A conformal field theories (CFT) is a quantum field theory that is invariant under conformal transformations (see [242][252][253] for comprehensive textbook treatments). Conformal transformation define a group called the conformal group. In two dimensions the generators of the conformal group form an infinite-dimensional algebra. This algebra is known as the conformal algebra or the Witt algebra [242]. A special case of CFTs are those CFTs that are invariant under super-conformal transformations [252]. Such CFTs as often called super-conformal field theories (SCFTs). We will not always be explicit when referring to SCFTs and just call them CFTs. In the context of the quantum theory of a CFT the relevant algebra is the Virasoro algebra [242]. The Virasoro algebra is a central extension of the Witt algebra. For (quantum) SCFTs the relevant algebra is called the super-Virasoro algebra [252].

Our first observation about CFTs are that the bosonic string in conformal gauge is an example of a CFT. Likewise, the superstring in super-conformal gauge is an example of a SCFT. We can further observe that while CFTs are QFTs, their scale invariance means they must analyzed a bit different from the common QFT approach. Since there is no scale in a CFT, there are no asymptotic states. This means that the traditional S-matrix apparatus cannot be constructed. Instead of calculating S-matrix elements between asymptotic states, one calculates correlation functions. Another feature of CFTs is that they are usually not defined by Lagrangians (although some are). There is also an extended notion of elementary fields in a CFT. Basically any definable operator field is a fundamental field [250]. From the mathematical side a CFT can be seen as a representation

of the Virasoro (or super-Virasoro) algebra. We calim that this is a useful and important perspective on what a CFT is. Unfortunately a full explanation of representational properties of CFTS as well as substantiating this claim falls beyond our scope (see [242] for some details on this).

9.12 Foundations of perturbative string theory

The purpose of this section is to suggest one way to look at string theory that provides some useful theoretical and conceptual insights. The first aspect we emphasize is how string theory represents a generalization from one-dimensional QFT defined on world-lines, to two-dimensional QFT defined on worlds sheets. The second aspect we emphasize is a more abstract view of perturbative string theory then the very visual and geometric view based on strings propagating and interacting in some ten-dimensional background space. The view we present in this section does not represent a new view on string theory. This view is at very least implicit in many presentations of string theory. Some researchers, such as Witten [254] and Schreiber [255] make these views more explicit. We want to put this view center stage, and use it as a foundational tool for understanding both the power and the possible shortcomings of string theory.

In ordinary QFT we can derive Feynman rules from Lagrangian based actions and the path integral [254] [142]. The Feynman rules can be used to calculate the S-matrix elements at any given order. For theories that are either non-divergent or renormalizable, it is possible to make all these S-matrix elements finite and well defined. In principle we could disregard that the Lagrangian and the path integral are the origin of the Feynman rules, and just let the Feynman rules for calculating S-matrix elements define the theory. This would then be a different theory than the original QFT, and in particular it would be a theory that was defined exclusively by a perturbation series. However, these two theories would agree when it comes to perturbative amplitudes.

It turns out that this perturbative structure we just outlined can also be derived by using what is known as the worldline formalism [256][257][258]. In the worldline formalism one defines a (0+1)-dimensional QFT living on graphs. The type of graphs allowed defined the possible interaction types of the theory. Using this setup it is possible to define basic n-point functions (correlators) of this (0+1)-dimensional QFT. (Note that we have reduced QFT to something simpler and substantially more primitive to enable the generalization step of the next paragraph.) The point of this is a follows: The S-matrix elements at a given loop order can be calculated from a sum over such correlators for all distinct graphs with a given loop order. To make the theory fully defined one needs to add decorations at each vertex specifying vertex amplitudes. These do not follow from the QFT defined on the graph. Thus we can define a purely perturbative (3+1)-dimensional QFT by summing correlators of a (0+1)-dimensional theory defined on graphs.

In perturbative string theory something very similar happens. We consider string theory to be a theory of S-matrix elements. These S-matrix elements are defined summing correlators of (1+1)-dimensional CFT or SCFT [255]. The S-matrix elements at each loop level n are defined by correlators of the CFT defined over a surface of genus n. Convincing arguments suggest that this expansion is UV-finite at each loop order [209][211]. The series is not convergent but it is asymptotic. The expansion only defines on-shell perturbative amplitudes for asymptotic states.

There is no known underlying theory from which one can derive derive this perturbative expansion. Possibly there could be such a theory and possibly there could be no such theory. String field theory is one attempt at creating such an underlying theory but it seems to fall short of something approach a full underlying theory.

We will make some further comments on this setup. This setup does not say that string theory is a CFT, but it says that string theory is an S-matrix theory where the S-matrix elements are build from the correlators of a CFT. At the outset the setup might seem a bit ad hoc, but with the parallel presentation of the worldline formalism of QFT, this setup seems to be a somewhat

natural generalization to one higher dimension. To view string theory as a natural generalization of perturbative worldine QFT to (1+1)-dimensions has some explanatory power. It explains why string theory can be order by order finite by eliminating the graph divergences by conformal transformations of corresponding two-dimensional manifold. The operator-state correspondence in CFTs explains why gravity must be a part of string theory. This also explains the why string theories are much more unique than QFTs. There is only one string "diagram" at each loop order and the vertex factors are dictated by the free theory. In contrast QFTs have a factorially increasing set of diagrams and there is an infinite number of different QFTs.

The geometric picture of strings propagating in some given space-time has great intuitive appeal. This view is also limiting. We find it more natural and enabling to assume the CFT view as the fundamental definition of perturbative string theory. We can view the CFTs themselves as representations of two copies of the Virasoro algebra [242]. This means that string theory is a theory defined by S-matrix amplitudes calculated from correlators of representations of $V \oplus V$.

The CFT view further impacts the geometric view of strings. While some CFTs can be derived from theories of strings propagating in an external spacetime, this is not the case for all of them [254]. Depending on the conditions that are made on the fields of the CFT one can create theories that are in our definition consistent string theories, but where strings propagating in space are nowhere to be found. There are CFTs that do not define any ordinary geometric space, and as such do not permit an interpretation of strings linving in space. CFTs themselves can define an extended concept of geometry through interactions with noncommutative geometry [259, 260]. This line of thinking brings us outside of conventional string theory and we will stop here, returning to this subject in our final summary in chapter 10.

9.13 D-Branes

We have previously mentioned Dirichlet boundary conditions for open strings, and suggested that these seemingly unphysical conditions are actually realized in string theory, and that such Dirichlet constrained endpoints define higher-dimensional non-perturbative physical entities called D-branes (see figure 9.4). In this section we try to substantiate this claim further and also look at some consequences of introducing D-branes.

Circle compactification of bosonic closed string

One way to motivate D-branes is by studying how open strings are affected by T-duality (in this section we mostly follow the exposition of [208]). To do this we need to know more about T-duality. To find out more about T-duality we study string theories with one dimension compactified on a circle.

We start by looking at the (25+1)-dimensional closed bosonic string on the space $\mathbb{R}^{24,1} \times S^1$. That is, the bosonic string compactified on a circle of radius R. Let us denote the compactified dimension by the scalar field X^{25}. We first recall some facts of the uncompactified closed string. Being closed means that as we let the parameter σ of the function X^{25} go from σ_0 to $\sigma_0 + l$ the map X^{25} must return to the same point. We write this boundary condition, that literally defines the meaning of being a closed string, as

$$X^{25}(\tau, \sigma + l) = X^{25}(\tau, \sigma). \tag{9.13.1}$$

In other words the map X^{25} is map of S^1 into $\mathbb{R}$.

We now consider the compactification of the target of the map X^{25}. Curling $\mathbb{R}$ into a circle means to quotient the space by the equivalence relation $x \sim x + 2w\pi R$, where $w \in \mathbb{Z}$. Since we have changed the meaning of "being the same point" by taking the quotient, this also changes the meaning of the closed string condition, which says that $X^{25}(\tau, \sigma + l)$ must be same point as

$X^{25}(\tau,\sigma)$. The compactification means that the closed string boundary conditions as applied to X^{25} must be changed to

$$X^{25}(\tau,\sigma+l) = X^{25}(\tau,\sigma) + 2\pi wR. \tag{9.13.2}$$

This leads to states where the string wraps around S^1 one or more times.

When we write down how this modifies the mode expansion of X^{25}, we will be able to deduce what it means for the spectrum. The mode expansion is unchanged except for the part relating to momentum. The modified expansion for $\mu = 25$ is given by

$$X_R^{25}(\sigma^+) = \frac{1}{2}(x^{25} - \tilde{x}^{25}) + \left[\alpha' \frac{K}{R} - WR\right]\sigma^+ + i\sqrt{\frac{\alpha'}{2}}\sum_{n\neq 0}\frac{1}{n}\tilde{\alpha}_n^{25} e^{-\frac{in2\pi}{l}\sigma^+} \tag{9.13.3}$$

$$X_L^{25}(\sigma^-) = \frac{1}{2}(x^{25} + \tilde{x}^{25}) + \left[\alpha' \frac{K}{R} + WR\right]\sigma^- + i\sqrt{\frac{\alpha'}{2}}\sum_{n\neq 0}\frac{1}{n}\alpha_n^{25} e^{-\frac{in2\pi}{l}\sigma^-}. \tag{9.13.4}$$

We have added $\tilde{x}^\mu$ as an extra constant that will prove to be helpful in analyzing T-duality. The mass relation formula is modified from

$$\alpha' M^2 = 2N + 2\tilde{N} - 4 \tag{9.13.5}$$

to

$$\alpha' M^2 = \alpha'\left[\left(\frac{K}{R}\right)^2 + \left(\frac{WR}{\alpha'}\right)^2\right] + 2N + 2\tilde{N} - 4. \tag{9.13.6}$$

In addition the condition $N - \tilde{N} = 0$ is now replaced by $N - \tilde{N} = WK$.

It is clear that the replacement $R \to \alpha'/R$ coupled with $W \to K$ result in the mode expansion changing by

$$X_R^{25} \to -X_R^{25}, \quad X_L^{25} \to X_L^{25}. \tag{9.13.7}$$

This leads to the expansion

$$\tilde{X}^{25}(\tau,\sigma) = \tilde{x}^{25} + 2\alpha' \frac{K}{R}\sigma + 2WR\tau + i\sqrt{\frac{\alpha'}{2}}\sum_{n\neq 0}\frac{1}{n}\left(\tilde{\alpha}_n^{25} e^{-\frac{in2\pi}{l}\sigma^+} + \alpha_n^{25} e^{-\frac{in2\pi}{l}\sigma^-}\right). \tag{9.13.8}$$

This change in the expansion does not change the theory. The map itself is called the T-duality map and the symmetry of closed bosonic string theory is described by saying that the theory is self-dual under T-duality.

Circle compactification of bosonic open string

We now investigate what happens to bosonic open string compactified on a circle under a T-duality transformation. We first notice that the closed string T-duality transformation can be analyzed by just using

$$X_R^{25} \to -X_R^{25}, \quad X_L^{25} \to X_L^{25}, \tag{9.13.9}$$

which transforms the expression

$$X^{25} = X_L^{25} + X_R^{25} \tag{9.13.10}$$

to

$$X^{25} = X_L^{25} - X_R^{25}. \tag{9.13.11}$$

We add the same constant $\tilde{x}^\mu$ for open string that we did for closed strings. Applying the T-duality transformation to the open string we find that the expansion changes to become

$$\tilde{X}^{25} = X_L^{25} - X_R^{25} = \tilde{x}^{25} + p\sigma + i\sqrt{2\alpha'}\sum_{n\neq 0}\frac{1}{n}\alpha_n^{25}e^{-\frac{in\pi}{l}\sigma^+}\sin\left(\frac{n\pi}{l}\sigma\right). \tag{9.13.12}$$

In this expansion there is no term $p\tau$ which means there is no momentum in the X^{25} direction. In addition we see from the $\sin$ function in the expansion that the oscillatory part vanishes at the endpoints (or in other words this coordinate of the string is stationary at the endpoints). The interpretation of this is as follows. A T-duality transformation in open string theory exchanges von Neumann conditions for Dirichlet conditions [261]. This means that if require T-dualtity to be a symmetry of the theory Dirichlet conditions must be included in the theory [261]. The interpretation of the Dirichlet condition is that it corresponds to a physical object called a D-brane. An important fact about D-branes is that they can be identified with soliton solutions of the corresponding low-energy supergravity theory. We will return to this aspect in the section on AdS/CFT.

In light of the reasoning we have just presented, which indicates the necessity of allowing Dirichlet boundary conditions, we choose to see all open string as having Dirichlet boundary conditions. This just means that we imagine an open string of the bosonic string theory with von Neumann condition to have endpoints restricted to end on a space-filing D25-branes (which essentially means they are unrestricted).

Open strings and Chan-Paton factors

Open strings can be endowed with an extra discrete degree of freedom (see [208][142]). This consists of assigning each end of the (oriented) open string with a number from 1 to N. In bosonic open string the number N is arbitrary. Each choice defines a different theory. The endpoints labeled by these integers can be considered to charges. Indeed in the early hadronic days of string theory they were imagined to be quarks [142]. Putting Chan-Paton charges at the end of open strings result in a theory with $\mathrm{U}(N)$ gauge symmetry.

Open strings ending on D-branes define $\mathrm{U}(1)$ gauge theories on the brane worldvolume [208]. If there are N D-branes this symmetry becomes $\mathrm{U}(1)^N$. If the N D-branes are coinciding we get $\mathrm{U}(N)$ Yang-Mills theory living on the brane.

9.14 String dualities

In the period up to 1995 the various string theories were considered to be distinct separate theories. The only known relations between the theories were the relations based on T-duality (see the previous section for an introduction to T-duality). Outside of the T-duality relations the five theories were still islands to themselves. In 1995 Edward Witten suggested that all the know string theories were connected by dualities (see figure 9.5)and that the known string theories should be seen as different perturbative instances of an unknown common theory [239] (see figure 9.6). A central link between the five known perturbative theories was a sixth 11-dimensional theory, called M-theory, which is non-perturbative in the high-energy domain. The low-energy limit of this theory is 11-dimensional supergravity. The 11-dimensional connection was also suggested by Townsend [262].

We saw in the section on D-branes that T-duality relates two compactified string theories compactified on S^1, by exchanging their compactification radius R by α'/R (see the previous

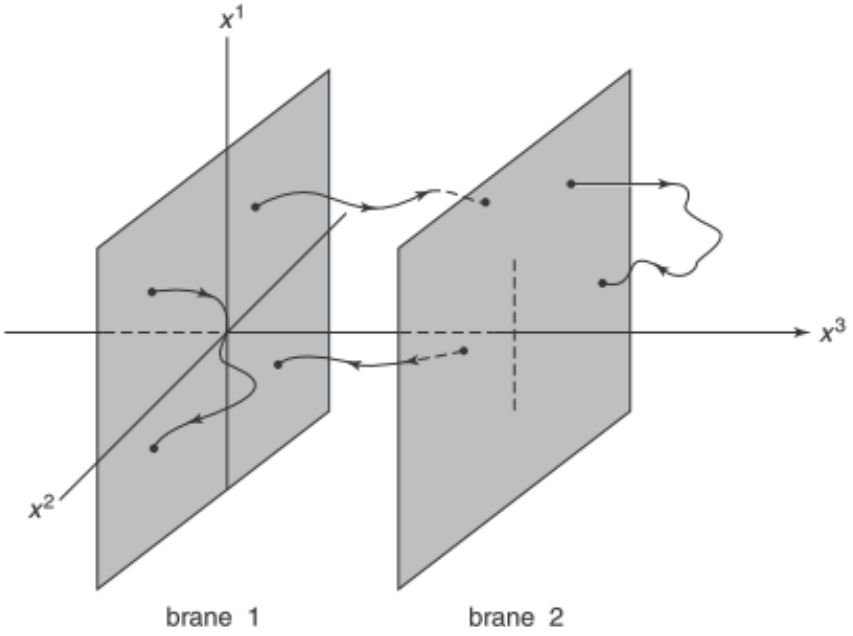


Figure 9.4: Open strings ending on D-branes. From [210]

section for the complete definition). Using this duality one can relate heterotic $E_8 \times E_8$ to heterotic $SO(32)$ when both are expressed on $M_9 \times S^1$ [263]. The same type of T-duality relation exists between type IIA and type IIB on $M_9 \times S^1$ [264]. T-duality arises partly from the string concept of winding numbers. A closed string can wind around a circle defining an integer winding number. Under T-duality the winding number is exchanged with the (quantized) Kaluza-Klein momentum.

The second duality we consider is S-duality. S-duality is a duality between a theory A at strong coupling g_S with another theory B at weak coupling $\frac{1}{g_S}$. S-duality represents a generalization of electric-magnetic duality as found in electromagnetism. A strong-weak type of duality is of course difficult to test in theories that are only defined perturbatively. At strong coupling the perturbative expansion is no longer well defined and the theory can no longer be computed. What makes it possible to still conjecture these dualities are special states found in supersymmetric theories, called BPS states, which have certain properties that are unmodified by perturbations, and that therefore can serve as probes or sign-posts in the non-perturbative domain. We will not explain how this occurs in any more detail. S-duality relates heterotic $SO(32)$ to type I [239], as well as relating type IIB to itself [265].

The final duality we will consider is called U-duality. Let us first note that the low-energy limit of type IIA at perturbative coupling can be expressed as 11D supergravity compactified on a circle. If we denote the radius of compactified dimension by R_{11} this can be translated to say that the $g_S \to 0$ limit of type IIA gives $R_{11} \to 0$. Conversely, we find that in the strong coupling limit of IIA we get $R_{11} \to \infty$. This is interpreted as showing that in the strong coupling limit of IIA a new dimension eleventh dimension appears [239]. This strong coupling limit is represented by a theory called M-theory, whose low-energy limit is 11-dimensional supergravity. A similar duality exist between heterotic $E_8 \times E_8$ and M-theory [266] [267]. In this case the compactified dimension is just a one-dimensional line element. M-theory in the sense we have defined it here is not necessarily any more fundamental than the previously known well defined ten-dimensional superstring theories. M-theory is only equivalent to type IIB, type I and heterotic $\mathrm{SO}(32)$ when these theories are compactified. We therefore choose to see all the most well-known vacua as

Property	Electromagnetism	General
Object	D0-brane	Dp-brane
Spatial dim.	0	p
World vol. dim.	1	p+1
Form field	$A_1 = A_\mu dx^\mu$	$A_{p+1} = \frac{1}{(p+1)!} A_{\mu_1 \dots \mu_{p+1}} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_{p+1}}$
Field strength	$F_2 = dA_1$	$F_{p+2} = dA_{p+1}$
Magnetic field strength	$\tilde{F}_{D-2} = *F_2$	$\tilde{F}_{D-(p+2)} = *F_{p+2}$
Electric charge	$Q = \int_{S^2} *F_2$	$Q = \int_{S^{(p+2)}} *F_{p+2}$
Magnetic charge	$Q = \int_{S^2} F_2$	$Q = \int_{S^{(D-(p+2)=}} \tilde{F}_{D-(p+2)}$
Action	$Q = \int A_1$	$Q = \int A_{p+1}$

Table 9.8: The table show the generalization of various construction in electromagnetism for point-particle to higher dimensional objects and fields. These constructions form the basis of assigning charges and n-form fields to higher-dimensional objects in string theory.

various tips of the moduli space as shown in figure 9.6. Combining the dualities we have mentioned here with compactification leads to many more relations between the various string theories on various compactified spaces (see e.g. [268, p15]).

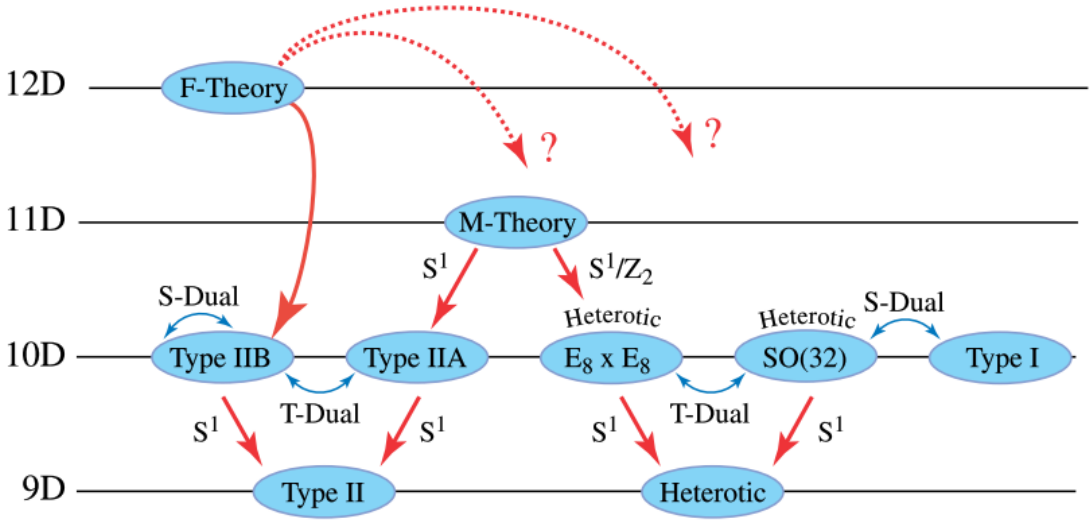


Figure 9.5: The canonical string dualities and relations to F-theory and M-theory. Illustration from [269].

9.15 Lower dimensional string theory

Fermionic string theory is most easily defined in ten dimensions. Our ordinary day-to-day experiences, as well as the current foundations of established physics, clearly stipulate that spacetime has four dimensions. Nobody is disputing that these are the facts about our experiences and our established theories. Does this mean that we know for a fact that spacetime is four-dimensional?

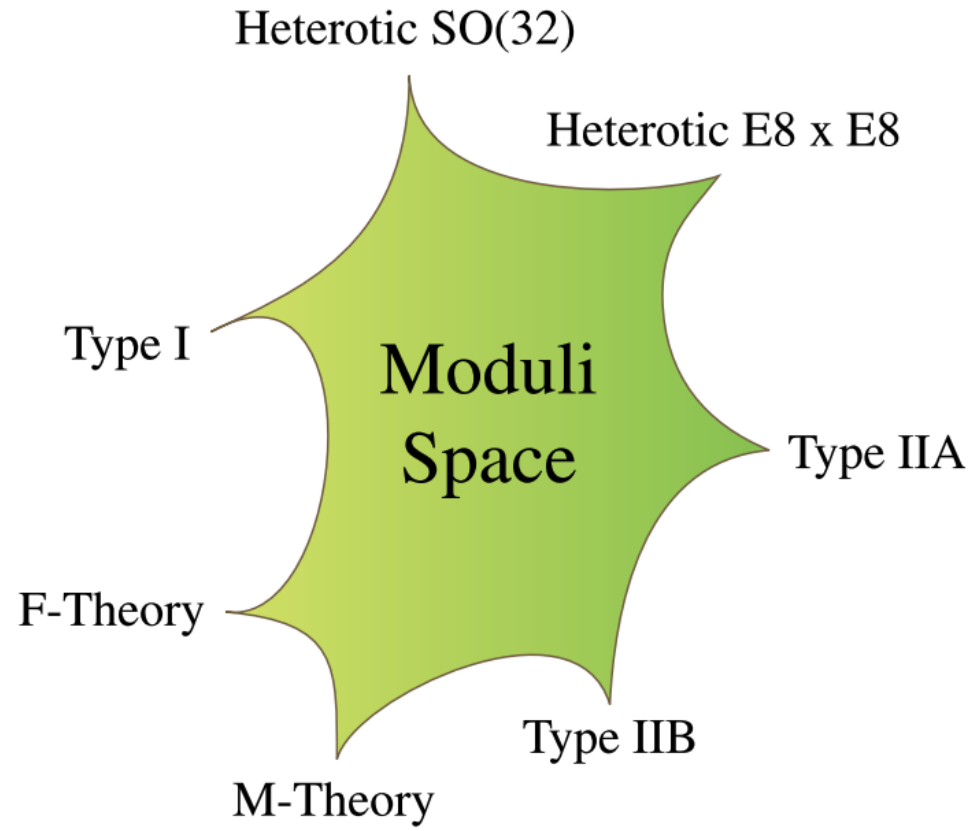


Figure 9.6: The canonical string dualities. Illustration from[269].

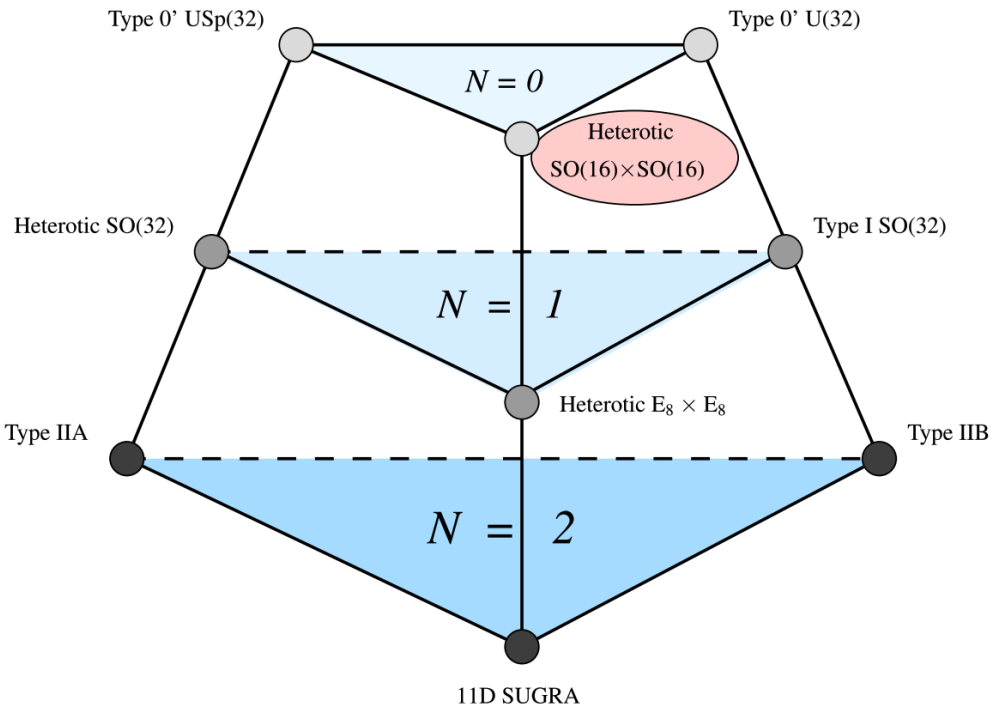


Figure 9.7: An expanded view of string dualities including non-supersymmetric string theories. From [270].

The short answer is that is does not. It is possible to construct models where spacetime has more than four dimensions but still appears four-dimensional in everyday experiences and currently available physics experiments. We therefore need a more refined statement. We could put this as there are only four observable macroscopic spacetime dimensions, which means any additional dimensions must somehow must be rendered unobservable at energies so far tested.

In the context of making a ten-dimensional theory appear four-dimensional, there are basically three ideas on how to achieve this (see [271] for the first two, and [211][272] for the third). The first idea is to express spacetime as a product $M = M_{3,1} \times M_6$ where $M_{3,1}$ is the ordinary four-dimensional spacetime, and M_6 is a six-dimensional compact spatial manifold. This is essentially the same old Kaluza-Klein idea, and this was first considered in string theory in [273] and [274]. Because of quantum theory, the spatial resolution of any experiment is inversely proportional to the energy, and we can therefore hide the dimension of the compact space by postulating that they are too small to be detected at currently available energies.

The second idea is the idea of sequestering particles to live on a lower-dimensional subspace [275]. Spacetime itself is ten-dimensional but the "content" in the form of regular matter, that we as beings made of quarks and leptons can interact with, is actually confined to live in some lower-dimensional subspace. A concrete scenario is that the space we live in is actually a brane with three macroscopic spatial dimensions, or that it is a three-dimensional hypersurface defined by the intersection of several branes (see [271][210]). In such a scenario we need not assume that all aspects of our world is confined to such a surface, and indeed in the most naturally occurring examples this is not the case. Gravity for instance is usually free to propagate in all spatial dimensional and is thus not confined to the brane. When using intersecting branes only the fermionic matter part of the theory is limited to the space defined by the intersection, and the gauge fields can propagate on the full extent of the brane they belong to. This means that we have to use the brane idea in combination with our first idea of an effectively invisible compact space.

The third idea is of a somewhat different character. In this case, some or all of the extra dimensions form an abstract internal space that cannot necessarily be interpreted as geometric space in the sense of spacetime. Models of this kind typically use the CFT formulation of string theory and put specific conditions on its field such that the resulting theory is effectively four-dimensional [211]. It can bee seen as an extension of the construction of the heterotic string theories [211]. Examples of such models are the free fermionic models [272]. It is of course a matter of opinion whether this third idea should be considered to be a compactification at all since in this case there are no proper dimensions to get rid of. We include it amongst compactification ideas since we find it to be quite closely related to "proper" compactifications.

In the following sections we will present a few aspects of the first and second ideas, i.e. the compactification and the brane idea, that we introduced above. We will start with the compactification idea. This is natural since most implementation of the brane idea will also require use of the compactification idea.

9.16 String compactifications

General aspects

In this section we will discuss how to achieve four-dimensional string theory by expressing spacetime as a product $M = M_{3,1} \times M_6$ where $M_{3,1}$ is ordinary four dimensional spacetime and M_6 is a six dimensional compact spatial manifold. The early history of string compactifications begun in 1975 with the first use of the Kaluza-Klein idea in string theory [273][274]. A major milestone was the introduction of Calabi-Yau manifolds [237] and orbifolds [276][277] in 1985. Figure 9.8 presents an overview of the general idea for making contact with ordinary particle

physics using compactification. We will denote M_6 as K. The manifold K can take many different forms. We will limit ourselves to five different cases.

The simplest case we present is to assume that K is a product of circles. This means that K is a six torus T^6. Such compactifications are called toroidal. The next simplest examples will be so called toroidal orbifolds, which are tori quotiented by the action of some finite group with fixed points. Because of the fixed points such quotient are generalization of the manifold concept to include a finite number of singular points.

A different way of extending the toroidal compactifications is by using orientifolds (introduced by Sagnotti in [278]). Orientifolds are parity related projections defined by some hyperplane. Such compactification spaces will be called toroidal orientifolds.

The fourth case sets K equal to a Calab-Yau manifold. The first three examples all have the benefit of being relatively easy to deal with in explicit form. Calabi-Yau manifolds are very different in this respect. It is difficult to give explicit examples of Calabi-Yau manifolds which limits the available techniques. Calabi-Yau manifolds are mostly dealt with by using their topological aspects, usually by techniques from algebraic topology (see e.g. [209]).

The fifth case extends the Calabi-Yau compactifications by also including orientifold planes (see [271] or [211] for textbook treatments).

Definition 9.16.1 — Calabi-Yau manifold. A Calabi-Yau 3-fold is a three-dimensional compact complex manifold that satisfies the following conditions:

1. It is Kahler.
2. It has vanishing first Chern class.
3. It is Ricci flat.

Definition 9.16.2 — Orbifold. An orbifold is a generalization of a manifold allowing singular points. Instead of being modeled on subsets of $\mathbb{R}^n$ is is modeled on subsets of $\mathbb{R}^n$ modulo the action of a finite group. When such an action has a fixed point it defines an orbifold.

Definition 9.16.3 — Orientifold. An orientifold is a generalized orbifold that includes orientation reversal in the orbifold group.

Before we list a few more detail about each compactification setup we need to consider some general aspects. First let us note that there is a long tradition in physics for trying out some given theory on various geometric background. The description of compactification we have give so far falls into that tradition. But in the case of string theory this not quite sufficient. Even though string theory is always defined on a specific background, this is not how string theory should be envisioned. String theory in the most ambitious sense should dynamically determine the background space it propagates in. As we have seen, string theory in its most general formulation does not assume that there is any such geometrically well-defined background at all. But let us assume for the moment that we are in regime where string theory provides a well-defined geometric background, then this background should not be a random choice by the modeler, but something that string theory imposes. Unfortunately string theory is unable to specify such dynamics at present, the best we can do is apply our previous analysis of how string theory consistency puts restrictions on the geometric background. All the compact space we will consider are Ricci flat in fulfillment of the requirements that we will explore in section 9.20. For orbifolds and orientifolds which are not manifolds the criteria holds for the manifold on which the finite group acts.

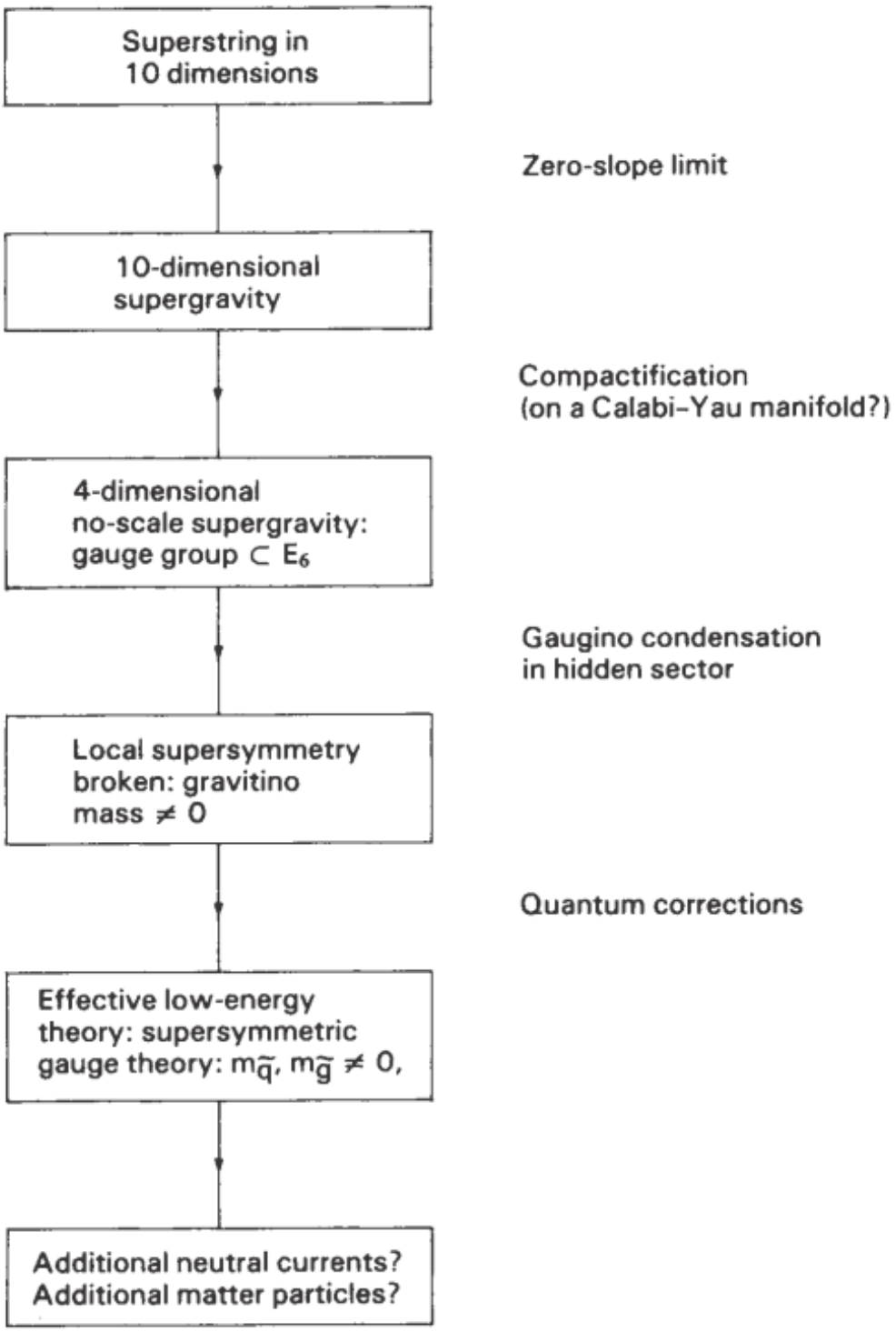


Figure 9.8: A schematic description of achieving realistic four-dimensional physics from superstrings by Calabi-Yau compactification of heterotic $E_8 \times E_8$ string theory. Illustration from [279].

Toroidal compactifications

Using a torus for the compact space K is the simplest possible solution (see [271] or [211] for textbook treatments). Compactifying on a torus preserves all the supersymmetry of the original

theory. This means that compactifying $D = 10, \mathcal{N} = 1$ heterotic theories leads to $D = 4, \mathcal{N} = 4$ and compactifying $D = 10, \mathcal{N} = 2$ type II theories leads to $D = 4, \mathcal{N} = 8$. All theories with extended supersymmetry are non-chiral and therefore not viable as realistic models of particle physics. This aspect can be corrected by orbifolds.

Orbifolds

Orbifolds are generalizations of manifolds that allow isolated singularities [276][277] (see [271] or [211] for textbook treatments). A simple example of an orbifold is R with the action of $\mathbb{Z}_2 = e, g$ defined as

$$g.x = -x, \tag{9.16.1}$$

and the group multiplication is given by $ee = e, eg = ge = g, gg = e$.

Let us now see the same concept for a torus. Let e_1 and e_2 be to independent vectors in $\mathbb{R}^2$. Let these vectors define a lattice Λ by $n_1 e_1 + n_2 e_2$. We define the torus as quotient of $\mathbb{R}^2/\Lambda$. We can now use the same $\mathbb{Z}_2$ group as before, and we define an analogous action on $\mathbb{R}^{\not\vDash}$ by

$$g.\vec{x} = -\vec{x}. \tag{9.16.2}$$

There are four fixed points under this action: $(0,0), (0,1/2)(1/2,0), (1/2,1/2)$. These points are all given relative to the lattice basis vectors. The resulting space can be visualized as in figure 9.9.

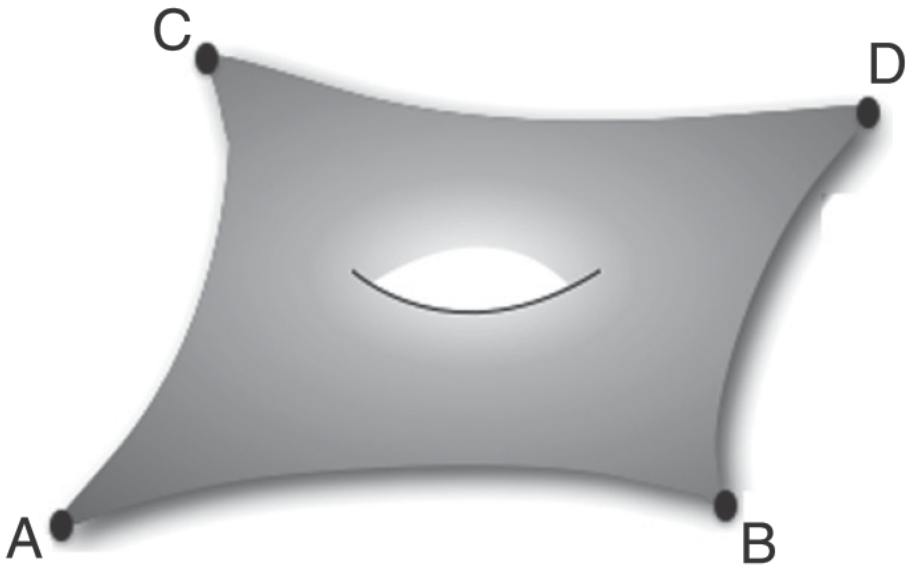


Figure 9.9: An artistic rendering of the orbifold $T^2/\mathbb{Z}_2$ defined in the text. The point A,B,C and D are meant to illustrate singular point of the manifold/orbifold. Illustration from [271]

The two-torus orbifold can be generalized to T^6. To achieve stable models we want the orbifold to lead compactified theory with $\mathcal{N} = 1$. This means that the orbifold must have $\mathrm{SU}(3)$ holonomy. This can be achieved by imposing restrictions on the finite group and on its action on T^6.

Calabi-Yau compactifications

In the early days of string phenomenology the favored approach towards achieving realistic models was compactification of the $E_8 \oplus E_8$ heterotic string on a Calabi-Yau manifold [237]. Calabi-Yau

manifold are chosen as the compact space because they break $3/4$ of the supersymmetry. For $D = 10, \mathcal{N} = 1$ heterotic theories this leads to (in the typical setup) $N = 1$ supergravity coupled to E_6 chiral and gauge multiplets at low energies. For $D = 10, \mathcal{N} = 2$ type II theories this leads to $D = 4, \mathcal{N} = 2$.

The number of fermion generations is determined by the Euler characteristic (half the absolute value) [209]. The multiplicies of the various fields are either fixed or determined by the topological parameters $h^{1,1}$, $h^{2,1}$ and h^1 (see appendix I for definitions). We list the full massless spectrum of heterotic and type II theories compactified on Calabi-Yau manifolds in figures 9.10 and 9.11. We include this just to give a brief idea of what the resulting spectrum looks like and how the topological parameters $h^{1,1}$, $h^{2,1}$ and h^1 influence the spectrum. For a full analysis of the spectrum see [211][280][271].

Multiplets	Component fields	Multiplicity
Gravity	$g_{\mu\nu}, \bar{\psi}_{\mu\dot{\alpha}}\eta, \psi_{\mu\alpha}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}$	1
Chiral	$\Phi, B_{\mu\nu}, \lambda_\alpha\eta, \bar{\lambda}_{\dot{\alpha}}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}$	1
Chiral	$g_{ij}, g_{\bar{\imath}\bar{\jmath}}, \psi_\alpha\eta_{\bar{\imath},\bar{\jmath}}, \bar{\psi}_{\dot{\alpha}}\eta_{i,\bar{\jmath}\bar{k}}$	$h^{2,1}$
Chiral	$g_{i\bar{\jmath}}, B_{i\bar{\jmath}}, \psi_\alpha\eta_{i,\bar{\jmath}}, \bar{\psi}_{\dot{\alpha}}\eta_{\bar{\imath},\bar{\jmath}\bar{k}}$	$h^{1,1}$
Vector	$A_\mu^{(248)}, \bar{\chi}_{\dot{\alpha}}^{(248)}\eta, \chi_\alpha^{(248)}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}$	1
Vector	$A_\mu^{(1,78)}, \bar{\chi}_{\dot{\alpha}}^{(1,78)}\eta, \chi_\alpha^{(1,78)}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}$	1
Chiral	$(A_i^{(8,1)}, \chi_\alpha^{(8,1)}\eta_{\bar{\imath}\bar{\jmath}}) + (A_{\bar{\imath}}^{(8,1)}, \bar{\chi}_{\dot{\alpha}}^{(8,1)}\eta_{\bar{\imath}})$	$h^1(\text{End}\, T)$
Chiral	$(A_{\bar{\imath}}^{(\bar{3},\overline{27})}, \chi_\alpha^{(\bar{3},\overline{27})}\eta_{\bar{\imath}}) + (A_i^{(3,27)}, \bar{\chi}_{\dot{\alpha}}^{(3,27)}\eta_{\bar{\imath}\bar{\jmath}})$	$h^{1,1}$
Chiral	$(A_{\bar{\imath}}^{(3,27)}, \chi_\alpha^{(3,27)}\eta_{\bar{\imath}}) + (A_i^{(\bar{3},\overline{27})}, \bar{\chi}_{\dot{\alpha}}^{(\bar{3},\overline{27})}\eta_{\bar{\imath}\bar{\jmath}})$	$h^{2,1}$

Figure 9.10: Masless spectrum of heterotic strings on Calabi-Yau. Illustration from [211]. Note the types of multiplets and how some of the multiplicities are related to topological parameters. Since we will not use the details of the component fields of this table we refer to [211] for an explanation of the notation.

The gauge symmetry of heterotic Calabi-Yau compactifications is determined by the vector bundle defined on K. The default vector bundle is given by the holonomy group $\mathrm{SU}(3)$ [271]. It leads to the symmetry breaking pattern $E_8 \times E_8 \to E_8 \times E_6$. The extra E_8 is considered to be a hidden sector. It is also possible to define other vector bundles on K that will provide other breaking patterns. If we use $\mathrm{SU}(4)$ we will get $E_8 \times \mathrm{SO}(10)$, and with $\mathrm{SU}(5)$ we will get $E_8 \times \mathrm{SU}(5)$ [271].

The setup of a supergravity coupled to a grand unified theory, with an extra hidden sector, has been a popular setup for many years. With dynamical supersymmetry breaking occurring in the hidden sector by some unknown mechanism which is communicated to the visible sectors by gravity or other very weakly coupled mechanism [281]. This is a natural setup to achieve the minimal supersymmetric standard model (MSSM) with soft terms arising from the dynamical supersymmetry breaking [282]. Potentially the details of the supergravity sector and the supersymmetry breaking can be derived from string theory leading to predictions for soft terms in the MSSM. Mostly this possibility has not been realized, and it has been partially supplanted by simplifying ad hoc assumptions like minimal supergravity (mSUGRA) models. In mSUGRA the parameters are assumed to behave very simply at high-energies from which some concrete prediction for low-energy MSSM phenomenology can be derived [281]. It is probably fair to consider mSUGRA as more of a useful parameter organizing paradigm than an actual proposal for unified physics.

Multiplet	Component fields	Multiplicity
	Type IIA	
Gravity	$g_{\mu\nu}, \psi_{\mu\alpha}\eta, \bar{\tilde{\psi}}_{\mu\dot{\alpha}}\eta, \bar{\psi}_{\mu\dot{\alpha}}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}, \tilde{\psi}_{\mu\alpha}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}, (C_1)_\mu$	1
Hyper	$\lambda_\alpha\eta, \bar{\tilde{\lambda}}_{\dot{\alpha}}\eta, \bar{\lambda}_{\dot{\alpha}}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}, \widetilde{\lambda}_\alpha\eta_{\bar{\imath}\bar{\jmath}\bar{k}}, \Phi, B_{\mu\nu}, (C_3)_{ijk}, (C_3)_{\bar{\imath}\bar{\jmath}\bar{k}}$	1
Hyper	$\psi_\alpha\eta_{i,\bar{\jmath}\bar{k}}, \overline{\psi}_{\dot{\alpha}}\eta_{\bar{\imath},\bar{\jmath}}, \bar{\tilde{\psi}}_{\dot{\alpha}}\eta_{i,\bar{\jmath}\bar{k}}, \tilde{\psi}_\alpha\eta_{\bar{\imath},\bar{\jmath}}, g_{ij}, g_{\bar{\imath}\bar{\jmath}}, (C_3)_{i\bar{\jmath}\bar{k}}, (C_3)_{\bar{\imath}jk}$	$h^{2,1}$
Vector	$(C_3)_{\mu i\bar{\jmath}}, \bar{\psi}_{\dot{\alpha}}\eta_{i,\bar{\jmath}}, \psi_\alpha\eta_{\bar{\imath},\bar{\jmath}\bar{k}}, \tilde{\psi}_\alpha\eta_{i,\bar{\jmath}}, \bar{\tilde{\psi}}_{\dot{\alpha}}\eta_{\bar{\imath},\bar{\jmath}\bar{k}}, g_{i\bar{\jmath}}, B_{i\bar{\jmath}}$	$h^{1,1}$
	Type IIB	
Gravity	$g_{\mu\nu}, \bar{\psi}_{\mu\dot{\alpha}}\eta, \psi_{\mu\alpha}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}, \bar{\tilde{\psi}}_{\mu\dot{\alpha}}\eta, \tilde{\psi}_{\mu\alpha}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}, (C_4^+)_{\mu ijk}$	1
Hyper	$\lambda_\alpha\eta, \bar{\lambda}_{\dot{\alpha}}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}, \tilde{\lambda}_\alpha\eta, \bar{\tilde{\lambda}}\eta_{\bar{\imath}\bar{\jmath}\bar{k}}, \Phi, a, B_{\mu\nu}, (C_2)_{\mu\nu}$	1
Hyper	$\psi_\alpha\eta_{i,\bar{\jmath}}, \bar{\psi}_{\dot{\alpha}}\eta_{\bar{\imath},\bar{\jmath}\bar{k}}, \tilde{\psi}_\alpha\eta_{i,\bar{\jmath}}, \bar{\tilde{\psi}}_{\dot{\alpha}}\eta_{\bar{\imath},\bar{\jmath}\bar{k}}, g_{i\bar{\jmath}}, B_{i\bar{\jmath}}, (C_2)_{i\bar{\jmath}}, (C_4^+)_{\mu\nu i\bar{\jmath}}$	$h^{1,1}$
Vector	$(C_4^+)_{\mu i\bar{\jmath}\bar{k}}, \bar{\psi}_{\dot{\alpha}}\eta_{i,\bar{\jmath}\bar{k}}, \psi_\alpha\eta_{\bar{\imath},\bar{\jmath}}, \bar{\tilde{\psi}}_{\dot{\alpha}}\eta_{i,\bar{\jmath}\bar{k}}, \tilde{\psi}_\alpha\eta_{\bar{\imath},\bar{\jmath}}, g_{ij}, g_{\bar{\imath}\bar{\jmath}}$	$h^{2,1}$

Figure 9.11: Masless spectrum of type II strings on Calabi-Yau. Illustration from [211]. Note the types of multiplets and how some of the multiplicities are related to topological parameters. Since we will not use the details of the component fields of this table we refer to [211] for an explanation of the notation.

Orientifolds

We will not say much about orientifolds. Orientifolds can be used to break supersymmetry [271]. In Calabi-Yau compactification of type II this provides an opportunity to break $\mathcal{N} = 2$ down to $\mathcal{N} = 1$.

9.17 Brane worlds

The brane worlds scenario is based on sequestering a specific part of physics on a D-brane or an intersection of D-branes. The sequestration reduces the dimensionality available to the sequestered physics but as we have mentioned the non-sequestered parts (at least the part that interacts with the sequestered part) must also appear four-dimensional and therefore we must still compactify along the $M = M_4 \times K$ schema.

In a brane world compactification the gauge theory is created by having several sets of N_i coincident D-branes each providing a $\mathrm{U}(N_i)$ gauge theory. The fermions of brane world models are provided by intersecting branes. At the intersection chiral fermions appear in the spectrum. Brane world compactification have been mostly applied to type II string theories. For IIA D6-branes are used and for IIB D7 with D3-branes or D9 with D5-branes have been used. We will only consider a brief sketch of the IIA model based on D6-branes.

The IIA model starts out with a model using one stack of three D6-branes (N_1 in figure 9.12) plus one more stack of two D6-branes (N_2 in figure 9.12). These intersecting stacks provides a gauge symmetry on the brane of $\mathrm{U}(3) \times \mathrm{U}(2)$. This theory can be seen as a gauge theory with $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1) \times \mathrm{U}(1)$. This simple model does not have the correct fermion spectrum. One way to get the fermion spectrum of the standard model is to add four more D-branes (N_3, N_4, N_5, N_6 in figure 9.13). The extra branes are needed to produce fermions that transform differently under the various gauge groups of the standard model as indicated by the labels on the

open strings in figure 9.13 (see [210] or [271] for the details of such models). Symmetry breaking along the lines of the Higgs model have been difficult to implement in such models [210][271].

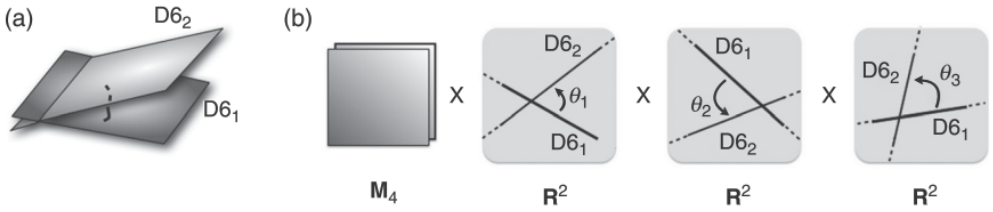


Figure 9.12: A general setup for intersecting D6-branes. Illustration from [271].

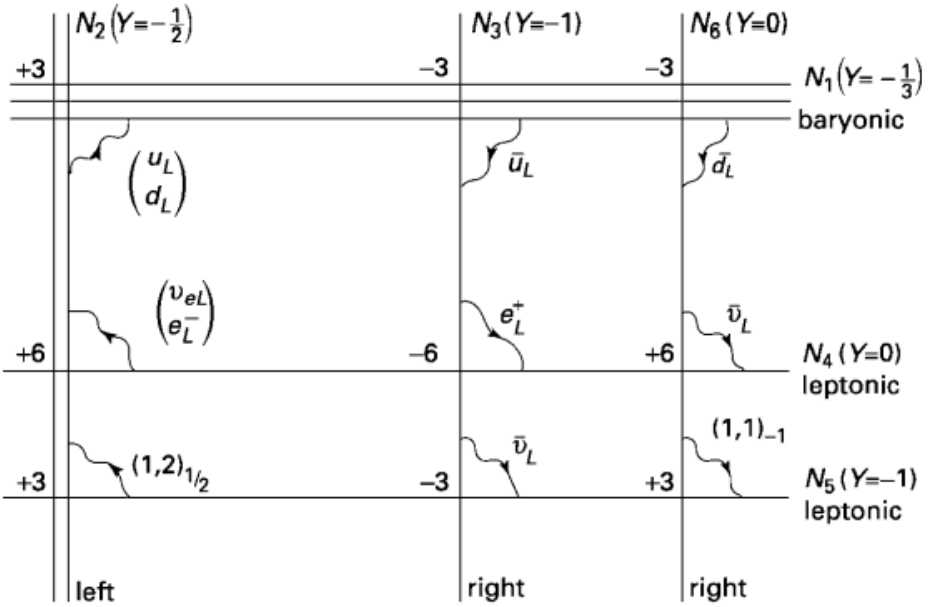


Figure 9.13: A combination of D-branes giving a particle spectrum equal to the standard model. The basic model uses the stack N_1 and N_2. To get the proper fermion representations $N_3 - N_6$ is added to the model. Illustration from [210].

9.18 Moduli stabilization

This section deals briefly with the so called moduli stabilization problem [271][283]. It concerns how to specify the complete set of details for the compactification spaces. This is important to achieve realistic low-energy theories and to be able to make specific predictions regarding masses and mixings.

All the compactification models we have considered involve, as a part of the solution, writing the total space as $M = M_4 \times K$. We have used both tori, orbifolds and Calabi-Yau manifolds as our compact manifold K. When we specify K we have let several aspects been unspecified. We have said thing like K is the torus T^6 or things of a similar nature. The point is that there are many different tori which are homeomorphic to T^6. These different tori have different radii of the circles they are the product of. The circle radii is the simplest example of a free parameter of the compactification geometry.

The totality of free parameters for a given class of manifolds forms a moduli space (loosely speaking, the proper definition of a moduli space is not always possible). The individual parameters are called moduli. The moduli space concept involved in compactification is analogous to the moduli spaces that parametrize the geometries for each genus in the topological perturbative expansion that we presented when discussing string interactions. In mathematical terms the moduli space usually has some natural intrinsic structure. E.g. the moduli space of a class of manifolds could itself have a natural manifold structure. In physical terms the moduli corresponds to massless scalar fields that have long range interactions. Unless these interactions have exceptionally weak couplings these massless fields are in violation of experimental upper limits on long range forces beyond gravity and electromagnetism.

It is important to find a mechanism that can give these massless moduli fields a high enough mass. The moduli fields are massless because they are associated with a flat potential, and the basic idea is to introduce mechanisms that can make this potential non-flat. There are two major ideas to make this happen.

For type II theories the idea is to introduce background fields stemming mostly from the p-form fields. In the proper setting these introduce a potential for moduli and "stabilizes" the moduli. This mechanism is called flux-compactification [271][283].

For the heterotic theories the idea is to use a bundle defined on K. Note that there is already a bundle defined on K to function in breaking the gauge group of the heterotic theory down to a product of some smaller sub-groups. In the moduli context this bundle is given a second task in the form of stabilizing moduli moduli fields [284].

9.19 String theory landscape

> ... the large number of possible types of compactification which exist e.g. in the theories of superstrings should be considered [..] a virtue of these theories, since it increases the probability of the existence of mini-universes in which life of our type may appear.
>
> Linde

From the earliest suggestion of semi-realistic four-dimensional models of string theory, it has been clear that going from a potentially unique theory in ten dimensions to a four-dimensional theory, might introduce a large number of choices [285]. Some researchers viewed this multitude as something that one could conquer by finding the one correct vacuum. A quote from a textbook by Kaku (1990)[241] illustrates this viewpoint: "The theory has an embarrassment of riches. There are apparently *thousands* of ways to break down the theory to low energies." (note: the emphasis on thousands is by Kaku). In Kaku's next textbook (1991)[286] the "thousands" has increased to "millions" but the sentiment is basically the same. The central question is to find the proper vacuum and derive the standard model. A few researchers accepted early on that string theory could not predict unique models (see [285] and references therein).

For those seeking a unique vacuum the invention of flux compactifications became the final nail in the coffin for such ambitions. Flux compactification have made it clear that the number of possible vacua are in the range of 10^{500} and above [287]. This has led to a reevaluation of the paradigm where one envisions deriving standard model predictions from string theory [287][288]. The interpretation of this has not yet fully converged on a single viewpoint, but most string researchers seems to be moving towards an understanding of the physical parameters of our current world as being environmentally or stochastically determined. This means that string theory does not select a particular vacuum but that this is determined by random initial conditions (or something similar). This is often phrased in the context of eternal inflation and multiverse ideas [289][11].

In the eternal inflation scenarios, inflation is a process that once started somewhere in the universe will always continue in other parts of the universe (see figure 9.14). This leads to an infinite number of bubble universes where our observed universe is just one such bubble [290][11]. One then couples this idea to string theory landscape, and imagines that this would lead to different string theory vacua in the different bubbles. This provides for a different set of explanations for physical parameters that appear unnatural or hard to explain. Things like the cosmological constant and the Higgs mass can then be considered more or less arbitrary numbers determined by stochastic processes [291]. One can then suggest that the "special" values these parameters have in the observable universe is explained by providing the necessary conditions for large scale structure formation and complex chemistry (which again is likely to be necessary for life and intelligence) [11].

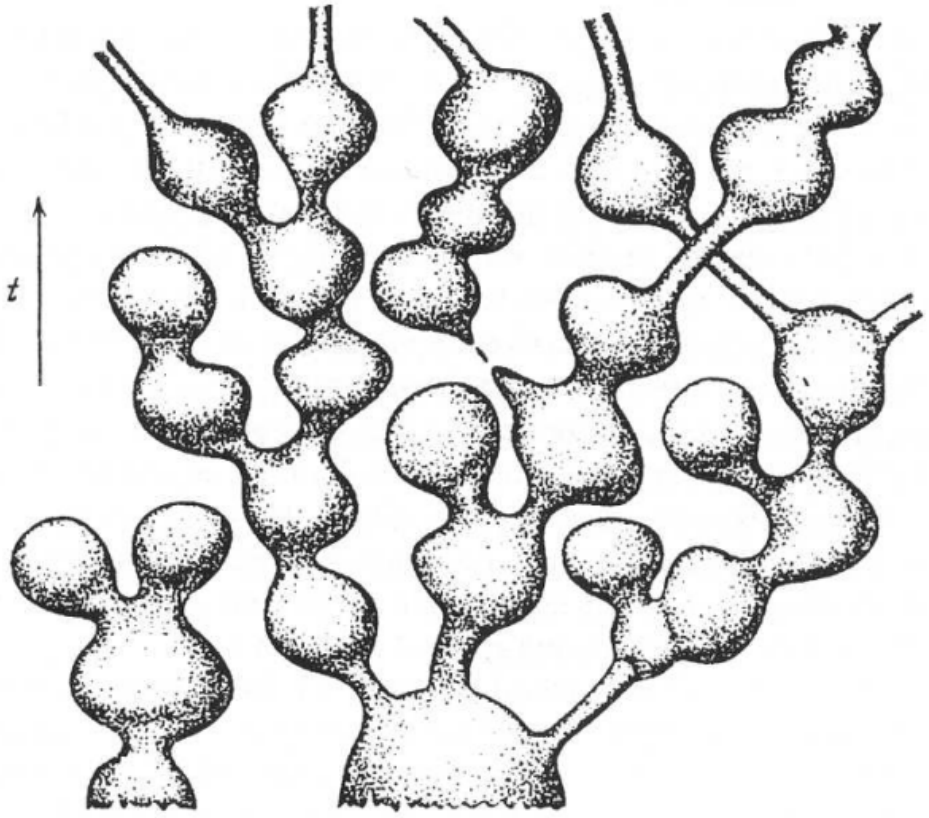


Figure 9.14: The idea of a multiverse produced by eternal inflation. The time direction is from down to up. Illustration from [290].

9.20 Gravity in string theory

String theory is said to contain gravity theory and to define a quantum theory of gravity. In this section we want to explore some important aspects of how string theory relates to gravity. In later sections we will approach this question in the context of black holes, AdS/CFT and entanglement entropy.

The first aspect we want to explore is the low-energy effective field theory limit of string theory. We will start with this more general formulation of the question and then afterwards specialize to the aspects relating to gravity. In the general introductory section, and in the Kaluza-Klein history

section, we have already alluded to the result and discussed a few of its consequences. Even so we have not yet had time to properly define this concept. It is therefore due time that we more carefully define our terms. I.e. what does low-energy effective field theory limit actually mean? We can make some progress towards defining this idea by observing that the energy scale of string theory is determined by the parameter α'. The parameter α' defines the string mass scale as

$$M_s = \alpha'^{-1/2}. \tag{9.20.1}$$

From this equation we can conclude that taking the limit $\alpha' \to 0$ means that excited string states will be infinitely heavy. Recall that in the analysis of the string spectrum, all excited states had values of M^2 proportional to M_s $(1/\alpha')$. (For the present discussion we ignore tachyonic states.) Thus, taking the $\alpha' \to 0$ $(M_s \to \infty)$, should imply that only the massless states of the theory are relevant. This conclusion gives us the (sub)spectrum of particle states that we need to deal with in the low-energy limit.

What else do want to know about the low-energy limit? We certainly want to know how the states are interacting. In string theory, all we know about the interactions is captured in the perturbative expansion for on-shell S-matrix elements. This means that to determine how the massless states interact we must calculate S-matrix elements. S-matrix elements by themselves are useful but not so easy to summarize. What we want to do is to find some (preferably well known) field theory, that has the same spectrum as the massless modes of string theory, and to some approximation produces the same scattering amplitudes. This is what we shall mean by the low-energy effective field theory limit.

The next question is of course: How can we find such an equivalent field theory without having to calculate all possible scattering amplitudes? There is actually a well defined method to achieve this, at least in Minkowski space. The method is described in detail in chapter 16 of [211].

The conclusion from performing such an analysis is that, if we ignore higher-order terms, the low-energy effective field theories that reproduce string scattering amplitudes are either, type IIA and IIB supergravity theories with no additional fields, or type I supergravity coupled to vector super-multiplets [211]. The type I supergravity is the limit of both type I and heterotic string theories. Given what we know about the spectrum of string theories, this result is essential to establish the consistency of the approach. A theory containing massless states of string theories, which includes massless spin-2 fields, and sometimes massless spin-1 fields, must be a theory of supergravity in the spin-2 case, and a Yang-Mills theory in the spin-1 case, as these are the only consistent low-energy theories with that particle content (see chapter 3 or [44]).

The above discussion answers the general question of the low-energy limit. We now want to comment more specifically on how this relates to gravity. In the sense of low-energy effective field theory, string theory include ten-dimensional supergravity. Ten-dimensional GR can be considered a low-energy approximation to supergravity (at least when we include breaking local supersymmetry), and in this sense string theory includes GR as a low-energy limit. The bosonic string has 26-dimensional GR, extended by some additional fields, as its low energy limit [142]. This does not quite complete the analysis. All we have found is the ten-dimensional low-energy limit of superstrings. Further analysis shows that compactifications does not change the gravity part of this conclusion (some of the relevant details can be found in the section on Kaluza-Klein models). See also figures 9.10 and 9.11 that show the gravity multiplet when string theory is compactified on a Calabi-Yau manifold.

We also want to explore the interrelation between gravity and string theory from another viewpoint. GR is a theory of dynamical spacetime, but string theory is defined on a fixed spacetime. We need to try and make sense of this aspect. First of all, although string theory is defined on a fixed background spacetime, this does not mean that it can only be defined on one particular such background spacetime. We have used Minkowski space as the background spacetime for defining

string theory, but this can be generalized this to a curved background spacetime. On a general curved background the Polyakov action is changed to [142]

$$S = -\frac{T}{2}\int d^2\sigma\sqrt{-h}h^{\alpha\beta}\partial_\alpha X^\mu(\tau,\sigma)\partial_\beta X^\nu(\tau,\sigma)g_{\mu\nu}. \tag{9.20.2}$$

In this action the metric $g_{\mu\nu}$ defines some curved background spacetime that the string propagates in. Let us try and investigate whether string theory puts any constraints on this background spacetime.

If we express the generalized Polyakov action in conformal gauge we get

$$S = -\frac{T}{2}\int d^2\sigma\partial_\alpha X^\mu(\tau,\sigma)\partial_\beta X^\nu(\tau,\sigma)g_{\mu\nu}. \tag{9.20.3}$$

We want to determine the conditions on $g_{\mu\nu}$ for this action to be invariant under Weyl transformations. Weyl invariance implies scale invariance which again implies a vanishing β-function. This is the RGE definition of scale invariance. We can therefore investigate the question of Weyl invariance by studying the β-function. The one-loop β-function is given by [142]

$$\beta_{\mu\nu} \propto R_{\mu\nu}. \tag{9.20.4}$$

It is clear that preserving conformal invariance must mean that beta functions vanish. In this case demanding vanishing beta functions has interesting consequences. At the one-loop level it implies that the background spacetime must be Ricci flat. Or in other words, the background spacetime must satisfy Einsteins equation in vacuum ($R_{\mu\nu} = 0$) [142].

As a third point of contact between gravity and string theory, we want to expand on the background consistency requirements that we just presented. We found that demanding Weyl invariance sets requirements on the background metric, but we can integrate this requirement into our understanding of string theory in another way as well. We can rewrite the curved metric $g_{\mu\nu}$ as a perturbation $h_{\mu\nu}$ of the flat Minkowksi metric $\eta_{\mu\nu}$. We write

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}. \tag{9.20.5}$$

We can now put this ansatz into the path integral and see what we can gather. We write this as

$$Z = \int DX^{stione}Dh_{\alpha\beta}e^{-\frac{T}{2}\int d^2\sigma\partial_\alpha X^\mu(\tau,\sigma)\partial_\beta X^\nu(\tau,\sigma)(\eta_{\mu\nu}+h_{\mu\nu})} \tag{9.20.6}$$

$$= \int DX^{stione}Dh_{\alpha\beta}e^{\frac{T}{2}\int d^2\sigma\partial_\alpha X^\mu(\tau,\sigma)\partial_\beta X^\nu(\tau,\sigma)(\eta_{\mu\nu})} \tag{9.20.7}$$

$$\times\left\{1+\frac{T}{2}\int d^2\sigma\partial_\alpha X^\mu(\tau,\sigma)\partial_\beta X^\nu(\tau,\sigma)h_{\mu\nu}\right. \tag{9.20.8}$$

$$\left.+\frac{1}{2}\left[\frac{T}{2}\int d^2\sigma\partial_\alpha X^\mu(\tau,\sigma)\partial_\beta X^\nu(\tau,\sigma)h_{\mu\nu}\right]^2+\cdots\right\} \tag{9.20.9}$$

The extra part that is give by

$$\frac{T}{2}\int d^2\sigma\partial_\alpha X^\mu(\tau,\sigma)\partial_\beta X^\nu(\tau,\sigma)h_{\mu\nu}, \tag{9.20.10}$$

is nothing but the vertex operator corresponding to emission of a graviton. We can therefore sketch a connection where the string theory curved background is actually determined by a coherent set of gravitons [142]. This is an attractive line of reasoning and is certainly a potentially consistent interpretation of the background aspects we have just presented. However, this is a very loose conceptual idea that has not yet been defined operationally. There are no suggestions as to what kind of coherent states might be able to produce such a curved background.

9.21 Black hole thermodynamics

> ... the program for counting the states of black holes amounts to treating the black hole as a bound state of solitons, quantizing the solitons at weak coupling, and extrapolating the resulting collective coordinate degeneracy back to strong coupling via an appeal to supersymmetry.
>
> Balasubramanian

Before we progress in our presentation of string theory we need to digress slightly to present a historical context on black holes. Black holes are specific solutions of Einsteins equations, or in other words they are specific pseudo-Riemannian geometries that satisfy the equations of GR. The first such solution was found by Schwarzschild in 1916 (a translated version is given by [292]). In highly simplified terms one can say that Schwartzchild black holes are spherically symmetric solutions were gravity in the most central region is so strong that the radial direction changes from being spatial to being temporal. The three-dimensional surface defined by the this change in the radial coordinate is called the event horizon. The future of anything inside the event horizon lies inside the horizon, and furthermore the ultimate future of any causal curve inside the horizon is the singularity at "the center" of the black hole. Even with later generalization to include black holes with spin and charge, classical black holes are essentially featureless [293][45]. They can be characterized by just three numbers specifying their mass, spin and electric charge [45]. Classically, nothing escapes from inside the horizon, and specifically the classical black hole does not emit electromagnetic radiation.

Based on studies of classical black holes suggestion were made that some properties of black holes correspond to properties of thermodynamic systems. Hawking proved that the black hole horizon area can never decrease [294]. Soon after the four laws of black hole mechanics were suggested [295]. At the same time Bekenstein developed general ideas of the entropy of black hole horizons (see Bekenstein [296, 297, 298]). When Hawking extended these studies to quantum fields in classical black hole backgrounds, these thermodynamics analogs became even clearer. Various properties, like entropy and temperature, could be identified with specific properties of the black hole solutions (see Hawking [299, 300]). Unruh in related work studied accelerated observers and quantum fields (see Unruh [301]). The idea of thermodynamic analogs in black hole system is of prime interest for a microscopic understanding of space and time, since thermodynamics in ordinary systems is known, by the methods of statistical mechanics, to stem from the underlying microscopic features of the systems in question. This suggest that the thermodynamic properties of black holes could be a first clue to the underlying microscopic states of black holes. The entropy of black holes was found to be unusual in two respects. First, the entropy of a black hole is much higher than ordinary objects of similar mass, and second, the black hole entropy is proportional to the black hole surface area instead of the volume (see [302] for a review).

The thermodynamic picture of black holes was completed by assigning them a temperature. Reasoning more carefully about quantum fields in this classical background leads to the conclusion that black holes radiate electromagnetic radiation in accordance with this temperature [300]. The black hole then necessarily looses mass/energy in the radiation process and will eventually evaporate completely. A crucial feature of this radiation is that it is believed to be perfectly thermal. That is, it is represented by a mixed state that is perfectly thermal. This is in contrast with ordinary thermal radiation which are just pure states that are thermal on average. We naturally assume that any collection of objects in a pure state can be used to enlarge a black hole. When throwing such objects into the black hole, they disappear behind the horizon of the black hole. On the outside of the horizon only perfect thermal radiation is seen, and eventually the black hole disappears leaving nothing else behind. This means that we have converted pure quantum states into mixed quantum

states [303]. This conversion is not compatible with unitary time evolution. Thus, the straight forward application of quantum mechanics to states living on a black hole background leads to a contradiction. This apparent contradiction is called the black hole information paradox. The evaporation of a black hole seems to destroy the information contained in pure states leaving only thermal radiation behind [304].

9.22 Gauge/gravity duality

> ...hidden within every non-Abelian gauge theory, even within the weak and strong nuclear interactions, is a theory of Quantum Gravity.
>
> Horowitz and Polchinski

Introduction and history

In 1997 Maldacena proposed that a five-dimensional[1] string theory was equivalent to a four-dimensional QFT [305]. In this context, to be equivalent means that these two theories define exactly the same physics, including the same states, the same observables and the same dynamics. This means that a (n+1)-dimensional theory of gravity is equivalent to an n-dimensional theory without gravity. Maldacena's conjecture later became known as the AdS/CFT conjecture, or more generally as gauge/gravity duality. Many more suggestions of such dualities between QFTs and gravity theories have been presented beyond the initial AdS_5/CFT_4 gauge/gravity duality.

The gauge/gravity duality conjecture is a groundbreaking idea. It has influenced many other fields and remains an active research topic (see e.g. [306][307][308][302]). Our prime interest here is that it relates a theory with gravity to a theory without gravity. The theory of gravity is defined on a space that asymptotically must be AdS but otherwise it is not restricted to a specific background metric. The gauge/gravity duality therefore gives a partially background independent definition of string theory. In addition, the idea of describing d-dimensional physics by a (d-1)-dimensional theory, is an idea that could contain insights into the very origin of spacetime.

There are several historical roots to Maldancena's conjecture. We have already mentioned the work of Hawking and Bekenstein on semi-classical black holes, and the seeds of holographic ideas contained within those contributions. In 1974 't Hooft suggested that scattering amplitudes of the high N limit of $\mathrm{U}(N)$ gauge theories behave like string scattering amplitudes (or dual model scattering amplitudes) [309]. 't Hooft along with Susskind also presented more general holographic ideas concerning gravity and black holes [310] [311]. Holographic in this context means that the entropy a theory defined on a spatial volume V is proportional to the area of ∂V instead of being proportional to the volume of V. Or in other words that a d-dimensional theory of gravity could be modeled by using a (d-1)-dimensional theory[2].

Already in 1986 a publication appeared that related (2+1)-dimensional gravity (with AdS asymptotics) to (1+1)-dimensional conformal field theory [312]. However these ideas were not picked up and developed further before much later. A much more recent thread (bordering on being an immediate precursor) was the discovery of D-branes, and in particular the realization of a correspondence between open string descriptions and closed string descriptions of D-branes [238]. Basically this says that there exists two corresponding (but not necessarily equivalent) description

[1] Note that the string theory in question is actually defined on the ten-dimensional space $AdS_5 \times S^5$ but that we can often regrad this as string theory compactified on S^5.

[2] Let us consider a theory on a lattice where each lattice site can be in n. Let the lattice be three-dimensional and cubic with lattice separation ϵ and with side lengths $L = N\epsilon$ for some integer N. The volume of the lattice is $V = L^3\epsilon^3$, and the number of states are given by n^{L^3}. The entropy is given by $S = log(n^{L^3}) \propto L^3 = V/\epsilon^3$

of D-branes. One is in the form of open strings ending on D-branes, and the other is in the form of closed strings representing the gravitational effect of D-branes on the surrounding spacetime. We have already seen that open strings ending on D-branes are represented by gauge theories, while closed strings represent a theory of gravity. The construction of a model of black hole entropy in string theory by using D-branes is another closely related thread [313]. Before we get to the actual AdS/CFT conjecture we will briefly review some of these threads.

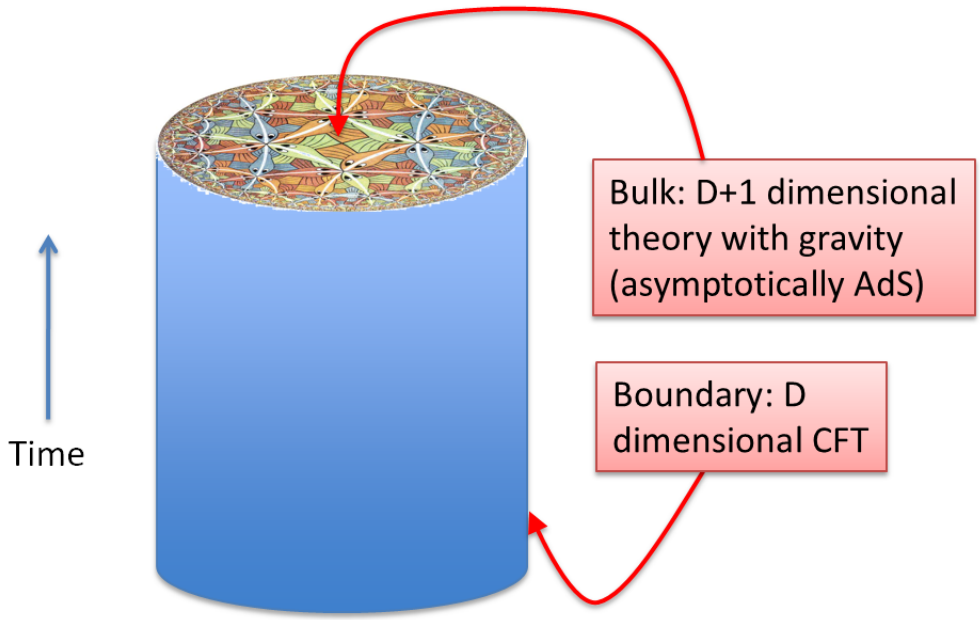


Figure 9.15: Illustration of the $AdS_{(3)}/CFT_2$ correspondence. The Escher drawing represents the metric properties of the AdS space. Illustration from [314].

Large N_c expansion

The theory of quantum chromodynamics (QCD) is a Yang-Mills theory with gauge group $\mathrm{SU}(3)$ (see e.g. [315] or [43] for a basic treatment of QCD). The 3 in $\mathrm{SU}(3)$ represents the number of colors in QCD. We can easily imagine generalizing QCD to a theory with a large set of colors. Such a theory would have a gauge group $\mathrm{SU}(N_c)$, where N_c is the number of colors. As first emphasized by 't Hooft [309], the large N_c limit of such theories is of particular interest for string theory, and we will now see why this is the case.

Let $\Phi = \Phi^a T_a$ be a matrix valued scalar field, with T_a being Lie algebra matrices of $\mathfrak{u}(N_c)$ in the fundamental representation. Let us define a Lagrangian by setting

$$\mathcal{L} = -\frac{1}{2}\,\mathrm{Tr}(\partial_\mu \Phi \partial^\mu \Phi) + g_{YM}\,\mathrm{Tr}\big(\Phi^3\big) + g_{YM}^2\,\mathrm{Tr}\big(\Phi^4\big). \tag{9.22.1}$$

By a simple rescaling $\Phi \to g_{YM}\Phi$, this Lagrangian changes into

$$\mathcal{L} = -\frac{1}{g_{YM}^2}\left\{\frac{1}{2}\,\mathrm{Tr}(\partial_\mu \Phi \partial^\mu \Phi) + \mathrm{Tr}\big(\Phi^3\big) + \mathrm{Tr}\big(\Phi^4\big)\right\}. \tag{9.22.2}$$

Let us define the 't Hooft coupling as $\lambda = N_c g_{YM}^2$. We will be interested in the limit of N_c being large while λ is held constant. We focus on how the various components of the Feynman diagrams scale with N_c and λ. The propagator is given as

$$\langle \Phi_j^i(x) \Phi_l^k(y) \rangle = \delta_l^i \delta_j^k \frac{g_{YM}^2}{4\pi^2 (x-y)^2}, \tag{9.22.3}$$

which means it contributes a factor $g_{YM}^2 = \frac{\lambda}{N_c}$ to the amplitude (the deltas disappear and we are ignoring numerical factors). The three-vertex and four-vertex factors are both given as g_{YM}^{-2} and therefore contribute a factor $\frac{N_c}{\lambda}$. For each closed loop there is a factor of N_c coming from the trace.

We choose to represent Feynman diagrams using the 't Hooft double line notation where the end of each line is labeled by an index from the indices of $\Phi_j^i(x)$(see figure 9.16). Let us now consider just vacuum diagrams of the type shown at the bottom of figure 9.16. By a simple set of rules vacuum diagrams can be interpreted as a triangulation of a closed surface (see figure 9.17). Diagrams that correspond to a triangulation of the sphere are called planar diagrams and those that correspond to a torus of genus $g \geq 1$ are called non-planar.

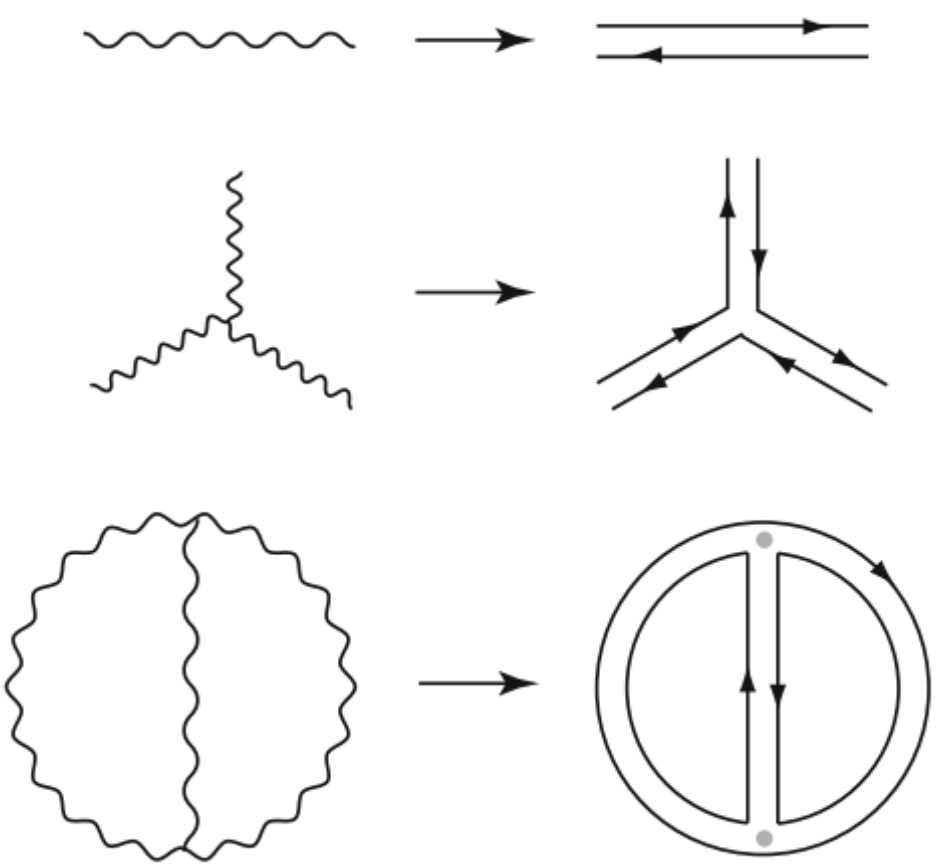


Figure 9.16: Illustration of 't Hooft double line notation for Feynman diagrams. Illustration from [302].

We denote the number of vertices as V, the number of propagators as E, and the number of closed loops as F. For a each vacuum Feynman graph we get

$$\mathcal{A} \propto \left(\frac{\lambda}{N_c}\right)^E \left\{\frac{N_c}{\lambda}\right\}^V N_c^F = \lambda^{E-V} N_c^{V-E+F}. \tag{9.22.4}$$

The point of our terminology V, E, F is that using the double line Feynman diagram notation each such diagram corresponds to a triangulation of a closed surface with V vertices, E edges, and F faces (see figure 9.17). We can now observe that the exponent for N_c is $V - E + F$, which is exactly the definition of the Euler characteristic of a triangulated surface (with V vertices, E edges, and F faces). We have that

$$\xi = V - E + F = 2 - 2g, \tag{9.22.5}$$

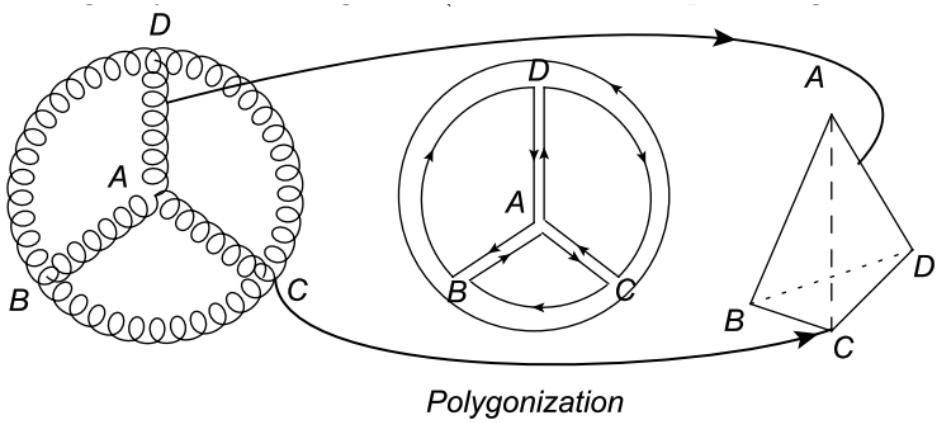


Figure 9.17: Illustration of how 't Hooft double line vacuum diagrams can be interpreted as a triangulation (or polygonization) of a closed surface. The diagram on the left is shown with curly QCD lines which are meant to represent the ordinary propagator lines of the Φ theory we are considering. In the illustrated case we see a planar diagram which corresponds to a triangulation of the sphere. Illustration from [316].

where g is the genus of the surface. This implies that the perturbative expansion will be an expression of the form

$$\mathcal{A}_n = \sum_g N_c^{2-2g} f_g(\lambda). \tag{9.22.6}$$

Here $f_g(\lambda)$ is a λ dependent sum over all possible triangulation of a surfaces of genus g. When setting $g_s \propto 1/N_c$ the interpretation of equation 9.22.6 is that it has the same structure as the topological perturbative expansion that we found in string theory. Specifically the expansion in genus for large N_c theories corresponds directly to the expansion in genus in string theory (see equation 9.10.11), and the sum over triangulations of large N_c theories corresponds to the integration over geometries in string theory. These correspondences suggest that large N_c gauge theories have amplitudes of the same form as string theories. Even more strongly we might suggest that in this limit such theories *are* (or are equivalent to) string theories. At this level the correspondence is quite heuristic, but as we will see, the AdS/CFT conjecture provides a concrete implementation of this correspondence.

Open and closed string descriptions of D-branes

D-branes are objects in string theory defined by the endpoints of open strings with Dirichlet boundary conditions. The proper handling of T-duality requires that open strings with Dirichlet boundary conditions be included both in type I theories and type II theories (see section 9.13). We define Dp-branes as D-branes with p spatial dimensions.

We have seen that the low energy limit of superstring theories are supergravity theories. Supergravity theories have a special class of classical solutions called solitons [317]. The solitons of supergravity theories represent extended objects with different possible dimensions. A supergravity soliton with p spatial dimensions is called a p-brane. The p-brane solitons are generalizations of Schwarzschild and Reissner-Nordström black holes in higher dimensions sourced by p-dimensional sources [317]. In comparison, the four-dimensional Schwarzschild solutions are black hole solutions for point like sources. Some p-branes are so called extremal. Extremal means that their charge is equal to their mass (in natural units). This is the maximal mass a black hole can have in GR (or supergravity) for a given charge. For supersymmetric theories with extended supersymmetry and central charge Z, the mass M of the states in a representation satisfies the Bogomol'nyi-Prasad-Sommerfield (BPS) bound $M \leq |Z|$. States that saturate the BPS bond and are called BPS states,

and extremal p-branes are BPS states [317].

A major discovery by Polchinski was that extremal p-branes can be identified (in a certain sense) with Dp-branes [261][238]. This correspondence means that we interpret the Dp-brane and the p-brane description to be a description of the same object but in two different domains. The Dp-brane description is appropriate in the small g_s limit, and the p-brane description is appropriate in the large g_s limit. This is the p-brane/Dp-brane correspondence. This correspondence is often rephrase as an open-strings/closed-strings correspondence. The large g_s limit corresponds to a situation where the p-brane represents an object with black hole like gravitational interaction governed by closed strings. The small g_s limit corresponds a situation where the excitations of the inert Dp-brane is represented by the dynamics of open strings endpoints. The description of open string endpoints can be given in terms of a $\mathrm{U}(1)$ gauge field and a scalar field living on the D-brane itself. For n coincident D-branes the corresponding QFT living on the D-brane is a $\mathrm{U}(n)$ Yang-Mills theory (see e.g. [318]).

Note that the two corresponding descriptions that we have given are not identical. They are descriptions of the same object in two different domains. We shall shortly see that both the calculation of black hole entropy and the AdS/CFT conjecture is based on exploiting this idea. In string terminology the objects we have variously described both as extremal p-branes and as Dp-branes, are usually just referred to as Dp-branes. One then says that in such and such limit this Dp-brane can be described by closed strings (the p-brane or gravity viewpoint), and in another limit it can be described by open strings (the D-brane or gauge theory viewpoint). We shall adhere to this terminology.

Black hole entropy

After Bekenstein, Hawking and others had launched their ideas, it seemed clear that the entropy of a black hole was indeed proportional to the area of the black hole horizon. And while the temperature of a black hole could be partially understood in terms of the Hawking radiation, there was no corresponding explanation for the origin of the entropy. String theory was a conjectured theory of quantum gravity and it was hoped that it could provide such an explanation. In 1996 this hope found at least partial fulfilment when Strominger and Vafa managed to construct a state based on D-branes, representing a charged 5-dimensional black hole with non-zero horizon [313]. The Bekenstein-Hawking entropy of the Strominger-Vafa system is given by

$$S_{BH} = 2\pi\sqrt{\frac{Q_H Q_F^2}{2}}. \tag{9.22.7}$$

From the corresponding description in terms of open strings ending on D-branes they were able to calculate the number of available microscopic states corresponding to this black hole. They found the microstate based entropy to be given by

$$S_{\text{Microstates}} = 2\pi\sqrt{\frac{Q_H(Q_F^2+1)}{2}}, \tag{9.22.8}$$

which agrees with the black hole entropy in the limit of large charges. This was a concrete example of using corresponding open-string and closed-string description of D-branes. This correspondence is not a duality, the descriptions are not identical. However, because of the BPS nature of the D-brane states, the counting of microstates can be carried over from the weak coupling domain to the strong coupling domain. We refer to [317] or [208] for a fuller treatment of black holes in string theory.

Maldacena conjecture

Maldacena's insights, which led to the AdS/CFT conjecture, came from looking closely at the corresponding open-strings/closed-strings description of a specific black hole related configuration

of N coincident D3-branes in type IIB string theory. We have already explained how open-strings/closed-strings description is a form of gauge/gravity correspondence. The specific setup used by Maldacena leads to the idea that in this case the gauge and the gravity descriptions are not only corresponding but are actually dual. Specifically, Maldacena conjectured that type IIB string theory on the space $AdS_5 \times S^5$, is equivalent to $\mathcal{N} = 4$ SU(N) Yang-Mills theory defined on the flat four-dimensional Minkowski space of $\partial(AdS_5)$ [305] (see figure 9.15).

Let us now give a brief derivation of the AdS/CFT correspondence that Maldacena constructed, and see how it leads one to suggest the AdS/CFT duality. The setup is based on a set of N coincident D3-branes in type IIB string theory. We start by considering type IIB theory in (9+1)-dimensional Minkowski space, with spacetime coordinates $(x^0, \ldots, x^9)$. Let us then consider N D3-branes which are all coincident and extended along the dimensions $x^0, \ldots, x^3$, with coordinates in the transverse dimensions being given by $x^4 = \ldots = x^9 = 0$. The action for this this system consists of three parts. We write it as

$$S_{tot} = S_{open} + S_{closed} + S_{int}, \tag{9.22.9}$$

where the subscripts denote the open string sector action, the closed string sector action, and the action for interactions between the open and closed sectors.

Consider first the D-brane world-volulme gauge theory limit. This is the limit where g_s is small and N stays constant, which makes $g_s N \ll 1$. We then consider the low-energy limit where $E/E_S \to 0$. In this limit, any string excitations beyond the massless level decouples and are irrelevant, which means that only massless states contribute. This means that S_{open} is well approximated by a (3+1)-dimensional U(N) Yang-Mills theory living on the D3 brane(s). Type IIB string theory has $\mathcal{N} = 2$ supersymmetry in (9+1)-dimensions, and this corresponds to $\mathcal{N} = 8$ on a $(3 + 1)$-subsurface. However, the D-branes reduce the supersymmetry by half [306], and we get $\mathcal{N} = 4$ supersymmetry in the $(3 + 1)$-dimensional $D3$-brane theory. Thus S_{open} is given by $\mathcal{N} = 4$ U(N) theory. For reasons we won't go into we can extract a free U(1) component from this theory which means that the S_{open} action is given as $\mathcal{N} = 4$ SU(N) plus a free $\mathcal{N} = 4$ U(1) theory. The S_{closed} action is the IIB supergravity action, but since we are taking the low energy limit $\alpha' \to 0$ we find that $\kappa \propto \alpha'^2$ also goes to zero, and therefore we are just left with free IIB supergravity. From this it also follows that there are no gravitational interaction between the closed strings and the D3-branes and we get $S_{int} = 0$. When we collect our result we find that

$$\begin{aligned} S_{tot} &= S_{open} + S_{closed} + S_{int} && (9.22.10)\\ &= \text{Super Yang-Mills } \mathcal{N} = 4,\ \text{SU}(N) && (9.22.11)\\ &\quad + \text{Free } \mathcal{N} = 4,\ \text{U}(1) && (9.22.12)\\ &\quad + \text{Free IIB supergravity on } \mathbb{R}^{9,1}. && (9.22.13) \end{aligned}$$

Let us now look at the p-brane gravity limit. This is the limit where $g_s \to 0$ and $N \to \infty$, which makes $g_s N \gg 1$. Also in this case consider the low-energy limit. We start again with

$$S_{tot} = S_{open} + S_{closed} + S_{int}. \tag{9.22.14}$$

This time S_{open} and S_{int} are both zero. The S_{closed} decomposes into a a near-horizon IIB supergravity on $AdS_5 \times S^5$ and a far away IIB supergravity on $\mathbb{R}^{9,1}$. This is simply the supergravity solution corresponding to the given D-brane configuration. In addition there is a free $\mathcal{N} = 4$ U(1) theory. Summing up the result for this case we get

$$\begin{aligned} S_{tot} &= S_{open} + S_{closed} + S_{int} && (9.22.15)\\ &= \text{IIB supergravity on } AdS_5 \times S^5 && (9.22.16)\\ &\quad + \text{Free } \mathcal{N} = 4\ \text{U}(1) && (9.22.17)\\ &\quad + \text{Free IIB supergravity on } \mathbb{R}^{9,1}. && (9.22.18) \end{aligned}$$

It is now natural to speculate that IIB supergravity on $AdS_5 \times S^5$ and super Yang-Mills with $\mathcal{N} = 4, \mathrm{SU}(N)$ on ∂AdS_5 might be exactly equivalent (at least in some domains of the parameter space). Extending this away from the low-energy limit this becomes the statement that IIB string theory on $AdS_5 \times S^5$ is equivalent to super Yang-Mills with $\mathcal{N} = 4, \mathrm{SU}(N)$ on ∂AdS_5 This speculation is the Maldacena conjecture, later known as the AdS/CFT conjecture. While this conjecture is certainly deep and startling, we can see from our study of the large N_c expansion of Yang-Mills theories that it is not totally unexpected that large N_c theories could be dual to string theory. We can also verify that the symmetries of the conformal theory, which are the $\mathrm{SO}(2,4)$ conformal symmetry and the R-symmetry group $\mathrm{SU}(4)$, matches up the with the string theory symmetries, which are the $\mathrm{SO}(2,4)$ isometry group of AdS_5 and the isometry group $\mathrm{SO}(6) \simeq \mathrm{SU}(4)$ of S^5.

There are several immediate consequences of the Maldacena conjecture, and we will discuss a few of them. As the dual theories of string theory on AdS and CFT on Minkowski space are defined in quite different languages, there must exist a dictionary translating the terms of one theory to those found in the other. Several items for such a dictionary was provided soon after Maldacenas suggestion by Witten in [319] and Gubser et al. in [320]. In the next section we will investigate a very important part of the dictionary which relates entanglement entropy on the boundary to areas in the bulk [321, 322].

In AdS/CFT a question arises as to where the extra bulk dimension comes from. The most prominent suggestion is that this dimension arises as the energy (or scale) dimensions of the RGE. The boundary theory is a scale invariant CFT which makes it possible to have a macroscopic dimension arising from the RGE scale parameter (see [323] and reference therein). The S^5 dimensions are treated differently. In this case on can match up the Kaluza-Klein modes of the bulk with fields in the boundary [306].

One can set up the AdS/CFT conjecture so that it applies to the bulk situation of a black hole in (asymptotic) AdS space. In this case the CFT states dual to this bulk spacetime is a thermal state [324]. While it is not possible to calculate the entropy of such a state, it proposes an understanding of what the black hole horizon entropy is the entropy of. It suggestes that the black hole entropy is the entropy of the CFT thermal state on the boundary of AdS. While this is not likely to be accepted at the final word in this matter, it is a concrete starting point for further investigations. This also suggest that the black hole information loss paradox is (re)solved in favor of no loss of information. This is due to the fact that in the AdS/CFT perspective the quantum gravity theory governing black hole evolution is equivalent to a unitary QFT (in the form of a CFT). This eqiuvalence prohibits information loss. However, some researchers suggest that this argument is too simplistic to be able to claim that the matter is settled [304].

The AdS/CFT conjecture relates two different theories which are each defined on separate physical domains with regards to couplings. The couplings are different but related on the two side of the correspondence. For a given set of corresponding couplings usually only one of the theories is in a domain where calculations can actually be made. It is therefore important to know about the various physical domains (or limits) of the AdS/CFT relation and what calculational approximations can me made in those physical domains (or limits). There are two free parameters on each side of the correspondence. (In this discussion we will refer to the number of D-branes N by the parameter N_c.) On the CFT side we have the parameters N_c and $\lambda = g_{YM}^2 N_c$ representing the number of colors and the 't Hooft coupling respectively. On the AdS side we have the string coupling g_s, and the ratio α'/L^2 of the string length squared $\alpha' = l_s^2$ and the AdS radius L. The connections between the parameters are given by

$$4\pi g_s = g_{YM}^2 \tag{9.22.19}$$

$$\frac{L^4}{\alpha'^2} = \lambda = g_{YM}^2 N_c. \tag{9.22.20}$$

The g_s parameter is the expansion parameter for the string theory perturbative expansion. Let us for the moment refer to the theory on the AdS side as gravitational theory (for reasons that will become clear). If g_s is very small, only the tree level diagram contributes and we are in the classical domain. If g_s is larger, but still smaller than one, we are in the domain of perturbative gravitational theory. If g_s is close to one or larger, we are in the non-perturbative domain of gravitational theory. The ratio α'/L^2 determines whether excitations beyond the massless level are relevant. When α'/L^2 is close to zero we can treat the string as being well represented by the low-energy effective point particle theory. When α'/L^2 is larger, the massive excitations of the string starts to play a role, and it is no longer well approximated by the low-energy effective point particle theory. We present and overview of these connections in table 9.9.

One thing we can extract from table 9.9 is that the CFT side in some cases must be said to define the AdS side. This is often mentioned by saying (in casual terms) that "the AdS/CFT conjecture gives a non-perturbative definition of string theory/quantum gravity". The non-perturbative aspect presents an interesting opportunity for interfacing AdS/CFT and string theory with LQG. LQG could provide a definition of the AdS side of AdS/CFT duality in the case where the AdS theory is represented by "non-pertubative supergravity" [144]. Using the dimension independent reformulation of LQG that we briefly look at in section 8.12 supergravity on AdS (or asymptotic AdS) can be quantized non-perturbatively.

Name	**CFT side**	**AdS side**	**Ads theory**
Definitional A	Arbitrary N_c, λ	$\alpha'/L^2 \neq 0$, $g_s \geq 1$	Non-perturb. string th.
Weak form	$N_c \to \infty$ and λ large	$\alpha'/L^2 \to 0$ and $g_s \to 0$	Class. sugra
Strong form	$N_c \to \infty$ and λ fixed	$\alpha'/L^2 \neq 0$ and $g_s \to 0$	Class. string th.
Definitional B	N_c fixed and $\lambda \to \infty$	$\alpha'/L^2 \to 0$ and $g_s \geq 1$	Non-perturb. sugra

Table 9.9: Various limits of the AdS/CFT conjecture. By the labels "definitional A" and "definitional B" we mean that the AdS side of the theory is strictly speaking none-existent (or at least undefined/unknown). There are no non-perturbative definitions of string theory or quantum supergravity. We have set $g_s \geq 1$ as a limit for the theories to be perturbative. In practice any theory where it is not the case that $g_s \ll 1$ is not perturbative. We have not included specific limits where string theory or supergravity is perturbative. Abbreviations: class. = classical, sugra = supergravity, perturb. = perturbative, th. = theory

9.23 Entanglement entropy

Introduction

The discovery of gauge-gravity duality has led to many new insights into both sides of the duality. On the quantum gravity side, the problem of black hole information loss is in principle solved by linking quantum gravity to a unitary QFT on the boundary. On the QFT side, the understanding of many strongly coupled system have been enhanced by enabling previously impossible calculations to be made by performing equivalent calculation in classical gravity. Still, even in the light of all these achievements, the understanding of the deeper origin of gauge-gravity duality is still lacking. Why are gauge theories dual to gravity theories? One possible approach to a deeper understanding is to look for universal ingredients of quantum theories (like QFTs and CFTs) that can be related to generic features on the gravity side. This could bring us closer to understanding exactly what quantum properties of the boundary theory brings about the dual relationship with gravity theory. One such universal quantum property is entanglement.

Entanglement entropy in QM

In quantum theory all composite systems can have entangled states. Entangled states are quantum states of composite system that cannot be expressed as a single tensor product of states of the individual subsystems. This means that in such states the sub-systems can no longer be regarded as independent systems, but must be seen as being linked together. This section details how to understand and calculate a specific property of entangled states called the entanglement entropy. Before we get to that we want to make a brief comment on where the phenomena of entanglement itself originates.

In chapter 2 we argued that any physical system, whether it is classical or quantum, can be specified by an algebra (and its associated algebra states). In this setup, composite systems, again both classical and quantum, are specified by tensor products of algebras (see section 2.6). For commutative algebras this leads by the commutative Gelfand theorem to Cartesian products of functions spaces. In the noncommutative case it leads to tensor products of operators operating on tensor products of Hilbert spaces. This shows why entanglement is a unique and omnipresent property of quantum system but is not present in classical systems (see section 2.6).

Entanglement is most easily understood in the context of simple systems like a pair of spins. The use of the word spin in this context is conventional but not important, a pair of spins is just a system consisting of two two-state sub-systems. Let the Hilbert space of these two subsystems be given as $\mathcal{H}_A = \mathbb{C}^2$ with a basis $\{|\uparrow\rangle_A, |\downarrow\rangle_A\}$, and $\mathcal{H}_B = \mathbb{C}^2$ with a basis $\{|\uparrow\rangle_B, |\downarrow\rangle_B\}$. The Hilbert space of the composite system is then $\mathcal{H}_{tot} = \mathcal{H}_A \otimes \mathcal{H}_B$. This is a concrete realization of the axioms for composites system that we mentioned in the previous paragraph. The consequence of this tensor product structure is that, while all states $|\psi\rangle_{tot}$ in $\mathcal{H}_{tot}$ can be expressed as

$$|\psi\rangle_{tot} = \alpha\,|\uparrow\rangle_A \otimes |\uparrow\rangle_B + \beta\,|\uparrow\rangle_A \otimes |\downarrow\rangle_B + \gamma\,|\downarrow\rangle_A \otimes |\uparrow\rangle_B + \delta\,|\downarrow\rangle_A \otimes |\downarrow\rangle_B\,, \tag{9.23.1}$$

there are states $|\psi\rangle_{tot}$ in $\mathcal{H}_{tot}$ that cannot be expressed as

$$|\psi\rangle_{tot} = (\alpha\,|\uparrow\rangle_A + \beta\,|\downarrow\rangle_A) \otimes (\gamma\,|\uparrow\rangle_B + \delta\,|\downarrow\rangle_B). \tag{9.23.2}$$

Such states, i.e. those that cannot be written as a single tensor product, are called entangled states. Note that this property is independent of the chosen basis (unlike e.g. the property of being a superposition). The canonical example of an entangled states is given by the Bell states. One example of a Bell state is

$$|\psi\rangle_{tot} = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_A \otimes |\downarrow\rangle_B - |\downarrow\rangle_A \otimes |\uparrow\rangle_B\right). \tag{9.23.3}$$

In many cases representing states by density operators is preferable to representing states by elements in the Hilbert space. In this formalism a state $|\psi\rangle \in \mathcal{H}$ is represented by the linear operator $|\psi\rangle \otimes \langle\psi|$, where $\langle\psi| \in \mathcal{H}^*$. In physics the density operator is often written using dyadic notation as $|\psi\rangle \otimes \langle\psi| = |\psi\rangle\,\langle\psi|$. We will use either notation as we see fit.

The density operator formalism can be used to extend the ordinary quantum state concept to more general quantum states. Ordinary quantum states, which are defined as elements $|\psi\rangle \in \mathcal{H}$, are called pure states. (More properly pure states are defined as one-dimensional subspaces or rays.) We designate the density operators corresponding to such pure states as pure state density operators. We can then generalize the density operator to allow density operators that are convex sums of pure state density operators. Density operators that are not equivalent to pure state density operators, are called mixed states (mixed state density operators). The intuitive interpretation of these mixed states is that the coefficients of the convex sum represent classical probabilities that the quantum system is in the corresponding pure state. In physics terminology this is often referred to as an ensemble, but we prefer the more mathematically correct term probability distribution.

States of composite systems can be described by tensor products of the density operators describing the individual systems. Imagine a system A in the state ρ_A, and a system B in the state ρ_B. The systems A and B are assumed to be separate and independent. The state of the composite system of A and B is given by $\rho_{AB} = \rho_A \otimes \rho_B$. Let us now imagine going in the opposite direction, where we are given a density operator ρ_{AB}, describing the composite system, and we want to find a meaningful description of system A by a density operator ρ_A. The density operator for A is given by the so called reduced density operator which is defined by

$$\rho_A = \mathrm{Tr}_B\, \rho_{AB} = \sum_i \langle b_i |\, \rho_{AB}\, | b_i \rangle\,. \tag{9.23.4}$$

As indicated by the explicit sum over the basis states $|b_i\rangle$ of $\mathcal{H}_B$, the operator Tr_B takes the partial trace over the system B ($\mathcal{H}_B$). Let us now focus on the state of subsystem A for our example Bell state. We easily calculate that the reduced density operator ρ_A is given by

$$\rho_A = \frac{1}{2}\left(|\uparrow\rangle_A \langle\uparrow|_A + |\downarrow\rangle_A \langle\downarrow|_A \right). \tag{9.23.5}$$

A general feature of entangled states is that the corresponding reduced density operator for a subsystem is generally a mixed state. Note that we can construct a pure state that reproduces the measurements that would results if we restricted those measurement only to the basis $\{|\uparrow\rangle_A, |\downarrow\rangle_A\}$, but that for general measurements there are no pure states that reproduce the measurements that would result from a mixed state.

Entropy is a measure of the amount of uncertainty in a probability distribution, or equivalently a measure of the information we would gain by measuring the random variable associated to a probability distribution [325][35]. For a classical system the relevant measure of entropy is the Shannon entropy, while for a quantum system it is the von Neumann entropy [35]. The von Neumann entropy of a density operator is defined as

$$S(\rho) = -\,\mathrm{Tr}\log\rho. \tag{9.23.6}$$

We can let the entropy of a subsystem be a measure of the entanglement of the subsystem with the rest of the system. When seen in this perspective we call the von Neumann entropy of the subsystem the entanglement entropy of the subsystem.

Entanglement entropy in field theories

The same concept of entanglement entropy that we have defined for quantum mechanical system, can also be applied to systems described by quantum field theory. What we are after is some kind of analog of the entanglement entropy for geometrically defined subsystem. The first result in this direction were by Srednicki [326] and Bombelli et al. [327]

We want to consider entanglement entropy for an ordinary free scalar field theory defined on four-dimensional Minkowski space. We consider a fixed time-slice $\mathbb{R}^3$ and we ignore dynamics. Let us select some spherical area A of $\mathbb{R}^3$, which divides $\mathbb{R}^3$ into two regions A and B with $B = \bar{A} = \mathbb{R}^3 - A$. We would then want to define the entanglement of the two regions A and B. To be able to do this we need to make the Hilbert space of the theory decompose in bipartite manner as $\mathcal{H}_{tot} = \mathcal{H}_A \otimes \mathcal{H}_B$. The theory must also be regularized in both the UV and IR as the entanglement entropy will be a divergent quantity. The UV regularization consists of defining the theory on a lattice with lattice spacing ϵ. The IR regularization consisting of defining a spherical ball of radius $L > R$ and then redefining the region B to be the complement of A in the radius L ball. We then view the system as a set of coupled harmonic oscillators, with one oscillator sitting at each lattice point. The techniques demonstrated for QM case can now be applied. As for the QM case what we want to calculate the von Neumann entropy of the ground state density matrix after we trace over

the states of subsystem B. (Equivalently one can trace over the states of A, which is what is usually done in practice.) Applying this to the system ground state $\rho_0 = |0\rangle \langle 0|$, and using a ball-shaped region A or radius R, we get results of the form [326] [327][328][329]

$$S(A) = c\frac{R^2}{\epsilon^2} + \ldots \propto \frac{Area(\partial A)}{\epsilon^2}. \tag{9.23.7}$$

The result is symmetric between A and B. We call $S(A)$ the entanglement entropy of the subsystem defined by the region A. Note that this notion is really the entanglement entropy for some state associated to the region A, and that the entanglement does depend on the state. In many cases one is implicitly referring to the ground state, but this is not always the case.

We can also consider the special case of (1+1)-dimensional of the scalar field example, which means that we are looking at a QFT defined on say the real line. In this case we do not expect the area rule to hold, since the boundary in this case would just be two individual points. We define the region A to be a conneted region of length R, and we are still using the regulators ϵ and L. The results one gets are of the form [330][331]

$$S(A) = \frac{1}{3}\log\frac{R}{\epsilon} + \ldots. \tag{9.23.8}$$

For conformal (1+1)-dimensional theories with central charge c, one gets [330][331]

$$S(A) = \frac{c}{3}\log\frac{R}{\epsilon} + \ldots. \tag{9.23.9}$$

(See [328][329] for a review.) We will soon see how this is generalized to arbitrary regions and arbitrary states in holographic CFTs.

Ryu-Takayanagi area formula

In the previous subsections we have considered entanglement entropy states associated to spatial subareas in field theories. This has been done for generic scalar field theories, and we have not assumed that these field theories are dual to a theory of gravity. Let us now consider the situation where such a dual theory exists. Imagine now AdS_3/CFT_2 where a (1+1)-dimensional CFT is defined on the boundary of (2+1)-dimensional AdS (the bulk). Let us consider spatial slices of both the boundary and the bulk. We have already stated that for the ground state the entanglement entropy of a length R subsystem is given by

$$S(A) = \frac{c}{3}\log\frac{R}{\epsilon} + \ldots. \tag{9.23.10}$$

The subsystem A is just a line segment and the boundary of A is given by its two endpoints a, b. A geometric question one can ask is: What is the minimal length curve through the bulk that starts at a and ends at b. It turns out that the answer to this question is given by same formula as the one given for the entanglement entropy [321]. The regularization ϵ now represents a regularization that defines a new surface inside AdS which is close to the true boundary since otherwise the length would be divergent (we will not give the exact details, see [321]). This correspondence between entanglement entropy and path length for CFT_2/AdS_3, in combination with the ideas from black hole entropy, could lead one to speculate that in a setting where a CFT has a dual gravitational theory, there might be some general connection between the entanglement entropy of a region A on the boundary and some area functional in the bulk. This turns out to be a fruitful path to explore.

Consider a CFT living on the boundary of an asymptotic AdS space. The CFT is dual to a gravity theory on the asymptotic AdS space. Let Σ be a spatial slice of the boundary. Consider now

subdividing the spatial slice into regions A and $B = \overline{A} = \Sigma - A$. We assume that all states can be expressed as a sum of elements from

$$\mathcal{H}_\Sigma = \mathcal{H}_A \otimes \mathcal{H}_B. \tag{9.23.11}$$

We define the entanglement entropy of a state $|\psi\rangle$ with respect to the region A just as we did for a scalar field theory. Since the CFT is dual to a theory of gravity the state $|\psi\rangle$ is dual to a spacetime M_ψ on the AdS side. Ryu and Takayanagi suggested in 2006 that, in CFT systems with a holographic dual, the entanglement entropy of a state $|\psi\rangle$ with respect to the area A is related to the area of a minimal area surface A' in the space M_ψ (see figure 9.18 for the setup). We require that $\partial A = \partial A'$ and that A is homologous to A' [321, 322]. The Ryu-Takayanagi holographic entanglement entropy formula can be written as

$$S_A = \inf_{\partial A = \partial A'} \left(\frac{\text{Area}(A')}{4G} \right). \tag{9.23.12}$$

This formula is postulated to apply for any state $|\psi\rangle$ and for any region A. Note that even though the formula is the same for all states, the spacetime, or rather the AdS spatial slice, is (potentially) different for each state. The Ryu-Takayanagi formula connects the geometry of the bulk to the entanglement structure of the boundary. Ryu and Takayanagi gave some arguments in favor of this conjecture [321, 322], and further stronger arguments were presented in [332]. A covariant version is given in [333].

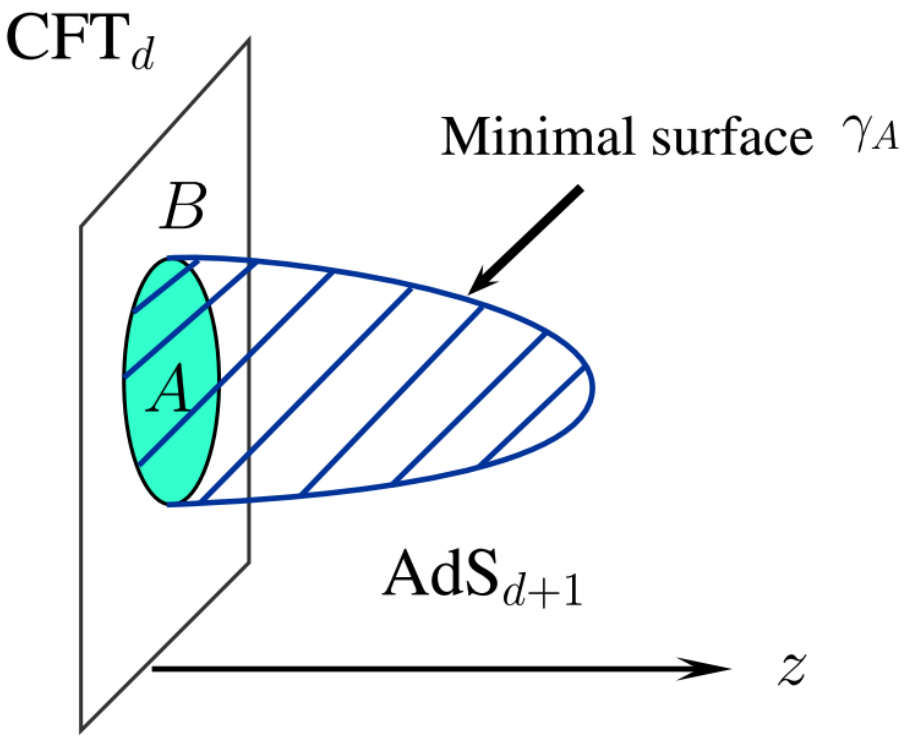


Figure 9.18: A geometric picture of the Ryu-Takayanagi holographic entanglement entropy formula. We have labeled γ_A as A' in the text. Illustration from [334].

Entanglement entropy as a purely quantum mechanical concept, without any reference to holography, satisfies several important properties [35][329]. These properties are independent of

the system in question. First of all the entanglement entropy with respect to a region A is equal to the entanglement entropy of the complementary region $B = \bar{A}$ (for pure states). We suppress the reference to the state ψ and write this simply as

$$S(A) = S(\bar{A}). \tag{9.23.13}$$

In addition entanglement entropy satisfies the property of strong sub-additivity, defined as

$$S(A+B+C) + S(B) \leq S(A+B) + S(B+C) \tag{9.23.14}$$

$$S(A) + S(C) \leq S(A+B) + S(B+C) \tag{9.23.15}$$

From strong subadditivity we can deduce that entanglement entropy also satisfies subadditivity, defined by

$$S(A) + S(B) \geq S(A+B). \tag{9.23.16}$$

Space through entanglement

We have seen that the Ryu-Takayanagi conjecture ties the entanglement structure of the boundary tightly to the geometry of the bulk. In this section we will study what we can deduce about the bulk from this conjecture.

One quite intuitive consequence of Ryu-Takayanagi is as follows. Consider two AdS/CFT systems where the boundary is S^n. The two CFT systems are represented by Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. Thus states of the combined system are represented by linear sums of elements of $\mathcal{H}_1 \otimes \mathcal{H}_2$. However, even if the systems 1 and 2 are independent and non-interacting we can still define states where the systems 1 and 2 are entangled. The question is then: What is the bulk interpretation of the entangled states? The suggestion is that entangled states of CFT1 and CFT2 are dual to a connected spacetime. Thus in the dual picture of the bulk, the entanglement of the two independent non-interacting boundary theories, leads to a sort of fusion between their dual spacetimes. To strengthen the case for this interpretationo let us recall Maldacena's model of a maximally extended AdS black hole [335]. In a maximally extended black hole there are two separate asymptotic regions that are only connected through the interior of the black hole. Maldacena suggest that corresponding to the two asymptotic regions there are two separate and non-interacting CFTs and that the state dual to the spacetime is a state that entangles the two CFTs. One can then make the case that the proper interpretation of the AdS/CFT setup for a maximally extended black hole is that it is the (cross-boundary) entanglement of the state that holds the two regions of spacetime together.

Let us follow Raamsdonk [336] and expand on this idea in a similar setup. Consider now a CFT, with a gravity dual, where we for simplicity imagine the CFT to be defined on a sphere. We let a vertical plane $\tilde{A}$ divide the sphere into two hemispheres called A and $\overline{A}$, and we assume $\tilde{A}$ is the extremal surface corresponding to region A. We consider the entanglement entropy with respect to A of the state $|\psi\rangle$ dual to the spacetime. From Ryu-Takayanagi this is equal to (proprotional to) the minimal area of the surface $\tilde{A}$ dividing the bulk. Now, imagine perturbing the state $|\psi\rangle$ of the CFT such that the entanglement between A and $\overline{A}$ is decreased. This must then decrease the minimal AdS surface that ends on $\partial A = \partial\overline{A}$. In the limit of the entanglement going to zero the two hemispheres pinch off and disconnect [336] (see figure 9.19). By manipulating the states and reducing their entanglement entropy, one can make space itself disintegrate into smaller disconnected parts. This suggest that the structure of entanglement entropy is responsible for the appearance/occurrence of connected semi-classical space.

GR from entanglement

We now consider the so called first law of entanglement [338]. To state this law we use the following setup. Consider a CFT defined on the space $M = \mathbb{R}^{D,1}$ and let ρ be a density matrix corresponding

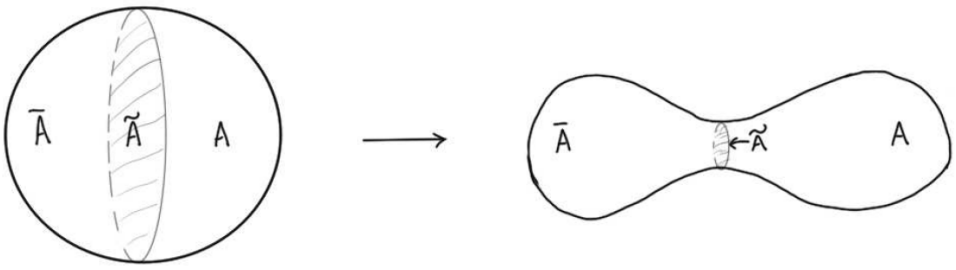


Figure 9.19: Disconnecting bulk space by reducing entanglement entropy in the dual boundary theory. Illustration from [337].

to the state $|\psi\rangle$. Let $A \subset M$ be some spatial region of M. We denote the reduced density matrix with respect to A by

$$\rho_A = \mathrm{tr}_{\overline{A}} \langle\psi|\psi\rangle \,. \tag{9.23.17}$$

Any such states defines an operator H_A by setting

$$\rho_A = e^{-H_A}. \tag{9.23.18}$$

The operator H_A is called the modular Hamiltonian. Let us denote the entanglement entropy of ρ_A by $S(A)$. The first law of entanglement can be expressed as

$$\delta S_A = \delta\langle H\rangle. \tag{9.23.19}$$

By considering a one-parameter family of CFT states $|\psi(\lambda)\rangle$, representing perturbations of the state $|\psi(\lambda = 0)\rangle$, one can derive interesting bulk consequences. We will not have time to consider this in detail but we list the main conclusions. If we consider the dual theory of a boundary CFT respecting this equality the first law implies the linearized classical Einstein equations must be satisfied in the bulk [339]. Considering changes beyond the first order in the same setting one can also derive the universality property of gravity [340]. By universality we mean the universal coupling of the gravitational field to the stress-energy tensor.

Geometry from quantum foundations

We want to end our discussion of entanglement entropy by considering it in a more general context. The viewpoint in the previous subsections have been mostly tied to the AdS/CFT conjecture and string theory. In this section we try to escape from those restrictions and look at a more general picture of spacetime geometry. We also want to tie what we have been discussing to research within LQG and other areas.

The AdS/CFT duality and the Ryu-Takayanagi area conjecture indicates that there is a strong link between entanglement structures in some quantum theories on fixed backgrounds and geometric theories of gravity. It is not clear if the Ryu-Takayanagi formula should be thought of a assigning entanglements entropies to regions of the bulk, but this is one interesting interpretation. This interpretation has to be made with some care as the minimal surface in the bulk is not unique and jumps discontinuously upon continuous changes of the boundary region [341]. Still it is worthwhile to pursue this interpretation while keeping such caveats in mind. Given this interpretation we can find connections to entanglement entropy structures in LQG. In LQG regions of space can be given assigned an entanglement entropy based on calculation that are similar to those used for calculating black hole entropy in LQG [342][343][344]. A postulate that has been made that the appearance of semi-classical spacetimes is conditioned upon the entanglement entropy area law to

hold [345][346]. Or in other words that it is a general property of semi-classical geometries that entanglement entropy of a region is proportional to area of the region. This is a vast and powerful generalization of the Bekenstein-Hawking formula for the entropy of black hole horizons. The area law criteria has been used in LQG to facilitate the search for a state approximating semi-classical spacetime [346].

From a different perspective, Jacobson has shown that if one assumes that the entanglement entropy area law is valid one can (given some other assumptions) derive the theory of general relativity [347][348][349]. This sees at least superficially similar to the derivation of GR from applying the Ryu-Takayanagi formula in the AdS/CFT setting, which we discussed previously.

We have mentioned the idea that it is the appearance of the area law can serve as a signal for the presence of semi-classical spacetime. It is also clear that there are many "geometries" beyond such semi-classical one. LQG gives a quantum description of them and so does potentially string theory through AdS/CFT (we do not consider matrix models). It is still natural to ask if there is a more generic description of those "geometries" which are not geometries. It is possible that ncg could serve a role in this capacity but for now this is just speculation.

Raamsdonk, Swingle and other have emphasized how space in a sense emerges from the entanglement structure of a quantum theory [350][336]. However, the quantum theories in question are typically CFTs or lattice like solid state models (tensor network models). In these quantum models there is no dynamic space but a background space is still defined. Thus one can say that dynamic space arise from the entanglement structure of quantum theories defined on fixed spaces. It would be interesting to try and extend this emergence of space to an even more abstract setting. Would it be possible to start from an abstract quantum theory defined with no reference to any concept of space and derive emergent space from the entanglement entropy structure of this quantum theory? This type of geometry from quantum theory construction have been attempted in recent papers [351][352][353] (see also [354]). Cao et al. in [351] uses mutual information to construct a metric on a discrete space made of points defined by subspaces of the Hilbert space. This discrete space is then embedded in a continuous manifold and in this way space emerges from quantum entanglement. We will explore some aspects of this in section VII of chapter 10.

9.24 Notes

History

A brief history of string theory goes as follows. In the 1960s physicist where attempting without success to make a QFT that could handle hadron scattering. The lack of progress led some people to question if QFT was really the right tool to describe fundamental particles. A program called the S-matrix program, initiated much earlier by Heisenberg, focused on developing criteria that could lead to the construction of scattering amplitudes independent of a QFT. A specific variation and elaboration of the S-matrix idea was the concept of dual models of hadronic scattering. Dual model theory puts an extra criteria of s-t duality on the S-matrix amplitudes (see [142][228]). This duality demand soon proved to be of value.

String theory is usually said to have begun when Veneziano in 1968 published an expression for a scattering amplitude that satisfied the s-t duality criteria, although no concept of strings were involved at the time [355]. Later it was realized that Veneziano's amplitude could be generalized in many ways, and Susskind [356], Nambu [357] and Nielsen [358] all realized that it could be derived from a model of scattering quantum strings. Supersymmetry was (implicitly) invented in the early 70's in the process of finding a way to add fermions to the purely bosonic string models [224]. At the same time as these founding moments of string theory, QFT was on the rebound. The success of QCD demonstrated that QFT could handle hadronic scattering, and interest in string theory as a model of colliding hadrons faded quickly.

String theory was soon headed for new and bigger targets. In the hadronic context the presence of a massless spin-2 state in string theory had been puzzling and undesirable. In 1974 Scherk and Schwartz revived string theory by suggesting that the spin-2 state was actually a graviton, and that string theory was a theory of quantum gravity [359][360]. By this time most physicist were now mostly focused on QFT again, and with the invention of supersymmetry, QFT models with local supersymmetry soon followed [226]. These so called supergravity theories were new prime candidates for a theory of quantum gravity. It was hoped that supergravity in combination with Kaluza-Klein ideas could lead to a unique unified theory (see section 9.20 for more details). The interest in string theory as theory of quantum gravity was therefore minimal

By the early 80s supergravity theories were known to contain divergences [230] and the Kaluza-Klein supergravity program was also beset with other problems, and there was room for a new approach. The only known string theory at the time that could support gauge groups was the type I string. It was known that this theory was anomalous for some choices of the gauge group and this was believed to be true in general. When Green and Schwartz in 1984 [233] showed that type I string theories could contain interesting gauge groups that were free of anomalies, the time was finally right, and interest in string theory rose quickly. Based on the list of possible gauge groups from Green and Schwartz, two new string theories called heterotic string theories were soon invented by Gross, Harvey, Martinec and Rohm [234, 235, 236]. Heterotic string theories were theories of closed strings, which meant they included gravity in addition to non-Abelian gauge groups. This made heterotic string theory a candidate for a unified theory of everything. Candelas, Horowitz, Strominger and Witten showed that heterotic string theories compactified on Calabi-Yau manifolds provided a promising way to deduce the standard model and its supersymmetric extensions [237].

The next big set of events came in the mid 90s. Before these events, some of the perturbative string theories were known to related, but in general the existence of five different string theories was considered a problem. This all improved when M-theory was discovered [239]. All the perturbative string theories could now be related by compactifications and dualities. In addition, and at about the same time, Polchinski and co-workers uncovered that open strings with Dirichlet boundary conditions are actually required in string theory for a consistent implementation of T-duality [238][261] (see also [264][361]). These fixed open string endpoints define physical non-perturbative higher-dimensional objects called Dp-branes. Dp-branes can be identified with extremal p-brane solitons in the corresponding low-energy limit supergravity theories. Thus the Dp-branes admit two corresponding descriptions in terms of open strings with Dirichlet conditions, and in terms of the closed string gravitational field of the supergravity solitons.

The open-closed correspondence for D-branes was soon used by Vafa and Stomminger to give a derivation of black hole entropy based on counting of available microscopic states [313]. In 1997 Maldacena (with elaborations from Witten and Gubser et al.) used D-branes to relate type II string theory in AdS space to a four-dimensional gauge theory [305][319] [320]. This AdS/CFT conjecture introduced the ideas of holography and gauge-gravity duality to string theory, and it provided the first hints of a possible non-perturbative formulation of string theory. In principle the AdS/CFT conjecture also resolved the black hole information paradox in favor of unitary evolution and preserving information [313].

D-branes turned out to also provide many new possibilities for phenomenology. Gauge theories could be introduced in string theory not just from being present in the 10-dimensional world, like in heterotic string theory and type I theory, but also via D-branes in the brane-world scenario. Braneworld models based on intersecting D-branes, instead of just coincident branes, could also produce chiral fermions [362].

The process of compactification is described by specifying some compact six-dimensional manifold. The complete set of such manifolds in any given class can often be parameterized by a moduli space. The coordinates or parameters of the moduli space (some of which can be discrete)

are referred to as moduli. The exact compactification space is not completely specified unless one list the values of all of these moduli. In the low-energy theory, such moduli corresponds to scalar fields, and an unrestricted moduli corresponds to a massless scalar.

Figuring out how to handle the moduli of compactified string theories was a long standing problem. Introducing non-trivial background fields on the compactification manifold led to a solution of this problem [363]. Compactification with non-trivial background fields define what is know as flux compactification. Flux compactifications introduced a way in which the moduli fields of the compactification could receive non-trivial potentials. The non-trivial potentials determines the parameters of the compactified space, and avoids producing a model with a plethora of massless scalar fields.

Flux compactifications solved in principle the question of how low energy particle physics parameters could be determined, but at the same time it revealed a previously hidden enormity of different compactifications. The enormous set of different compactification are known as the string landscape, a landscape with 10^{500} or more valleys, each of which represent a possible low energy string theory [287]. The string landscape coupled with eternal inflation led to the multiverse idea, and this has led to the introduction of environmental models for low energy physics parameters. This includes a sort of "it-is-what-it-has-to-be" perspective on the cosmological constant [287].

In the 2010s the gauge-gravity duality concept has been studied intensely and greatly expanded upon. One major new idea was singling out the property of entanglement entropy of boundary field theories as being of particular importance for determining dual geometric features of the bulk [336]. Ryu and Takayanagi proposed a formula relating the entanglement entropy to the area of minimal surfaces in the bulk [321, 322]. An important earlier idea in this direction was presented by Maldacena in 2001, where he described the spacetime of a maximally extended black hole as being dual to two entangled CFTs [335].

Raamsdonk extended this idea to more general geometries, postulating that entanglement could be the glue that holds spacetime geometries together [336]. This was expanded by Swingle from a condensed matter tensor network perspective as well as by many others [350, 364]. All of this indicated that the gravitational theory of the bulk could be derived from basic information theoretic aspects of entanglement entropy when applied to field theories that were assumed to respect the Ryu-Takayanagi holographic entanglement formula.

Reasoning about the entanglement of Hawking radiation, the equivalence principle, and the AdS/CFT induced unitary black hole development paradigm, has led to further ideas and potential paradoxes concerning black holes. Chief among these are the firewall proposal from Almheiri, Marolf, Polchinski and Sully (AMPS) [365] and the ER=EPR idea from Susskind and Maldacena [366]. No clear consensus on these issues have yet emerged.

Some general facts about supersymmetry and spinors

Several aspect of this chapter builds on specific fact about supersymmetry and spinors. We record some of them here (see [243] or [246] for a review).

In Minkowski space of even dimensions $D = 2m$ there is a unique corresponding Clifford algebra with a unique faithful representation of dimension $2^{D/2}$. This representation is reducible and can be split into two sub-representations of opposite chirality. For odd dimensional Minkowski space the corresponding Clifford algebra has two faithful irreducible representations, both of dimensions $2^{(D-1)/2}$. Only representation of even dimensional Clifford algebras support chirality.

Weyl spinors can be defined in dimensions 2, 4, 6, 8, 10 and 12 (all even dimensions). Majorana spinor can be defined in dimensions 2, 4, 10 and 12 (2 and 4 mod 8). Majorana-Weyl spinors can be defined in dimension 2 and 10 (2 mod 8). Weyl spinor have half the number of components of a Dirac spinor.

Supersymmetry is an extension of ordinary Lie algebra based symmetry concepts to include

symmetries based on graded Lie algebras. The two most important such extensions is the extension from the Poincaré Lie algebra to the super-Poincaré algebra, and the extension from the conformal algebra to the super-conformal algebra. The super-Poincaré algebra have representations that are direct sums of ordinary Poincaré representations, these representations always include an equal number of fermionic and bosonic states. This means that supersymmetry imposes relations between fermionic degrees of freedom and bosonic degrees of freedom.

In addition to the basic super-Poincaré algebra, there are also extended super-Poincaré algebras. The basic super-Poincaré algebra is referred to as having $\mathcal{N} = 1$ supersymmetry and the extended super-Poincaré algebras are referred to as having $\mathcal{N} = n$ for some positive integer $n > 1$. Representation of extended super-Poincaré algebra are always non-chiral.

For theories with $\mathcal{N} = 1$ supersymmetry we can informally describe the most relevant multiplets as a follows. The chiral multiplet consists of a spin-1/2 Weyl fermion (or spin-1/2 Majorana fermion) and a spin-0 complex scalar field. The vector multiplet consist of a spin-1 vector field and a spin-1/2 Majorana spinor. With extended supersymmetry the multiplets are bigger, but we will not need the details.

Theories with local supersymmetry contains gravity and are called supergravity theories. In the $\mathcal{N} = 1$ the gravity supermultiplet consist of a spin-2 graviton represented by metric tensor and a spin-3/2 gravitino.

VI

Extended theories - Evaluation

10. Evaluation and conclusion

> In spite of their empirical success, GR and QM offer a schizophrenic and confused understanding of the physical world. The conceptual foundations of classical GR are contradicted by QM and the conceptual foundation of conventional QFT are contradicted by GR. Fundamental physics is today in a peculiar phase of deep conceptual confusion.
>
> Rovelli

Summary

In this chapter we evaluate noncommutative geometry, loop quantum gravity and string theory. We evaluate them with respect to the two questions "How can we unify gravity with quantum theory?" and "How can we explain the standard model?". In addition we evaluate their foundational structure.

This chapter critically evaluates the three extended theories we have treated. We focus mostly on how well they are able to deal with the questions of explaining the standard model and unifying gravity with quantum theory. In addition, in light of the methodological preferences stated in chapter 1, we also touch upon the foundational qualities of the three theories. Except for the partial postdictions of GR and the SM from NCG, none of the theories have any empirical support, and therefore their evaluation lies outside the domain of empirical science. To make an evaluation we thus have to look at meta-theoretical aspects of these theories. Unfortunately (and probably unavoidably), such meta-theoretical based evaluations do not have an established method to it. It is therefore inevitable and intentional that this chapter will have a freer and more subjective style. Our goal is not to lay down the ultimate truth about these theories. Such a truth does not yet exist. Instead we try to provide useful perspectives that could motivate further study and research. We apologize in advance to all researchers. No critique is totally fair and correct without caveats and limitations. In the interest of clear communication and brevity we temporarily set aside the proviso of being totally correct and fair in the strictest sense of the words.

10.1 Evaluation of NCG

NCG sets up an exceptionally clean and rigorous model for physics. It builds on an extensive mathematical framework that has many interconnections to other areas of mathematics. The CLC models themselves can be characterized in multiple ways. Some suggest that they are similar in spirit to GR, and call them gravitational, but we prefer the term geometric. No matter which word one uses, a key quality is the singular nature of the dynamic principle. It is diffeomorphism invariance, or automorphism invariance, or spectral invariance that is the single principle dictating the theory. The basic setting of CLC models is that of metric geometry based on noncommutative spin manifolds, where the metric is implemented algebraically through the Dirac operator. CLC models manage to express the complete symmetries of a physical system as the automorphism (diffeomorphism) group of the geometry. This setup smoothly and elegantly captures the empirical fact that we observe the low energy symmetries of the world to be $\mathrm{Diff}(M) \ltimes G$ for some gauge group G. CLC models claim to be able to predict the gauge group G from first principles. Let us for the moment postpone the question of the prediction of the gauge group, and first focus on the merits and problems that are independent of this assertion.

GR, or rather GR with some higher-order modifications, can be deduced from a canonical triple, completely independent of any finite triple. Given a specific finite triple CLC models are able to deduce GR plus spontaneously broken Higgs-Dirac-Yang-Mills theories with minimal coupling of matter to gravity. This is certainly a neat and impressive achievement but does it have any deeper significance? Does it go beyond being a neat refactoring of already well established knowledge? There are several possible points one could make.

We first discuss the similarities with Kaluza-Klein theories. In the simplest Kaluza-Klein example we use the space $X = M \times S^1$ and derive four-dimensional GR and electromagnetism from five-dimensional GR. This seems in a sense remarkable but could also be considered slightly trivial in the sense that the isometry of S^1 is $U(1)$ which is what is needed for electromagnetism. The Kaluza-Klein approach can be viewed as a theory with five-dimensional diffeomorphism invariance (let's denote it $\mathrm{Diff}(5)$), where by some unknown mechanism the symmetry is broken to $\mathrm{Diff}(4) \times \mathrm{U}(1)$. However, it is not true that the diffeomorphism group of $M_4 \times S^1$ is $\mathrm{Diff}(4) \times \mathrm{U}(1)$.

The AC model setup is similar, but also substantially different. First of all, for a CLC model it *is* true that the diffeomorphism group of the AC space is $\mathrm{Diff}(4) \times \mathrm{U}(1)$. Secondly, a Kaluza-Klein theory is just demonstrably successful when then infinite set of extra Kaluza-Klein excitations are ignored, and while this is reasonable in a low-energy limit the untruncated compactified Kaluza-Klein theory is still a very different theory from ordinary four-dimensional GR and electromagnetism. The CLC models gets the spectrum right without requiring an infinite set of extra states to be deleted[1]. It therefore seems that CLC models is more than just a compact reformulation of gauge theory.

We would argue to the contrary that we find that the idea of expressing all of physics through the complete symmetries of a single space is a profound idea. It is not in any way obvious that metric geometry could be unified with gauge degrees of freedom in the NCG manner. And although we have argued that symmetry breaking scalar fields are a natural component of QFT, the automatic integration of these structures in CLC models are still significant (note that we are not talking about UV sensitivity naturalness here). In a CLC model the Higgs field is not treated as an extra component but it seems to be an absolutely necessary and natural part of the setup.

While CLC models are in principle flexible with regards to the finite triple input, they are also quite restrictive structures. Because of the use of algebras, only fundamental representations appear for fermions (strictly speaking the conjugate and the adjoint representation are also possible). Even

[1] To be precise the effective bosonic action is usually truncated as high energies. One might suggest that this is similar to the truncation of the Kaluza-Klein excitations.

though the presence of fundamental representation comes as no shock to any particle physicist, we must remark that within QFT there is no reason that fermions should appear in the fundamental representation, or any other simple or small representation. This is therefore a significant improvement in comparison to the unrestricted situation in QFT, where any representation may be used.

Let us return to the question of predicting the standard model. The "official" CLC prediction for particle physics is the $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R \times \mathrm{SU}(4)$ Pati-Salam model. The symmetries of the Pati-Salam model is then spontaneously broken to standard model symmetries. In addition to predicting the gauge group of the model, the number of fermions per generation and some mass relations is also predicted. The number of generations remains a free parameter. What can we make of these predictions? Do they signal that some significant insights have been uncovered? Depending on which set of arguments you are building your predictions (see chapter 6), the Pati-Salam model is either predicted uniquely, or as the simplest non-trivial possibility. The uniqueness argument is still too recent and too ad hoc to be properly evaluated. The argument predicting the model as the simplest non-trivial possibility is solid and deep. A very limited set of auxiliary assumption need to be made for this argument to work, and with some slight amount of leniency the prediction could be said to be very close to a mathematical deduction within the generic CLC model setup.

If the so called first-order condition is kept, CLC models predict the standard model as the simplest non-trivial possibility but with a Higgs mass in excess of the experimental bounds. Though this model is experimentally falsified, it is still relevant in the sense that, with the exception of the Higgs mass parameter, it does very easily predict many central features of the standard model.

The direct prediction of Pati-Salam and the standard model is very interesting. Both as a prediction in itself but also in the light of competing theories. String theory can be said to have partly resigned from the task of strongly predicting the low-energy world business, choosing to see such things as a more stochastic and environmental question. (It might be more accurate to say that string theorist have reinterpreted the role of string theory with regards to the question of unique low-energy models.) While there are certainly arguments that make such an environmental approach preferable, the uniqueness of the standard model (or Pati-Salam model) in CLC is in stark opposition to the stochastic approach. This does not preclude that noncommutative methods at some point could give rise to an environmental understanding of these question, nor that some form of specific predictions could be made in an environmental setup.

There are several weak spots in the CLC models. A simple one to point out is that they are all defined in a Riemannian or Euclidean setup (instead of using the Lorentzian or pseudo-Riemannian formulation). Central parts of CLC theory depend on a Euclidean signature to be consistent. It is hard to tell whether this should be cataloged as a technical problem or as a foundational problem. Recent result have introduced Krein spaces as a possible way to get to Lorentzian signatures [367].

Another issue is the formulation of action principles in CLC models. Usually action principles are integrals defined with "temporal" boundary conditions, specifying some form of initial and final conditions representing initial and final states. In CLC models the integrals are over the complete manifold leaving no room for such conditions. The significance of this is unclear [368].

An issue could be raise against the way the Dirac operator of the finite space is defined. In the standard setup this operator is a fixed finite dimensional operator, which means it is a matrix with constant components. Since the components of this matrix determines the Yukawa couplings and masses of the particles it is perhaps a bit to limiting to set this parameters as fixed and non-dynamic. One would like to think of these Yukawa couplings as something which values are determined by some dynamic principle.

We started by claiming that the general setup in NCG is quite clean and principled, but still it could be said that the true nature of all of its assumptions are not completely clear. It is not clear which of these assumption are technical assumptions and which should be categorized as more physical axioms. More importantly, it seems very demanding to conclude with certainty whether

any given assumption should be classified as ad hoc (or even opportunistic). An intuitive judgment on these matter is often more due to issues of notation and presentation, and is therefore not a reliable indicator. We will present no conclusion on this part but we emphasize the importance of trying to avoid ad hoc adjustment in a theory that derive much of its strength from the seeming natural appearance of physically relevant structures.

It is clearly unsatisfactory that the CLC models are not quantum models. Of course, nothing prevents the quantization of these models anymore than any other model, but as a fundamental theory of physics it is not quite satisfactory that it is formulated as classical theory. One would rather hope and expect that a true quantum theory would naturally emerge from the noncommutative setup. In a way the quantum part is the part where we would expect noncommutative geometry to really excel. However, if the CLC models where actually complete in the sense of including the quantum aspect it would be a full-featured suggestion for a unified model of quantum gravity and particle physics. In addition, since NCG is working at mathematically rigorous level, this would also include a mathematically rigorous definition of QFT. Meeting all these requirements is certainly a tall order, and it should be considered acceptable to approach this goal in increments. Currently CLC models focus on the lower energy domain where QFT lives on a more less fixed (GR determined) background, and the absence of an intrinsic quantum component is not a direct practical problem for CLC models.

Some progress have been made towards defining quantum gravity in the context of CLC model but so far no full featured models have emerged. Recent result have suggested a way to achieve volume quantization (in the sense of spacetime volume quantization) in CLC models [121][122][123]. This proves to have some interesting consequences for dark matter and dark energy. However, this approach seems to still be far from presenting a full dynamical model of quantum gravity. The approach is very recent and though mathematically elegant it seems at first sight to be almost simplistic in physical terms. In our view it is to early to have a firm opinion on this issue.

In a much grander setting one can envision a more complete theory based on a fully noncommutative space as suggested in chapter 6. The CLC models would appear as a low-energy limit derived from this fully noncommutative space. This would be a grand achievement but it is not known how to build such a model. We can also consider CLC-like models in other physical theories. It is possible that CLC models can appear as limits of theories like string theories (see e.g. [259, 260] for such models in the context of CFT). In a much more radical direction there are models based on a noncommutative version of the holonomy algebra of LQG (see the work of Grimstrup and Aastrup in [369][370] [371]). This all goes to show some of the intrinsic power of NCG. In situations where one can derive a noncommutative algebra representing the central aspects of some physical system, NCG provides a possibility to give a geometric interpretation of these data.

With regards to experimental confirmation there is no evidence of any deviation from the standard model, which implies that there is no evidence for any beyond the standard model physics. (Some researcher might object to this statement. We base our claim on the following observation: There are non-particle versions of dark matter, and neutrino masses can still be handled without introducing anything beyond neutrinos of both chiralities.) This also applies to the Pati-Salam semi-grand unified model. The Pati-Salam would in the simplest version introduce one mass scale for the breaking of $\mathrm{SU}(4)$ and another for breaking of $L - R$ symmetry. No indications of the presence of such scales have yet been uncovered. We will not go into any more of the experimental details of such scenarios.

In summary CLC is a strong contender for explaining many aspects of the standard model. It has so far not reached it potential with regards to quantum gravity or a deepening of our quantum understanding. NCG in general represents a tool for both creating new models but also for providing geometric insights into existing physical theories formulated in an algebraic language.

10.2 Evaluation of LQG

Loop quantum gravity can be characterized in two different ways. First it can be labeled as a set of general quantization techniques, that allow principles of quantum mechanics to be applied to connection based diffeomorphism invariant theories in a background free manner. Second it is the specific application of these techniques to the theory of general relativity in four dimensions. The techniques are based on combining gravitational principles with quantum principles in a straight-up minimalistic manner. The actual implementation of this straight-up philosophy is not so straightforward as we saw in chapter 8. We will focus on the second characterization first, that is LQG as the quantization of GR. From a research strategy point of view such a non-creative straight up combination of GR and QFT is something that should be attempted. It is hard to evaluate a solutions based on a more complicated set of principles unless one knows in some detail what using just a minimal set of principles can provide.

The straight-up combination approach brings us back to table 8.1 where we list the fundamental principles of QFT and GR. There is an obvious clash between the classical and the quantum principles that by default (or by history, or by experiment) is scored in favor of the quantum theory. (It should be noted that in the gravity case the preference for the quantum side is so far a purely theoretical preference.) However, in GR vs QFT case the clash and the difficulties does not end with the quantum vs classical divide. The traditional formulation of QM upon which QFT is based, represents time evolution as one-parameter families of automorphisms, or, in more conventional language, as evolution with respect to a global external time parameter. This present no problem when upgrading classical theories that also use such a global external time parameter, but this is not the case for GR, where there is no such parameter.

Before we enter into further details, let us also add a third problem to the list. The fixed Minkowski background of QFT provides a set of tools that are central in the construction and use of QFT. Building a quantum theory in a background free setting deprives the constructor of such tools. Therefore new tools must be invented. It is also worth noting that several aspects or features, which are applicable and important in the fixed background case, loose their validity and pertinence in the background free approach. This includes techniques like perturbation theory and concepts like unitarity.

Let us now return to the issue of the time parameter. The time parameter issue brings into question which principles should (be assumed to) survive when these two theories are combined. In particular what should happen with regards to time and dynamics. Perturbative quantum gravity answers this by siding with QFT, and applies the conventional "dynamics by external time parameter" paradigm. LQG takes a different stand here, and declares that being background free is one of the core tenets of GR, and cannot be disregarded in the union with QFT. Instead, one must accept both background freedom and QM, but interpret QM in a wider sense, as relational QM. In this view, time-evolution dynamics is just a useful special case of relational dynamics. From this platform, the assumption-free manner to see dynamics, is to see it as correlated changes in two or more physical properties. One does not need to assume that these correlated changes can be parameterized by a global external parameter.

We find little reason to doubt that the principled thinker would side with the relational idea on this question. There are few reasons to insist that a time parameter always exists, and such a parameter is not required to properly define change and dynamics. But not all choices are made on principles alone. The choice of a more pragmatic person would depend on beliefs held about the degree of practical necessity for time parameterization, and the amount perceived difficulties of doing dynamics the relational way. The relational way is certainly more difficult, and some would suggest that sticking with the more pragmatic time-evolution paradigm might still capture the important parts of GR.

This is certainly an issue where opinions part ways. In light of the research principles we advocated in chapter 1, we clearly side with principles over pragmatism. Principles seem better equipped to provide progress in a situation with little empirical guidance, and pragmatism is perhaps more suitable in a situation with ample empirical input. In light of this we declare that relational thinking seems to be at the core of GR, and that it is therefore important to preserve this idea. The Newtonian or Lorentzian manner of constructing a rigid platform for the drama of physics to play itself out, seems to be more a matter of convenience than a sound fundamental assumption.

LQG operates in line with the ideas of relational dynamics. The precepts of being background free and diffeomorphism invariant are taken to be the central features of GR. The LQG construction tries to preserve these principles at every step. The LQG construction is background free in the sense of not using a background metric. The basic fields are defined classically in a fully background free manner, following the template set up by GR. While the Hamiltonian D+1 split does make the diffeomorphism invariance non-manifest, it still respects the diffeomorphism invariance by parameterizing the arbitrariness of the split.

While there is no metric background in the definition of LQG, it is still the case that a non-metric differentiable manifold background is used in the construction. The physical meaning (if any) of this background space is not clear. This is much the same as what is done is classical GR, where there is also a background manifold without any clear physical interpretation. In classical GR, this background manifold is always assumed to be "covered" by a metric field, and as such is more hidden than in the LQG picture. In LQG one can more easily imagine some sort of "unformed" space where no metric (or connection) is defined. This would represent "raw or pure nothingness". By extension, such an idea could also be imagined in classical GR. Beyond the rather empty words of "raw or pure nothingness" (pun partially intended) there is no empirical interpretation of this background space. At the mathematical level it provides a convenient stage for the theory. This is not necessarily detrimental. As we suggested in chapter 1, physical theories are mathematical theories with a partial set of semantic rules. It is not required that all mathematical objects have semantic content. Indeed, in most theories there are clearly many objects that do not have, and perhaps should not have, such semantic content. Still, we raise the issue in this specific case because after all we are evaluating a theory which details the creation a quantum analog of space, and which emphasizes the background free nature of this construction, but which still utilizes an intangible non-physical non-metric differentiable background to erect this construction.

In defense of the LQG approach, one can state the following: All the aspects of the spin networks relating to the embedding in the background manifold are "abstracted" away when implementing spatial diffeomorphism invariance. Only the combinatorial aspects of the graph interconnections and coloring are left as relevant entities. Still we maintain that, to make this "abstracted-away" property more manifest, one would wish for a completely non-geometric formulation of the theory, involving only combinatorics and algebra.

It is intrinsic to the background free view of GR that the most suitable formulation of its quantum theory ought to be non-perturbative. Any perturbative approach must at some point declare some state to be the state to perturb, and for a theory of geometry, this results in introducing a background geometry. Thus, any perturbative approach is in opposition to a background free approach. Regardless of this specific argument which is closely tied to the nature of gravitational interaction, we enlist another argument. Even in a world without gravitation, perturbation theory is not the gold standard of theories. A non-perturbative theory is just a word for a theory formulated without basing it directly on calculational methods that only apply in specific and limited situations. Perturbation theory limitations include, amongst other things, the proviso that certain expansion parameters are within restricted ranges. A perturbative theory is then obviously a theory, the formulation of which, depends crucially on the use of said specific calculational methods. No one would choose perturbative over non-perturbative in an otherwise equal setting. Non-perturbative

theories does not preclude using perturbative methods but they do not depend on them for their definition. QFT is not a perturbative theory, but it is a theory which uses perturbative methods as calculational tool in many (or most) instances. (Lattice gauge theory could be considered a non-perturbative calculational tool within QFT.)

As we said at the outset, LQG can be seen as both a specific GR quantization, and also as a general set of methods for background free quantization of connection based theories. In any quantization it is of course essential which theory one chooses to quantize. There are some classical theories that captures the low energy physics well, but that are far removed from what we later come to see as the fundamental degrees of freedom of the associated quantum theory. These theories are e.g. thermodynamics theories that are averaged over microscopic degrees of freedom. Quantizing fluid mechanics is perhaps not totally without merit, but does not lead to atomic physics.

The question of what the true fundamental degrees of freedom for gravitation is not settled. Because of the weakness of gravity at accessible energies, it present itself in a manner that might be far removed from its true high energy degrees of freedom. It is suggested by some that gravity is a thermodynamic manifestation of an entropic force, or some other macroscopic approximation of something which is microscopically very different. It is certainly unreasonable to assume that our GR based understanding of gravity is explicitly mirrored in the fundamental degrees of freedom corresponding to this. It is very likely for GR to be an effective theory at some level. Whether the fundamental degrees of freedom are so different as to render a quantization partially meaningless is unknown. It is important that the general quantization methods of LQG are not dependent on the exact Lagrangian of the classical theory. We believe the primary importance of LQG lies in demonstrating a framework for doing quantum physics in background free manner. It presents a framework for quantizing the very stage on which physics lives.

Furthermore, we would argue that GR is the natural first target to apply these quantization method to. We certainly can claim that our best current understanding of gravity (or geometry) is that it complies with GR at sufficiently low energies. And that that the understanding of GR we presented in chapter 4 shows that GR is a theory based on clear and quite natural principles. It would therefore seem strange *not* to try and base a quantum theory on this understanding. Indeed the only truly essential requirement of any quantum theory of gravity would be that it reproduces the findings of GR in the relevant domain. There might certainly be other ways of achieving this beyond a quantization of GR, but a quantization of GR must be considered a natural starting point.

From these very general considerations we continue with more concrete aspects of the theory. LQG aspires to be the theory of quantum gravity and not just a theory of quantum gravity, and this makes the number of unforced choices in the development of the theory important. If we accept starting from a geometric diffeomorphism invariant theory formulated in terms of connections, the kinematical setup of LQG seems to be fairly unique. The choice of the holonomy-flux algebra as the fundamental classical algebra derives quite directly from mathematical necessities. When defining a theory one aims for an algebra that is non-distributional and mathematically well defined. The Ashtekar variables themselves lead to distributional Poisson brackets (i.e. containing Dirac delta functions). The form of the holonomy and flux smearings that removes these distributional aspects, are dictated by the one-form and two-form nature of the Ashtekar variables. The representation of the abstract quantum version of this algebra is unique given the assumptions of the F-LOST theorem.

The quantization of the abstract quantum version of the holonomy-flux algebra does not correspond to the quantization as specified in the Stone-von Neumann theorem. The LQG representation map is not weakly continuous [25]. This does raise questions about the correctness of this type of quantization. We believe that current result show that, for system where traditional quantization is applicable, the LQG type quantization produce result that approximate traditional results to any desired accuracy [25][165].

Let us also discuss some problems that occur after the initial quantization. Because of its diffeomorphism invariance, the classical theory is a fully constrained theory. There seems to be no option but to use the Dirac approach of performing a separate quantizing the constraints after first doing an unconstrained quantization. This leads to the problem of dealing with the constraints in the quantum theory.

The Gauss constraint and the diffeomorphism constraint are related to the use of triads and the arbitrary 3+1 split of the Hamilton formalism. The Hamiltonian constraint is the constraint that captures the specific nature of the theory. And this brings us right to a crucial sticking point, the Hamiltonian constraint is difficult to solve.

Before we consider the Hamiltonian constraint we make a note of some results that are derived in LQG without solving this constraint. The basic picture of space in LQG is derived just using the Gauss constraint. The results on the quantization of area and volumes are results found by using $SU(2)$ invariant states. The definitions of a specific area and volume are not meaningful after diffeomorphisms are considered. Even so, it is believed that these results do carry over to the fully theory. One imagines that by using relative positioning in reference to matter fields, one could define such area and volume operators, and achieve the same result. It seems unfortunate that one has not been able to further substantiate this. We must consider that part of the very attractive picture that spin-networks present, is not guaranteed to be analogously represented among true physical states.

It is not completely clear to us that it is the relational nature of the dynamics, implicit in a background free theory, that leads to the complex nature of the Hamiltonian constraint, but it seems likely that this is the case. The end result of LQG is a Hilbert space defined by constraints, which means it is a quantum theory of initial conditions. The Hamiltonian constraint encodes the dynamics of the theory in this theory of initial conditions by restricting the Hilbert space. It seems natural that this must lead to a complicated constraint. In the minisuperspace approach of LQC one can actually see that while the Hamiltonian constraint just defines acceptable physical states of the gravitational field and the scalar field, the correlation between the properties of these states is encoded right into those physical states.

While it is certainly an exaggeration that LQG is permanently stuck on the Wheler-deWitt (WdW) equation, there are some problems. For all the brilliance behind Thiemann's untangling of the Hamiltonian constraint, it still raises issues. There are presently too many ambiguities of unclear physical significance in the implementation of the Hamiltonian constraint. This threatens to undermine the celebrated uniqueness of the previous steps of the LQG construction. There is no reason to think that this is detrimental, and its unreasonable to expect quick fixes for such a complicated problem. Still, it is important to find good arguments to substantiate that this does not undermine the careful construction of the other parts of LQG.

Because of the difficulties in defining dynamics by solving the Hamiltonian constraint, other approaches to dynamics are certainly welcome. The path integral approach, in the form of spin foams, is one such promising alternative [149]. Of course, the issues relating to time parameters does not disappear in the spin foam formalism, but it still might be easier to achieve progress on dynamics. Though the spin-foam theory is not derived directly from LQG, there seems to be growing consensus that the spin-foam formalism does represent the same theory as LQG. Group field theory (GFT) is a further such approach that also holds promise of dealing with dynamical issues [372].

One should realize that it might be wishing for too much to seek to solve the Hamiltonian constraint in the general case. There is no "general" solution of GR and this is in a sense the corresponding classical problem. (By a general solution we mean something like the general solution of the wave equation.) Simplification are therefore the order of the day, and much progress has been made on the Hamiltonian constraint in simpler systems.

In cosmology, black holes and other systems with high degrees of symmetry, there are many new results from LQG [373][374][375]. Here one is able to devise explicit physical states and extract the dynamics of the system. The main issue with these solutions is that their relation to the full theory is not clear. Symmetry approximation are usually made on the classical side, and then analogies of the full LQG quantization is used to construct a quantum system. This is not demonstrably equivalent to dealing with the same system within the full theory.

Case in point is applying loop methods to FRLW cosmologies. In the minisuperspace approximation serious deficiencies have been uncovered [185][186]. The constraint algebra of the full theory indicates that the metric undergoes signature changes at high densities. This has been repeatedly been demonstrated without being properly acknowledge in the relevant reviews. Only very lately have this fact entered into reviews. This doesn't undermine the whole theory of LQG, but it is an important caveat when considering results for other highly symmetric systems.

Recent result have brought the symmetry reduced systems in closer contact with the full theory, and one can anticipate that proper derivation can soon be made. This would be huge leap forward. It would be the first instance of something physically meaningful being calculated in the full theory.

The matter part of the theory is underdeveloped. All matter relevant for the SM are representable in LQG, but so far there are few constraints [145]. It is certainly a possible scenario that such constraint will emerge and that LQG can contribute to the particle physics aspect of theoretical physics. Before entering more into that arena we need to discuss Lorentz invariance and the classical limit.

Regarding (local) Lorentz invariance it is unclear exactly how it is affected by the discrete structure of LQG. At least naively it seem to be violated by its discrete nature and the absolute distances found in spin-networks. Some ideas like double-special relativity where one introduces a second set of absolutes, in addition to the maximal speed (of light), are sometimes considered [376][377]. Since matter is defined on edges or vertices, and general states will have superpositions, it seems inevitable that additional assumptions on the matter states are necessary to avoid Lorentz violations [378]. It has been shown that even extremely small Lorentz symmetry violations in the Planck regime cannot be easily hidden from low energy processes [379][2]. One idea is that non-local effects reminiscent of interactions found in string field theory leads to a cancellation of such effects rendering them invisible at low energies [378].

It is not yet clear that GR emerges as the macroscopic limit of LQG. This limits is not just about taking the Planck constant to zero, but to see that classical like states are present in certain larger spin-network states. Separate consideration might be given to both a large set of nodes as well as high spin representations. We will not discuss the details of this, but rather comment on its importance. First of all, it is essential that a correct macroscopic limit be demonstrated. The only absolute demand of LQG is that it must reproduce what we already know. LQG cannot be relevant unless this is demonstrated. Results within the spin-foam approach are getting closer to achieving this [146].

Assuming for the moment that the correct limit can actually be obtained, another very important aspect emerges. We have argue that the non-perturbative nature of LQG is an essential part of how it captures the core of GR. Nevertheless, the non-perturbative framework disconnects LQG from the rest of high-energy physics. It is necessary to evaluate whether it is possible for LQG to reconnect to the perturbative framework is some limit, as an attempt to properly validate the LQG approach. One can certainly claim that the non-renormalizability of perturbative gravity is an artifact of an inappropriate approach to quantization of gravity. And by several arguments this is probably true. It is still seems prudent to give a detailed explanation of how the non-perturbative nature of LQG results in a finite quantum theory. This explanation must at some level relate LQG to

[2] It is interesting to note that the ability of the Wilsonian coarse-graining procedure to decouple the IR from the UV is dependent on local Lorentz invariance.

the disappearance of the perturbative divergences. At lower energies, perturbative methods are often appropriate and applicable, and LQG should attempt to provide an explanation of how it (implicitly) deals with the non-renormalizable divergences. The question might have to reformulated to make sense within the LQG framework. Possibly, one might even decide that it is not a proper question, but somehow a satisfactory resolution (if not solution) should be provided.

Connecting with the perturbative framework would also provide a major boost to make progress on the matter contents and the Lorentz invariance. Before some relation between the pure discrete LQG picture and the lower energy smooth background perturbative framework has been established, most of the crucial questions regarding matter and Lorentz invariance will remain speculations.

One very interesting aspect of LQG is it relationship with ideas of holography and entanglement entropy. The most celebrated result in this respect is the derivation of the Bekenstein-Hawking black hole entropy formula [344][344]. A much more wide-ranging proposal is given in [345]. In this case one considers general surfaces in a semi-classical background and suggest that they obey the same area law. This is also discussed in [343]. The area law appears to be a signal of the presence of a semi-classical domain of the theory [345][346]. The area law criteria has been used in LQG to facilitate the search for a state approximating semi-classical spacetime [346].

LQG has no empirical confirmation so far, and while such confirmations are certainly possible it is not likely to be possible in the near future. We therefore try to summarize its achievements on a theoretical basis, independent of whether it actually is a (correct) theory of nature. By this measure we find that the theory of LQG is a substantial theoretical achievement in the sense that is tackles some very fundamental problems with regards to quantization, quantum theory, background free systems and dynamics without time parameters. It demonstrates the feasibility of making a principled analysis of the question of uniting gravity with quantum mechanics. Independent of the detailed conclusion the increased understanding of the listed issues that LQG provides, promotes a deeper understanding of quantum mechanics and time.

LQG for all its achievements have not yet ignited the interest of the greater part of the physics community. This could be due to the relative isolation of the central question of LQG from other parts of physics. LQG focus on pure quantum gravity based on conceptual principes might seem irrelevant to the more pragmatic physicist. While the pragmatists of the theoretical physics community have to face up to the fact that it is principles and not pragmatism that will lead the way to the highest reaches of the physical domain, LQG could do well to make its efforts be more relevant for physics in general. Matter and particles, colliders and cross-sections are more prominent in the minds of many than the questions of quantized area operators and the existence of time. It thus seems prudent that LQG seek to uncover more of its own relation to other parts of fundamental physics. Even though force unification is correctly not a necessary axiom for any fundamental theory, physics unification most definitely should be. Quantum gravity need not be unified with the other forces in the gauge sense, but it must be put on the same platform or vice versa. The hints of the applicability of LQG to gauge-gravity duality [144], and the hints of specific (possibly stringy) limitations on LQG matter interactions are important progress in this direction [378]. LQG should continue to promote it principled approach to important questions. The examples of Einstein, Weyl and von Neumann are historical examples of how far principled reasoning can lead. Equally important, one must disseminate the understanding that physics has permanently left the pragmatic domain where a continuous stream of new scattering data will lay out a clear path forward.

10.3 Evaluation of ST

In this section we present an evaluation of string theory. We again focus on the three topic that were most central in chapter 9. What are the conceptual foundations of string theory? How does

string theory explain the standard model? And how does string theory provide a theory of quantum gravity?

A difference between string theory and the other theories that we consider is that string theory is the subject of numerous books and TV-programs, and is partly a theory known to the general public. This creates a backdrop of opinions and views which to a certain extent affects every discussion about string theory's merits. Whole books have been written, just arguing for or against the basic soundness of the theory. Discussions are often polarized and phrased in a language which are not conducive to fruitful contemplation. We do not want to add to such non-productive polarized exchanges. However, some of the argument raised in such exchanges contain valid points, and therefore should and will be repeated in this evaluation. Let us therefore first of all declare our affection for the truth, and science as method for seeking the truth, and to declare our intention to avoid language and formulations that do not contribute to truth seeking.

Let us now consider the foundations of string theory. We find that we are unable to summarize ST in the same compact manner as NCG and LQG. Looking at dozens of lecture notes, and a similar number of textbooks this problem seem to affect other writers as well. We will take this at face value, and conclude that it is difficult to summarize string theory, and that there are some natural reasons for this.

Some of reason can be understood by a historical context (see sections 9.24 and 9.20). Briefly stated: String theory started as a perturbation series describing satisfying some properties. This was later interpreted as being a perturbation series for scattering of strings. Even later the presence of a massless spin-2 state in the spectrum was taken to suggest that perturbation series could be the basis of a theory of quantum gravity. Two-dimensional supergravity was added to the string world-sheet to enable fermions. This also also produces ten-dimensional local supersymmetry. Later on, various non-perturbative objects, mainly in the form of D-branes, was added to the basic description, and the low-coupling dynamics of these were described by extra open string states. Basically this is how string theory is defined.

Based on this history, string theory as a scientific theory is more of discovery than an invention. In hindsight, part of its nature could be described as a generalization of perturbative QFT in the direction of supergravity. However, even though string theory is strongly related to QFT, and uses many QFT techniques, string theory is not quite a QFT. The underlying principles behind the string generalization of QFT are not yet clear. It results in an infinite number of massive higher spin states that are responsible for improving the UV behavior of the theory to the extent of making it order by order finite. A more general description of string theory is available by the use of 2-dimensional CFTs. Certain set of CFTs describe the known perturbative string theories where we can use the picture of strings propagating in background spacetimes. Other CFTs presumably describe other "string theories" that do not have an interpretation in terms of weakly coupled strings propagating in some background spacetime. We find the most compelling version of what string theory is, to be the view presented in section 9.12. This is the view where string theory is seen as the two-dimensional generalization of the world-line formulation of QFT. We will switch back and forth between the CFT view and the string view as we see fit.

CFT have captured the interest of mathematicians. It is possible that CFT defines an extension of algebraic geometry [380]. A related perspective on CFTs are they constitute the representation theory of (super-)Virasoro algebras (and various related algebras) [242]. One way to relate CFTs to geometry could be through the ideas of noncommutative geometry. Specifically, CFTs seem to define spectral triples that in NCG define noncommutative spaces (see e.g. [259, 260]. This could mean that string theory in a sense could be a generalization of geometry. Why this generalization in certain cases looks like string propagating in space is an unanswered question. Perhaps this idea is no stranger than the idea that certain aspects of GR and standard model looks like a noncommutative space.

String history, although informative, is just history, and any formulation of any unconfirmed theory is bound to be evolving. Thus, so as not to hold what we perceive as conceptual weaknesses of the current formulation too much against ST, we could elect to see ST as a hint that there exists a theory we for ease of reference will call by the arbitrary name X-theory. By this hypothetical X-theory we will mean a theory containing (many or all) the insights of string theory, but formulated with a clear set of principles in a non-perturbative background free manner. While this positive view seems justifiable one also has to consider the actual state of string theory as a theory in itself. Even though the many dualities have brought hope that the known perturbative definitions are connected, and might be derivable from a single theory, string theory itself is still formulated in the same manner as before. It is a set of rules and guidelines for doing string theory with no clear definition of underlying principles. ST is not yet a theory in the traditional sense of implementing well defined principles. It is difficult to estimate how far from uncovering such an X-theory ST is. It has not been established that an underlying X-theory exists, but this is of course a difficult task without having a defined X-theory at hand. There are likely to be varying opinions on how persuasive the string theory hints are, but hopefully a fair assessment is that there are many hints that something like X-theory might exist. Note that we define M-theory as the theory giving rise to a low energy description in the form of 11-dimensional supergravity. Some use M-theory as a name for X-theory, but we stick to the definition just given.

It is quite often said that ST contains so many remarkable insights and connections that this must be taken to be strong evidence in favor of the theory. This is a natural argument to make, and we have even used something similar in this text to argue that NCG contains important insights (with respect to "predicting" many aspects of the SM). In essence of could put this as the "it fits too well not to be true" argument. We will cast some more light on this in the subsection on ST and the standard model, but we introduce some general considerations here. In a theory with strong empirical predictions this argument is certainly relevant, almost to the extent of being the definition of the scientific method (the theory that fits the best is the "true" one). However, in the case of ST the evaluation is much harder. What exactly is it that fits so well? We are not saying there is nothing to this, but it must be considered with some caution. Usually the category of "things that fit well" are things like gauge theory, chiral fermions, low-energy GR limit and so on[3]. Things we could label as very general features of some models in ST. These aspect are both relevant and important, but it is very hard to measure their evidentiary value. Certainly there are many other theories than gauge theories, even within the domain of QFT, and therefore the presence of gauge theories is important. And likewise with the presence of chiral fermions. However, there are many other theories than gauge theories with chiral fermions within ST. It is difficult to go beyond the statement that ST is compatible with the existence of these phenomena. This is not really a negative statement with regards to ST, it is more of a critique of certain statements given in support of ST.

ST as a theory of quantum gravity

Let us now turn to the question of unifying gravity with quantum theory. Is ST a theory of quantum gravity? There are basically three distinct finding based upon which one can argue that ST describes quantum gravity.

The first and simplest is the basic analysis of the closed string spectrum. All string theories must contain closed strings, and closed strings always have a massless symmetric (0,2) tensor in their spectrum. It is natural to identify this state with a spin-2 graviton.

The second step involves analyzing dynamics. We want to formulate a low energy effective theory that reproduces string scattering amplitudes in that limit. Calculating the tree level scat-

[3]Sometimes more specific aspects like fermion representations and generations are brought up. However, these phenomena occur deep within sub-models of ST and the "fits-too-well" arguments based on these are not equally strong. See also the later discussion of ST and the SM.

terings amplitudes of the graviton state by using vertex operator on a sphere, one find that these are reproduced by the usual perturbative tree level calculations based on the EH action or the supergravity action. Since they are tree level amplitudes they are in practice just the prediction of the classical theory. It is necessary that the EH-action appear in this manner since all consistent low energy interacting spin-2 QFTs need to be of this form. This is however not quite the same as reproducing GR. What string theory reproduces is the spin-2 version of gravity at tree level. This is a linearized version of GR where the metric is separated into a background metric and a perturbation of that background.

The third step involves analyzing background fields. Instead of letting the string propagate in flat 10-dimensional Minkowski space one can add the massless bosonic fields to the string action, and experiment with various background settings. These background fields are in a sense fixed, but they must still obey consistency requirements. Or rather, we assume that they obey consistency requirements. These requirements lead to the conclusion that Weyl invariance requires a vanishing beta function. At the one-loop level this requires that the background spacetime obeys the classical Einstein equations. At higher loop order this produces corrections to the Einstein equations.

These three steps are certainly enough to make ST interesting from a quantum gravity perspective, but at the same time they are not enough to claim that ST is complete and finite theory of quantum gravity. The reason is that ST is no more progressed towards an understanding of quantum gravity than any theory of quantum gravity defined as a finite perturbative series around a flat Minkowski vacuum. For most practical purposes string theory is just a small correction to this perturbative expansions. It is of course important that the string theory expansion is likely to be free of divergences but it doesn't immediately change the picture it provides. The perturbative nature of the formulation makes it unclear whether the features of a background free geometric understanding of gravity is fully captured. Quantum gravity is expected to obliterate any concept of space at the smallest scale, and to make space an emergent notion that appears in certain macroscopic situations. It seems illogical to quantize geometric perturbation on a background known to be headed for destruction. One cannot go so far as to say the perturbative approach precludes any definition of quantum gravity, but it appears to confine it to less interesting domains of quantum gravity. For perturbation to hold background spacetime must hold, and perturbations must be small. This is the exact opposite of what is expected to happen at the Planck scale, where probably no coherent idea of space exists.

We move on from the strictly perturbative domain to consider further insight on quantum gravity provided by ST. The gauge-gravity dualities, and the idea of entanglement entropy, are two very interesting ideas that can be used to extract more of the potential quantum gravity information hidden in ST. In gauge-gravity duality it is postulated that type II ST on asymptotic AdS_5 space is equivalently described by a CFT on the four-dimensional boundary of AdS_5. This seems to give a definition of string theory that does not require perturbation theory to be valid. There is no requirement for a background metric, only the asymptotic form of the metric is specified. Unfortunately the CFT is required to be at strong coupling which limits the calculational value of the CFT. In principle this is a much stronger contribution towards a theory of quantum gravity than ST as a perturbative theory can deliver. It is a bit awkward in the sense of using theories on boundaries to describe to describe the theory of interest. One might naively ask where this boundary should be located in physical space. However, any quantum gravity description that implements the core insights we have about gravity, and that works, should be fine. Even though this is certainly progress, the asymptotic AdS demand is actually not a small limitation. First, the actual asymptotic background is not believed to correspond to the actual universe we live, and second, any such asymptotic requirement is a limitation. There are currently no solution for how to implement this in de Sitter space (or similar spaces). For all its potential, the actual quantum geometric picture to extract from this is not yet clear. What quantum geometric conclusions and construction can one

make with the AdS/CFT setup?

An interesting deeper understanding of the gauge-gravity duality comes from the idea that entanglement entropy, in which a the entanglement structure of CFT (with a gravity-dual) can capture much of the geometric information on the quantum gravity side. As a first approximations this has mostly been worked out for classical gravity. For classical gravity the entanglement entropy seems to dictate area laws and connectivity properties in the gravity bulk. Advanced correspondences have been found, even to the extent of the basic information theoretic laws of entanglement on the boundary leads to the requirement that Einsteins equations are true in the bulk. This certainly holds a promise of being a road to quantum geometry, or road to the emergence of geometry from quantum theory.

We want to mention one issues that one would expect in a theory of quantum gravity that are not yet prominent in ST research. This is the problem of defining dynamics. While it is in a sense natural that this is not prominent in ST since is defined on a fixed background, it is still assumed in many discussion of ST that the background is actually dynamic, but there is seldom any mention of problems of defining time evolution without a fixed background. We cannot seem to find any papers discussing the problem of time in the context of ST. This is also not discussed in the context of AdS/CFT which is often labeled as a non-perturbative background independent definition of quantum gravity. Although this label is somewhat premature it should suggest that discussing the problem of time should find a place in ST research.

ST as an explanation of the standard model

We now move on to gauge theories and the SM. For a long time it was assumed that gauge theories must be present in the ten-dimensional theory for ST to have any hope of producing the SM with chiral fermions. The Calabi-Yau based compactification of the heterotic $E_8 \times E_8$ was seen as the only promising avenue towards producing phenomenologically interesting models. Later it was realized that by defining D-brane worlds made from intersecting D-branes presented a different opportunity for deriving gauge theories with chiral fermions. The available tools for producing interesting models are much richer than ever before. It might seem strange that in addition to the choice of which perturbative theory to start with one can also choose between brane-worlds and compactified world, and add or remove D-branes at will.

The question of compactifications bring us to another issue, the issue of moduli. Moduli are parameters of the compactification manifold, and they correspond to scalar fields in the compactified theory. Several techniques have been developed to create potentials for these scalar fields so that they are no longer massless and have localized minima. This is certainly progress towards making the understanding of compactifications more complete. Given a scalar potential their vev values can in principle be determined, and couplings, mixings and masses resulting from this can be calculated. However, the introduction of such flux compactifications has lead to a worsening of a different problem. Already, early on in the study of compactifications it became clear that they were quite numerous. Early string textbooks express hopes of sifting through the possibly thousands of possibilities. A number that soon rose to millions. With the advent of flux compactification the number landed in the region of 10^{500} possibilities and beyond. This launched string theory into the whole landscape debate.

The string landscape is a pictorial references to peaks and valleys of flux induced potentials for moduli. Each valley is a possible stable perturbative vacuum. The various valleys produce different gauge groups, couplings, generations, masses and so on. Having a landscape with several valleys would be no shock, but having a landscape with 10^{500} changed the perspective on ST. In the eyes of many ST could no longer be expected to give definite predictions for low-energy physics. Perhaps ST could exclude some combinations, and suggest that some combinations are more common and so on, but that would be as far as it goes. For some this was a closing of a door and an opening

of another window. One could see the ST landscape as opening up for a more environmental way of thinking about physical laws. Specifically, such possibilities occurred when combining the landscape with eternal inflation and the multiverse idea. The idea was that inflation produces ever new universe continuously, and that in each of these ST lands in a different valley of the landscape. This produces a different set of physical laws in each universe.

The environmental take on the landscape idea puts a serious dent in the methodological framework we suggested in chapter 1. It would no longer be reasonable to expect ST to reproduce what we already know about physics. This makes the proper handling of the landscape a critical cross-roads for physics. Can we ever substantiate a theory that gives 10^{500} predictions for low-energy physics, and which potentially might be unable to even determine that the SM is one of possible ones?

String theory is more closely infused into the day-to-day particle physics research than the other approaches we are discussing. This is certainly positive and attest to the presence of many useful and powerful ideas with string theory. There are however some aspects of particle physics research that we want to address in this respect. Theses aspects are not strictly within string theory, and therefore not directly attributable to string theory, but at the same time these aspects are to some extent derived from string theory ideas, and are certainly influenced by string theory, and have also influenced string theory. So while string theory cannot be held accountable for all the things string theory ideas are used for, it is nevertheless relevant for our discussion.

There are two ideas in particle physics that have figured prominently on the agenda for the past forty years. These ideas are the idea of Higgs mass naturalness through low-energy global supersymmetry, and the idea of gauge unification through grand unified theories.

The GUT idea evolved naturally from the electroweak success with gauge "unification". It started first with semi-grand unified Pati-Salam model and the proper GUT of Georgi and Salam. All three known quantum forces were seen as originating from a single simple gauge group (with a single coupling constant). It expanded the Higgs spontaneous symmetry breaking idea and because of the simple gauge group quantization of electric charge was predicted.

The Higgs naturalness problem was perhaps first identified in 1976 by Gildener and Weinberg in the context of grand unified theories [381]. In 1978 Susskind presented technicolor as a possible solution [382], and by 1981 a supersymmetry solution was presented by Witten [383]. For the next thirty-five years the idea of finding superpartners at the electroweak scale came to dominate any discussion of beyond standard model physics (see e.g. [384]). With the exception of experimentally detecting Higgs this is the basic principle underlying the LHC design and research program. Arguments for weak-scale susy by solving the Higgs mass naturalness was soon supplanted by arguments of susy gauge unification, susy dark matter candidates. While superstrings is silent both about global supersymmetry and especially low-energy global supersymmetry, this neutral stance was not emphasized, since the leading paradigm from 1985 and on-wards was Heterotic string compactified on a Calabi-Yau, which leads to E_6 GUT with global $\mathcal{N} = 1$ supersymmetry (or something similar) [237].

All of these ideas fused into a nice paradigm that had additional attractive properties. The fundamental local supersymmetry of the string made possible transporting some important calculations from the weakly coupled domain to the strongly couple domain. The presence of low-energy supersymmetry stabilized the string compactifications and vastly simplified calculations. Both the weak scale sparticles and the WIMP[4] dark-matter particle would be detectable by the next generation detector. Such, "we will see it the next collider" reasoning stimulates perhaps the primal hunting instincts of many physicists and the ideas were soon universally popular. [385] is an example of a paper from 1999 that illustrates some of the expectations with regards to supersymmetry and string theory. In all the enthusiasm some aspect were soon relegated to the back pages. The arguments for

[4] weakly interacting massive partice

naturalness were not all that strong. It was certainly based on real concerns, and naturalness was an attractive feature, but there was no clear precedent that nature implemented naturalness at every level. There were several counter-examples and perhaps only one previous successful prediction based on this idea. Similarly, gauge unification and GUT was an appealing feature but not well motivated.

By the end of 2016 it is experimentally clear that if any of the ideas deriving from Higgs mass naturalness concerns are even remotely correct, they are certainly not correct in any way close to their original formulations. And regardless of the empirical correctness of consequences deduced from these ideas, we find that the faith bestowed upon these arguments was not based on a proper critical evaluation.

Let us now turn to string theory proper and briefly comment on how this historical period may have affected string theory. It is not unlikely that the level of faith in these ideas may have led string theory in a direction more heavily emphasizing fundamental supersymmetry, low-energy supersymmetry and gauge unification. Why should one consider models without supersymmetry when the Calabi-Yau idea is so convenient and agrees so well with the most prominent ideas of experimental particle physics? There are upsides and downsides to this. On the upside, string theory is not committed to low-energy supersymmetry or gauge unification, and perhaps even the idea of high-energy supersymmetry can be re-evaluated. On the downside, most of the work on string phenomenology are within this framework and would have to be reworked if it is to accommodate such a shift in position. In addition it displays the challenging nature of basing the setup of the low-energy paradigms of an unconfirmed theory on other unconfirmed ideas about how the standard model is extended at higher energies.

ST as a whole

In summary, ST is a potent framework for generating ideas regarding both quantum gravity and the standard model. As a theory in itself its current formulation is lacking in some respects. It is not disputed by anyone that its foundation is not clear. Most would also agree that it does not yet provide a theory of quantum geometry. The ideas of AdS/CFT and especially entanglement entropy seem to be a promising path towards further progress in this direction. As for explaining the SM ST is very powerful but at the same time so flexible as to render the explanatory force of the potential explanation in doubt. The landscape model must be taken seriously, but even though it could be correct, it is nonetheless fair to claim that at this stage it weakens the theory unless some decisive prediction can still be made. In light of our suggested methodological approach, devising a new theory based on string theory but with a clearer foundational basis could be one approach. The ideas of AdS/CFT and entanglement entropy might be closer unraveling the true foundations of the theory. Can one find a stronger arguments in favor of supersymmetry or should supersymmetry be dropped as fundamental axiom? Could supersymmetry perhaps be related to holography? Can holography itself produce hints of a world of extended objects such as strings and branes? The ideas of holography and entanglement should be investigated with an eye to establishing a stronger foundation for ST or creating what we have called X-theory.

10.4 Concluding remarks

Having reached the final section of this thesis we need to take stock. We have found that NCG, LQG and ST have all given important contributions to the questions of unifying gravity with quantum theory and explaining the standard model. They are all fascinating theories, full of ideas of immense depth and complexity. Still we must conclude that none of them have yet been able to fully answer these questions.

NCG has a clean and rigorous mathematical setup which should inspire other theories to

embrace its powerful mathematical machinery. NCG comes pretty close to explaining (or at least predicting) several features of the SM, but its vindication as a fundamental theory depends on experimentally discovering the Pati-Salam model. NCG aspires to contribute to gravity theory but the understanding of quantum gravity through the use of NCG is still in its early days.

LQG is a logical and straightforward attempt at combining gravity and quantum theory based on strong and clear principles. LQG provides an appealing and comprehensive quantum geometric picture and enlightens many foundational issues, but its contribution to the understanding of the SM is meager. Because of its pure focus on quantum gravity (often without including any matter components) it is somewhat isolated from other approaches. It lacks an effective handle on dynamics, and the classical limit is still elusive. The idea of entanglement entropy seems to provide an important insight into the structure of LQG as well as helping towards uncovering its classical limit.

ST is a powerful and flexible theory that connects to many other areas of physics. ST provides a rich platform from which to extract a deeper set of principles. Its lack of clear conceptual principles, and its reliance on perturbation theory and background dependent methods, as well as its overabundance of options, are things that should be improved for it to reach its full potential as a fundamental theory. Its quantum gravity approach is mostly limited to perturbation theory (with some exceptions like matrix theory) and does not yet provide a clear picture of quantum geometry. The quantum gravity part of ST is substantially improved and extended by gauge-gravity duality and entanglement entropy ideas. ST in combination with eternal inflation provides an environmental explanation for the parameters of the standard model. ST have made progress in the direction of understanding the SM as well as partial but important progress towards quantum gravity.

It seems that these three theories, although they all are powerful and ambitious, are unable to reach the complete answers to our designated questions on their own. In this final section we want to suggest some potential benefits from considering these theories together.

First and foremost we want to suggest that the pointless segregation of research on fundamental physics into sub-approaches with no communication between them is something that should be brought to an end. An essential source for new ideas and progress is interfacing these three approaches with one another. A few examples of such interfacing already exist.

Putting NCG within the context of another physical theory, either in LQG or ST, could bring new tools to the theory. One example of this is how NCG spectral triples have been applied to CFTs to try and extract the geometrical spaces that they define [259, 260]. A more detailed analysis of matter in LQG has uncovered necessary conditions to avoid (local) Lorentz invariance violations in LQG. This has again led to suggestion of introducing a form of non-local matter interaction that is similar to those found in string field theory [378]. NCG has been partly integrated with LQG by applying NCG to a noncommutative version of the holonomy-flux algebra leading to a brand new theory of quantum holonomies [370, 371]. Non-perturbative quantization in ST could help alleviate the Achilles heel of background dependence. One example of this would be to apply LQG to the supergravity in the AdS/CFT conjecture [144]. This would be a solid breakthrough towards connecting LQG and ST. This could also facilitating adapting perturbative techniques for use in LQG.

Another aspect of using LQG in AdS/CFT is that it connects the LQG and ST through the ideas of holography and entanglement entropy. Holographic ideas have been tried out and extended to LQG [386][387], and it has also been used to help in the search for semi-classical states [345][346]. LQG and ST seem to embody similar structure of entanglement entropy and this area is a promising candidate for connecting the two theories.

It is possible that even deeper connections can be made. It was beyond the scope of this thesis to enter into foundational issues of quantum mechanics beyond what we presented in chapter 2 (and

those ideas directly involved in the development of LQG). We just note that entanglement entropy ideas, as used in both LQG and ST, come from foundational research in quantum mechanics. Perhaps lines can be drawn between quantum foundational ideas, the relational definitions of quantum mechanics, holographic ideas and the emergence of spacetime (see the section "Postscript - Ideas for future research" for a more radical suggestion in this direction).

While it certainly may be the case that none of these three theories are getting anything right we have a strong hope that they are all getting something partially right. This means that if they are right about some of the same things one must be able find common ground among them. This could provide an invaluable insight into what parts of these theories are on the right track[5]. By stimulating an increase in NCG-LQG-ST cross collaborations, and adhering to the principles of seeking an ever better and more satisfactory foundational understanding, we are sure that both unifying gravity with quantum theory and explaining the standard model are problems that will yield to the force of many dedicated thinkers applying their minds in unison.

[5] Of course it is possible that they are wrong about the same things in the same manner which would make this a clue that would send us in the wrong direction, but we consider this to be an unlikely scenario.

VII

Appendices

Appendices

Postscript - Ideas for future research

In this extra section we want to present some ideas for future research. These ideas are slightly tangential to other parts of the thesis. To avoid confusion we place them as a postscriptum to the rest of the thesis. We will present our suggestions in the form of a template theory. The template is meant as a (slightly provocative) suggestion for which directions it might be most productive to conduct research. Note that while we write in a constructive mode, and this specific "construction" is new, we are not implying ownership of the individual steps. This template theory assembles and reshuffles ideas from many current research papers, and we will cite the most important ones. In this section we leave all restraints behind and let speculation and grandiose thinking rein freely. And in this respect we ask the reader for some latitude and indulgence on his or her part.

In this template theory we attempt to connect together many of the stands that we have woven in previous chapters. And while this template may initially seem far from the theories we have been reviewing there are some strong connections to what we have presented. The template is basically a generalization of structures of entanglement entropy and holography in ST and LQG that we presented in chapter 9. In addition, and beyond what we have touched upon in the thesis, it extends the reach of entanglement concepts to suggest that matter and gauge particles are emergent concepts. While on the surface of it emergent matter embodies a very different set of ideas, it is very natural to include such ideas since they are so closely related to the entanglement and quantum foundational based ideas presented for emergent space. Indeed when applying the entanglement and quantum foundational approach to space, applying a completely different approach to matter and gauge particles would seem unnatural. With regards to the emergence of "time" (or dynamics) we follow the approach suggested in section 8.8. The template will include some ideas from quantum foundations research that due to space limitations could not be included in the present text. We will also use this template to briefly exemplify how a focus on conceptual ideas and satisfactoriness could provide guidance and new ideas in physics research.

Building a foundation

To enable us to build our template theory we need a foundation. What should be present in the foundation of a theory of physics? As we suggested in chapter 1 we desire a foundation that is

transcendent and ethereal. Transcendent in the sense of it being both a foundation for physics but also a foundation for itself, and ethereal in the sense of building on the most universal and timeless notions we can think off. We want a foundation that in the least way possible induces further question about what are the foundations of this foundation. While we forego the demand that it be absolutely self-explanatory, we at least want to suggest components that has potential for some level of such self-explanatory capacity. We believe such founding elements exist in the form of logic and set theory, and in the theories of probability and information.

Let us quickly add these components one by one, with some brief comments on how these items could have some self-explanatory capacity. If we want to use the concepts of true and false, and we want to express true-and-falseness relations between various statements, we are immediately in the domain of logic. At the same time logic seems to be completely neutral to what subject we want to express. It is a foundational element in all knowledge gathering activity, and in particular in those using formal systems and methods. If we accept logic as a foundational element we can quickly extend its reach. We decide from further studies of logic, or from experience, or just by a reasoned choice, that the concepts that we find most useful can be expressed as finite or infinite collections, as relations, or as functions. These are all concepts that are most easily defined using the idea of sets. Formalizing our ideas about such collections adds axiomatic set theory to our list of foundational tools. To this list of tools we want to add the concept of probability, and we also want to add the concept of information. These last additions can in some sense be seen as extensions of logic, with extra expressiveness that enables us to quantify uncertainty. These additions can also be seen as basic assumptions about knowledge. We are thus so far armed with the tools of logic, set theory, probability theory and information theory.

The next step is to deduce the basic axioms of quantum systems from conceptual postulates formulated in terms of probability and information theory. There exist several reconstructions of quantum theory from conceptual principles [388][389][390][391][392][393][394, Ch9][395][396]. They are not completely satisfactory in that the sets of postulates they make are still more technical than purely conceptual, but we accept these reconstruction as models for what we would like to achieve: a natural deduction of the quantum postulates given just conceptual postulates formulated using the languages of information theory and probability. If one lacks faith in such reconstructions, one can alternatively use the quantum probability based axioms that we presented in chapter 2. From any such reconstruction we can deduce the traditional Hilbert space formulation of quantum theory. Note that at this point we are not including any notion of dynamics.

Now the basic palette that we need is set. Our basic tools consist of four abstract fundamental theories, in addition to a formulation of quantum theory. It is important to emphasize what we are *not* assuming at this time. We are not assuming any form of space or time or matter, and we have yet to consider dynamics. What we do have are Hilbert spaces, quantum states, tensor products, and so on.

Emergent space

Since the algebra of quantum theory is noncommutative it automatically brings with it entanglement (see chapter 2). Using entanglement we can do many things. Given some arbitrary decomposition of the Hilbert space (intuitively representing subsystems), and a pure entangled state for the complete system, we can always derive mixed states by focusing on some part of the Hilbert space decomposition (see section 9.23). This brings about many possibilities which we will now explore. In the following we just picture a general Hilbert space of states. We don't necessarily envision this system as representing any concrete physical system.

Imagine being given a decomposition of the Hilbert space. We then express all the subsystems of the decomposition as nodes of a complete graph (i.e. fully connected). Next we label each edge of the graph by the mutual information between the two subsystem (more precisely it is a function of

the mutual information). The hypothesis is now that the coefficients on the edges can be interpreted as metric information, and the graph therefore defines a discrete metric space. Furthermore we posit that this metric space can be used to define an approximating manifold in which we can embed the graph isometrically. Thus we have constructed a concept of approximate continuous space from pure quantum theory, just using the idea of entanglement (or mutual information) of subsystems (see [351][352][353] [354] for similar ideas).

There are of course several details to discuss here, even at this speculative template level. The entanglement properties, when expressed as weights on the edges of a graph (via mutual information), may not always define a metric, and even if it does there might not exist any embedding in a manifold. Not all states will correspond to a semi-classical geometry. There is nothing wrong with this, we expect quantum gravity to contain states which are not close to classical geometry. If we want to emphasize the emergence of semi-classical states we can phrase this as some sort of phase transitions (corresponding to varying the states of the Hilbert space), and let us posit the existence of various phases (sets of states), some of which allows a correspondence to be made with an approximating manifold space. Thus, geometric space arise direct from quantum entanglement structure in one or more specific phases (subset of states) of such an entanglement structure. At the intuitive level one could imagine a "condensation" of the graph resulting in the formation of a classical space structure (see [397] for something similar).

Emergent matter and dynamics

We now move one to consider matter and dynamics. We want to consider matter and dynamics as emergent aspects of the picture we already presented. While the idea of emergent matter is relatively undeveloped in the context of LQG and ST we regard is a a natural extension of the emergen space pircture. Let us consider matter first. Various geometric defects or kinks (or any semi-local or collective states) of the graph we just built could represent possible propagating degrees of freedom, which could be identified with fermions or gauge bosons (see [398][399][400] for similar ideas). These are not really particles in space, as there is no space, and there are no particles, but the entanglement structure produces something that looks like space and which has semi-localized "defects" that can propagate and interact. Propagation involves change, or dynamics, which we have not yet introduced. We imagine that the relational ideas presented in chapter 8 will enable us to define a sense of parameterized change, that in special cases can be seen as states evolving in time. Where does change come from? We suggest that it comes from the intrinsic correlations between potential observable properties that exists in the original quantum system.

We have not discussed what the subject matter is of the core quantum theory that we are building the whole template on. This must be a fairly abstract system, not relying on concepts of space for its definition. The details of this system could be important, or it could be the case that all that is needed is some form of general quantum system. We do not have any more specific speculation to offer in this direction at present.

Relation to existing theories

In this subsection we discuss how the three theories we have reviewed fit into this template theory. We will not attempt to give anything close to a complete picture of how this might work out, but we are instead content to suggest some ideas. We will mostly focus on ST and LQG.

We start with string theory where we have seen through the AdS/CFT conjecture and in all the results related to entanglement entropy, that a key part of string theory is that it reproduces, by the Ryu-Takayanagi conjecture, the entanglement structure of a conformal field theory. In section 9.12 we attempted to find a fundamental definition of perturbative string theory, and we saw that the main perturbative expansion of string theory is defined by using correlators of a conformal field theory. Furthermore, research on entanglement entropy suggests that the appearance of a

semi-classical spacetime is conditioned upon the validity of the area law on entanglement entropy [345]. These things taken together might suggest that a key element of string theory is how it captures the entanglement structure of a CFT. CFT states can be considered to be closely related to the states of the abstract Hilbert space of our template. Maybe it is the deep connection between string theory and CFTs that leads to the fact that string theory captures some of the essential entanglement structure relation. By this interpretation we consider the CFTs of AdS/CFT to be a sort of intermediate stage between the abstract quantum theory and the geometry it gives rise to.

By some of same logic we can fit LQG into the picture. If the template is close to the truth one of the essential features of LQG must be that it captures some of the correct entanglement structure. The spin networks with their edges piercing some surface seems to be very conducive to formulations in terms of regional entanglement entropy (see e.g. [346]). One promising avenue to explore further is the possible use of LQG in AdS/CFT. If one can get LQG quantization of supergravity to work in the AdS/CFT setup this will be a major breakthrough in terms of entanglement entropy structure as well relating string theory and LQG [144].

The role of NCG in this picture is much less clear. We can speculate that NCG might provide a very generic way of describing non semi-classical spaces with the correct entanglement structure.

Constructive use of satisfactoriness

We will now present some arguments of a slightly different nature, that relates to this template. We suggested in the preface and in chapter 1 that the "satisfactoriness" of a theory might be a useful tool in research. Even though this idea has the sound of something very subjective and anthropocentric, we believe it has merit. Isn't physics a subject driven by a desire to find a correct *and* satisfactory description/explanation of nature, and not just a correct one? We believe it is, and that we might as well turn this quest on its head and ask: what could a satisfactory explanation look like? We cannot claim that it must look like the template we just presented, but we can claim that it must possess the same properties as that template. It ought to be derived from a topic-neutral semi-self-explanatory foundation, and with a minimum of extra assumptions it must produce the common concepts we use to define physics.

A key feature of the template is that it leads to a conceptual reduction. By this we mean that there are far fewer fundamental concepts in this template than in current theories of physics. The concepts of space, matter, dynamics, time are all emergent. There is only abstract quantum theory and nothing else. While the template is purely speculative, it illustrates the appeal of the dual phenomena of emergence and conceptual reduction.

Where does the idea of "satisfactoriness" that we are suggesting come from? We find our "satisfactoriness" arguments to be no different from those used all the time in physics. Physicists argue for the unnaturalness of the Higgs mass, for the strangeness of the initial low entropy of the universe and so on. These are all arguments based on a concept of satisfactoriness. One reason that we have criticized some of these uses of satisfactoriness, has been that they have been used to deduce unique solutions based on satisfactoriness, while in fact satisfactoriness does not lead to unique solutions. It does instead lead to classes of proposals, and it excludes many proposals. Another reason for our critique is that satisfactoriness should be applied to the complete model of physics and not to sub-problems. It is very easy to propose a "more satisfactory" solution of a sub-problem by just transporting the unsatisfactory parts to another sector of the complete theory. With these caveats in mind, can we have any faith that nature confirms to a properly applied idea of satisfactoriness? Probably not, but physicist will not stop until a satisfactory explanation has been found, or all such satisfactory explanations have been excluded. We do not imply that nature is forced to comply to human conceptions of neatness, but it still makes sense to look for satisfactory explanation first, and by doing this the concepts of satisfactoriness can serve an extra purpose as an idea generating tool.

Role of quantum foundations

In our template theory we suggest deriving basic physical entities from foundational elements of quantum theory. We want to add some comments on how this could impact research.

In physics there is no consensus on the so-called interpretation of quantum theory. The Copenhagen interpretation, consistent histories interpretation, many-worlds interpretation etc. are all popular amongst some physicist. Most physicist consider the question of quantum interpretation to be more of personal preference without any real substantive content to it. In the same vein quantum foundation research is probably considered by most to be an odd side-line of physics research, with no relevance for anything actual and physical. These viewpoints are often supported by statements of the sort: "Shut up and calculate", "quantum theory cannot be modified", and "quantum theory has been tested without a fault for 80 years".

We want to pose an opposing viewpoint. First of all, in light of our template, considerations of the fundamental foundational axioms of quantum theory, which forms the basis of the whole template idea, seem to be very important. In this template, any question about physics more or less boils down to an understanding of quantum foundational principles. One can easily imagine that small changes to this foundational basis could produce important consequences. Perhaps it would no longer be possible to derive any idea of space if this conceptual basis is changed. Or maybe things like a preferred number of spatial dimensions depends initimately on the details of the foundation. Such issues supports the idea that a closer study of the basis of quantum theory is important. Secondly, while quantum theory is well tested in some domains, there are many domains and aspects that are untested. This applies in particular to situations involving strong gravitational fields. Thirdly, there are important unresolved issues within quantum theory, like the measurement problem and the emergence of a classical world, that cannot just be filed away as unimportant philosophical niceties. The various proposed solutions to these problems have potentially sizable and significant impacts on actual physics questions. We will give just one example. The question of black hole information loss have been a stimulating question for a lot of quantum gravity research. However, unfortunately, the impact of making changing to the foundations of quantum theory is usually not considered (see e.g. [401][402][403]).

Summary of the template

Summing up, we have seen evidence in the thesis that ideas of quantum gravity are closely related to entanglement entropy, and that there could be a sense in which geometry emerges from such fundamental quantum concepts. We suggest a template theory that goes all out in this direction, and derives several key physical structures from entanglement related aspects of quantum theory. In this template, space, time and matter are all emergent concepts. These ideas of emergence interface nicely with current research regarding entanglement entropy within ST and LQG, especially in a boundary/bulk setting. We further argue that the ideas of this template, as well as several other lines of inquiry, suggest that research on quantum foundations is likely to play a more central and constructive part of theoretical physics in the future.

The following appendices briefly but systematically present several mathematical topics that are required for a proper understanding of mathematical side of the thesis. Depending on background and preferences these appendices can be considered as extra non-essential appendices or as an integral part of the thesis. It is a systematic presentation but we focus mostly on key ideas, definitions and results. Proofs are mostly omitted. Not all definitions are results are directly used in the text (in the final version) but almost all of it is implicitly used. Mathematical structures are hierarchical and many definitions are required just to be used in other definitions. As there is quite a bit of mathematical material here we felt it would be disruptive to integrate all of it in the general text. We have prioritized staying on a focused narrative path in the main text. We have therefore avoided getting to mathematically involved in the main text. Still it is essential that these

definitions and results be available for the reader who wishes to read the text with greater attention to mathematical detail. A few definitions or results are duplicated in the main text. These are typically chosen on the basis of being both essential to the presentation and being less well known in the general physics community. We deviate from this rule (and place it in the appendix) when the material is too extensive to be presented briefly.

All the material is basically text-book material. References are given at the end of each appendix. We do not give in-text references for individual results.

Appendix A concerns some key foundational aspects of axiomatic set theory and logic which is used and alluded to in chapter 1. In appendix B some very basic aspects of category theory is presented. The framework of category theory is explicitly or implicitly used throughout the thesis. In appendix C a dense review of some central notions in algebra is given. Lattices and boolean algebras are used for the review of quantum physics. Clifford algebras are central in the foundation of QFT and NCG. Graded algebras are necessary to define supersymmetry. Appendix D on measure theory is used in chapter 2 on quantum theory also other chapters. Appendix E and F on topology and differential geometry is used throughout. Appendix G contains various useful information on Lie groups and Lie algebras. Appendix H, I and J presents various mathematical topics which are central to the understanding of NCG.

A. Logic and sets

We present a few facts surrounding formal systems, logic and the Zermelo-Fraenkel (ZF) axiomatic set theory. These structures serve as both the theoretical and also practical foundation of mathematics and mathematical physics. Specifically we are basing the whole setup of physics, as presented in chapter 1, on the idea of a formal system. We shall not directly challenge or deeply discuss this basis. We find it important to understand that while using ZF with standard first order predicate logic we are still able to treat quantum systems with "other" logics. As an example, we are able to discuss topos theory within ordinary mathematics, even though its internal logic is intuitionistic. There are also questions regarding infinity and "the continuum" which are rather explicitly answered (stipulated) by the axioms. We also find it important to substantiate the idea of math as a formal system, and mathematical practice as something that in principle is based on formal proof. We will do this by showing that axiomatic set theory is not that far removed from the ordinary mathematical objects that we all use. We briefly show how ordinary mathematics is built from axiomatic set theory. We indicate how the theory is able to build the real and complex numbers from the axioms. Most math books will build all other objects from scratch. Note that the intention of this chapter is not to actually use formal methods, but to indicate that using them represents a correct description of underlying structure of mathematical practice. Actual mathematics is still done by way of using informal methods. These informal methods are usually sufficient to indicate how a formal representation may be produced. For the most complex proofs computer based formal methods have already taken over for informal proofs.

A.1 Formal systems

A formal system is general purpose tool to represent reasoning, grammar, algorithms and more. Our only usage will be for representing logic and mathematics.

Definition A.1.1 — Formal system. A formal system F is a tuple $F = (A, S, L, NL, RD)$ where the elements are as follows.

1. An alphabet A (see definition below).

2. A grammar (or Language) S in the form of the set of wffs $S \subset A^\star$ ($A^\star$ = all possible strings made of letters from A).
3. A set of logical axioms $L \subset S$.
4. A set of non-logical axioms $NL \subset S$.
5. A set of rules of deduction RD as a subset of S^n.

The alphabet of a formal system can be split into several distinct sets.

Definition A.1.2 — Alphabet. The alphabet of a formal system is a union of the following sets.

1. A set of variables $A = (x_1, x_2, ...)$
2. A set of logical connectives (like $\wedge, \vee$ etc.), written as $\Omega = f_1^{j_1}, f_2^{j_2} \cdots$ where j is the arrity.
3. A set of punctuation symbols $\{ \quad (\quad) \quad , \quad \}$ etc.
4. The predicate symbols $\forall$ and $\exists$.
5. A set of (non-logical) relational symbols with arrity $G_1^{j_1}, G_2^{j_2}$....
6. A set of (non-logical) function symbols with arrity $g_1^{j_1}, g_2^{j_2}$....

We will see some examples in of formal systems in the next two sections.

A.2 Logic

Propositional logic (PL) is the simplest form of classical logic. It was not fully developed until completed by Boole and De Morgan in the later half of the 1800s. It concerns the logic of the combination of atomic statements by logical connectives like "and", "not" and "or". We can view PL as a formal system.

Definition A.2.1 — Propositional logic system. A propositional logic system is a formal system P supporting the use of PL. The alphabet A of such a system has only the first three component mentioned above.

1. A set of variables $A = (x_1, x_2, ...)$
2. A set of logical connectives (like $\wedge, \vee$ etc.), written as $\Omega = f_1^{j_1}, f_2^{j_2} \cdots$ where j is the arrity.
3. A set of punctuation symbols $\{ \quad (\quad) \quad , \quad \}$ etc.

The grammar of P is given by inductively defined set.

1. For any $x_j \in A$ x_j is a wff.
2. For any f_n^j and a set of wffs $(p_1, p_2, ..., p_j)$ $\left(f_n^j(p_1, p_2, ..., p_j)\right)$ is a wff.
3. The wffs produced by 1 and 2 are the only wffs.

The set of logical and non-logical axioms N and NL must also be specified as well as the set RD

PL is unable to formalize notion like "John is the farther of Patrick" which involves functions or relations. It also cannot express the quantifying over all individuals in statements like "all humans are mortal". Predicate logic (or first-order logic) is an extension of PL to include functions, relations and quantifiers.

Definition A.2.2 — First order logic. First order logic (FOL) is an extension of PL systems. The alphabet is extended by component 4, 5 and 6 in the above. The grammar of FOL is defined by the following inductive set. First we define terms.

1. Any variable $x \in A$ is a term.
2. For any set of terms $(p_1, p_2, ..., p_j)$ and a function symbol with arrity j $g^j()$ then $\left(g_n^j(p_1, p_2, ..., p_j)\right)$

is a term.

From the terms we can define wffs.

1. For any set of terms $(p_1, p_2, ..., p_j)$ and a relational symbol with arrity j $G^j()$ then $\left(G_n^j(p_1, p_2, ..., p_j)\right)$ is a wff.
2. For any set of terms $(p_1, p_2, ..., p_j)$ and a logical connective with arrity j $f^j()$ then $\left(f_n^j(p_1, p_2, ..., p_j)\right)$ is a wff.

A.3 Axiomatic set theory

Set theory starts with Georg Cantor in 1874, when he created what is (today) known as naive set theory. Naive set theory is based on the principle of unrestricted comprehension, which can be stated as follows, for any statement $\phi(x)$, where ϕ is a logical statement depending on x, the collection of all objects for which ϕ is true is a set. Briefly we can state that

$$\{x \mid \phi(x)\} \tag{A.3.1}$$

is a set. When Bertrand Russel in 1901 discovered that naive set theory leads to the contradiction known as Russel's paradox, it was just the first of several embarrassing paradoxes found in the foundation of mathematics. This was rightly perceived as a strong threat to the integrity of mathematics. Frege, who had just published his book on the axioms at the basis of arithmetic, felt that the basis for his system had suddenly been swept away. The need to find a sound basis and avoid contradictions was immediate. Russels and Whiteheads solution in Principia, using the theory of types, was successfull but cumbersome. Zermelo in 1908 created a first order theory without types, that was later improved by Fraenkel in 1921. Zermelo-Fraenkel + AC now forms the de facto basis of all mathematics. In the ZFC there is only one predicate symbol. It is a binary predicate symbolizing membership. The only concrete set that we can define is the empty set. The empty sets is defined by using a statement that is always false. All other sets are built from the empty set. We will see how this is done for the natural numbers, and then how the natural numbers form the basis of the integers, the rational numbers and the real numbers.

Definition A.3.1 — Zermelo-Fraenkel axiomatic set theory. ZF-set theory is a first-order logic with standard first-order logical with the added non-logical binary relation $\in$ and the following six (non-logical) axioms.

1. Axiom of extensionality
2. Axiom of replacement
3. Axiom of power sets
4. Axiom of unions
5. Axiom of regularity
6. Axiom of infinity

Definition A.3.2 — Axiom of choice. A choice function is a function f, defined on a collection X of nonempty sets, such that for every set A in X, $f(A)$ is an element of A. The axiom of choice is the statement: For any set X of nonempty sets, there exists a choice function f defined on X.

Definition A.3.3 — Continuum hypothesis. Two sets are said to have the same cardinality or cardinal number if there exists a bijection (a one-to-one correspondence) between them. The

Axiom	**Content**
Axiom of extensionality	Two set with the same elements are identical
Axiom of replacement	The image of a set under a definable function is a set
Axiom of power sets	The power set of a set is a set
Axiom of unions	The union of two set is a set
Axiom of regularity	Every set has a disjoint member set
Axiom of infinity	There is an infinite set

Table A.1: Axioms of ZF

Axiom	**Content**
Axiom of extensionality	$\forall x \forall y[\forall z(\, z \in x \leftrightarrow z \in y\,) \rightarrow x = y]$
Axiom of replacement	$\forall u_1 \ldots \forall u_k[\forall x \exists! y \phi(x, y, \hat{u}) \rightarrow \forall w \exists v \forall r(r \in v \leftrightarrow \exists s(s \in w \wedge \phi_{x,y,\hat{u}}[s, r, \hat{u}]))]$
Axiom of power sets	$\forall x \exists y \forall z[z \in y \leftrightarrow \forall w(w \in z \rightarrow w \in x)]$
Axiom of unions	$\forall x \exists y \forall z[z \in y \leftrightarrow \exists w(w \in x \wedge z \in w)]$
Axiom of regularity	$\forall x[x \neq \varnothing \rightarrow \exists y(y \in x \wedge \forall z(z \in x \rightarrow \neg(z \in y)))]$
Axiom of infinity	$\exists x[\varnothing \in x \wedge \forall y(y \in x \rightarrow \bigcup\{y, \{y\}\} \in x)]$

Table A.2: Axioms of ZF

axiom usually referred to as the continuum hypothesis is the statement: There is no set S whose cardinality $|S|$ is

$$\aleph_0 < |S| < 2^{\aleph_0}. \tag{A.3.2}$$

Where $\aleph_0$ is the cardinality of the natural numbers and $2^{\aleph_0}$ is the cardinality of the power set of the natural numbers.

Definition A.3.4 — ZFC. ZFC is the ZF-set theory with the added axiom
1. Axiom of choice

The following axioms are often listed as axioims but are actually redundant in ZFC.
1. Axiom of set existence
2. Axiom of pairing
3. Axiom of empty set
4. Axiom of separation (Specification / Separation / Selection / Restricted comprehension)

The axioms of above can be proved to be redundant by the following arguments The axiom of infinity implies that there exists a least one set. Thus the axiom of set existence is true and the theory is not empty. The axiom of replacement (in the version which does not demand totality) implies the axiom of separation. Using the existence of a least one set with axiom of separation, with a contradiction as membership criteria allows one to derive the empty set. The axiom of extensionality shows that this set is unique. The existence of a set with two members follows from the axiom of infinity. Using the axiom of separation allows one to deduce the uniqueness of this set. This is equivalent to the axiom of pairing.

A.4 Relations and functions

In this section we collect various definitions related to basic properties of relations and functions. We assume the following terms as known: domain, codomain/range, function, relation, injective, surjective, bijective, inverse of a function, inverse image.

Relational property	**Definition**
Symmetric	$(\forall x, y)\,(R(x, y) \implies R(y, x))$
Antisymmetric	$(\forall x, y)\,(R(x, y) \land R(y, x) \implies x = y)$
Asymmetric	$(\forall x, y)\,(R(x, y) \implies \neg R(y, x))$
Reflexive	$(\forall x)\,(R(x, x))$
Transitive	$(\forall x, y, z)\,(R(x, y) \land R(y, z) \implies R(x, z))$
Total	$(\forall x, y)\,(R(x, y) \lor R(y, x))$

Table A.3: Relation properties

Ordering	**Reflexive**	**Symmetric**	**Transitive**	**Anti-symmetric**	**Total**
Equivalence relation	+	+	+	-	-
Preorder	+	-	+	-	-
Partial order	+	-	+	+	-
Total order	+	-	+	+	+

Table A.4: Orderings

Definition A.4.1 — Directed preorder. A (upward) directed preorder is a set with a preorder relation where every pair of elements have an upper bound.

Definition A.4.2 — Directed poset. A (upward) directed poset is a set with a partial order relation where every pair of elements have an upper bound.

Definition A.4.3 — Projective family. A projective family (A_j, π_{ij}) is family of set A_j with indexes j in a directed poset $(I, \leq)$ such that

1. For every $i \leq j$ there are surjective maps

$$\pi_{ji} \colon A_j \to A_i \tag{A.4.1}$$

2. The maps π_{ij} are such that

$$\forall (i \leq j \leq k) \pi_{ji} \circ \pi_{kj} = \pi_{ki} \tag{A.4.2}$$

Definition A.4.4 — Projective limit. The projective limit of a projective family (A_j, π_{ij}) is the subset of $A_\infty = \prod A_j$ given by

$$\overline{A} = \{(fixthis) \mid (fixthis)\} \tag{A.4.3}$$

A.5 Mathematics from axioms

Mathematics can be easily and efficiently developed from the ZFC axiom set. Axiomatic and formal mathematics is not a far off prospect, but a very real and readily realizable way of developing mathematics. To partially demonstrate this, we will show how the Peano axioms (or equivalently the natural numbers) can be developed from the axioms. From this basis, the rational numbers, the real numbers, the complex numbers and so on are straight forward to define (with the benefit of hindsight). Many mathematical areas just need these number system, along with a system of logic, to be on a well defined formal basis. Some mathematical areas also need to use the axiom of choice (AC) and/or the continuum hypothesis (CH). Several well-known mathematical statements (or theorems) are equivalent to AC. Examples of this are Tychonow's theorem and Zorn's lemma.

We believe that formal systems and axiomatics will be essential in the future of both mathematics and physics. There is likely to be a near future where only mathematics with proofs checked by computer algorithms will be accepted for publication, and where the full set of mathematical knowledge is stored in a large database in computer readable form. Systems like Methmath [404] and Coq [405] already exists, and have encoded large parts of the mathematical knowledge base. This development is bound to spill over to theoretical and mathematical physics.

We now set out to build the natural numbers from the basic axioms of set theory. Our goal will be simply to establish the Peano axioms. Once having established the Peano axioms, we assume all the known properties of the natural numbers. These properties are all known to follow from the Peano axioms.

Definition A.5.1 — Peano axioms. The Peano axioms are given by

1. $0 \in \omega$
2. $\forall m \, (m \in \omega \implies m + 1 \in \omega)$
3. $\forall m \, (m + 1 \neq 0)$
4. $\forall m, n \, (m + 1 = n + 1 \implies m = n)$
5. $0 \in A \wedge \forall i \, (i \in a \implies i + 1 \in A) \implies \omega \subseteq A$

The construction builds on the empty set and the elementary properties of sets that we have established. The empty set is the only well defined set that we actually have, so any concrete constructions must build on this set. We start by listing a few facts that we have established. We know that the empty set is a set, we know that the union of two set is a set and we know that the power set of a set is a set. We use the notation $\mathcal{M}(a)$ for saying that a is a set and we state what we just said as

$$\mathcal{M}(\emptyset) \tag{A.5.1}$$

$$\mathcal{M}(a) \wedge \mathcal{M}(a) \implies \mathcal{M}(a \cup b) \tag{A.5.2}$$

$$\mathcal{M}(a) \implies \mathcal{M}(\mathcal{P}(a)) \tag{A.5.3}$$

This means that the following are examples of sets

$$\mathcal{P}(\emptyset) = \{\emptyset\} \tag{A.5.4}$$

$$\emptyset \cup \{\emptyset\} = \{\emptyset\} \tag{A.5.5}$$

$$\{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\} \tag{A.5.6}$$

From this we conclude that we have at our disposal set of the kind $\{\emptyset\}, \{\emptyset, \{\emptyset\}\}$ and so on. By the insights gleaned from the Peano axioms we know that we must define the number zero as well as a successor function that takes a natural number n and produces $S(n) = n + 1$. We define $0 \coloneqq \emptyset$, and we define $n + 1$ by $n + 1 = n \cup (n)$. This means that

$$0 + 1 = \emptyset \cup \{\emptyset\} = \{\emptyset\}. \tag{A.5.7}$$

We can now define the further numbers in the series by

$$1 \coloneqq 0 + 1 = \{\emptyset\} \tag{A.5.8}$$
$$2 \coloneqq 1 + 1 = \{\emptyset, \{\emptyset\}\} \tag{A.5.9}$$
$$3 \coloneqq 2 + 1 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \tag{A.5.10}$$

and so on. The ordering on the natural numbers can be defined by $a < b \iff a \in b$. From these definitions we can prove Peano's axioms for the natural number thereby proving that we have constructed the natural numbers (or something equivalent to them if you want to avoid ontological implications).

From the natural numbers we proceed to the integers. The missing feature of the natural numbers is the additive inverse which we will add by using the Cartesian product of the natural numbers and then imposing an equivalence relation on such pairs of natural numbers. We first define $\tilde{Z} = N \times N$. Then we define a relation $\sim$ on $\tilde{Z}$ by $(m, n) \sim (m', n') \iff m + n' = m' + n$.

Lemma A.5.2 The relation $\tilde{Z}$ defined by $(m, n) \sim (m', n') \iff m + n' = m' + n$ is an equivalence relation.

Proof. We first prove that $(m, n) \sim (m, n)$. By definition $(m, n) \sim (m, n)$ means that $m + n = m+n$, but this is always true hence we have shown than for all $(m, n) \in \tilde{Z}$ we have $(m, n) \sim (m, n)$. Next we show that $(m, n) \sim (m', n')$ implies that $(m', n') \sim (m, n)$. From the assumption we get (by definition) that $m + n' = m' + n$, but equality is a symmetric relation so this implies that $m' + n = m + n'$ which by definition means that $(m', n') \sim (m, n)$. Thus the relation is symmetric. Finally we show that the relation is transitive. Assuming that $(m, n) \sim (m', n')$ and $(m', n') \sim (m'', n'')$ we find (by definition) that $m + n' = m' + n$ and $m' + n'' = m'' + n'$. This means that $m+n'+m'+n'' = m'+n+m''+n$, which again implies that $(m+n'')+(n'+m') = (n + m'') + (n' + m')$, where we have used the commutative and associative properties. This implies that $(m + n'') = (n + m'')$ by the right-cancellation property which by definition means that $(m, n) \sim (m'', n'')$. Thus the relation is transitive. A reflexive, symmetric and transitive relation is an equivalence relation. ■

Let us denote a an equivalence class in , $\tilde{Z}$ by $\langle m, n\rangle$. On the set Z of equivalence classes of , $\tilde{Z}$ we define the following operations

$$\langle m, n\rangle + \langle p, q\rangle = \langle m + p, n + q\rangle \tag{A.5.11}$$
$$\langle m, n\rangle - \langle p, q\rangle = \langle m + q, n + p\rangle \tag{A.5.12}$$
$$-\langle m, n\rangle = \langle 0, 0\rangle - \langle m + q, n + p\rangle \tag{A.5.13}$$
$$\langle m, n\rangle \cdot \langle p, q\rangle = \langle mp + nq, mq + np\rangle. \tag{A.5.14}$$

This completes the definition of the integers Z. It is easily shown that these definitions reproduce the familiar features of the integers. We very briefly sketch the road ahead. The rational numbers will be defined in the same manner as the set of pairs of integers with an equivalence relation on them. Upon this set one again defines new addition and multiplication operations. The real numbers are defined as Cauchy sequences of rational numbers. The complex numbers are defined as pairs of

real numbers. This completes our demonstration of the path from ZFC to ordinary mathematics. The invention of this path was momentous occasion and a momentous achievement. Following in the footsteps of this path is a relatively easy process that puts basis of mathematical structures in a useful perspective.

A.6 Notes

Definitions of formal systems and a treatment of formal logic can be found in Mendelson [406]. Treatments of the Zermelo-Fraenkel axiomatic set theory can be found in Takeuti-Zaring [10], Suppes [407], Ciesielski [408]. The details of constructing numbers systems can be found in Rubin [409] and Suppes [407]. Some ideas about formalizing mathematics can be found in [410]. Some examples of systems of formalize mathetics are Mizar [411]. See [412] for historical details and references.

B. Category theory

B.1 Introduction

Category theory is a part of mathematics that tries to clarify, simplify and organize ideas, collected from many other parts of mathematics, to reveal what is common among what superficially seems to be unrelated. Category theory was introduced by Eilenberg and Mac Lane in the 1940s, mostly with a view towards formalizing constructions in algebraic topology. The function of category theory can be different in different settings.

Some view category theory in a utilitarian perspective, regarding it as a better language, giving a more consistent, abstract and general formulation of the traditional structures found in math. Some see category theory in a foundational perspective, as an attempt at a new starting point for all of mathematics, emphasizing different aspects of mathematical objects than the usual set-and-membership axioms of axiomatic set theory. Both of these views are interesting and they are not necessarily conflicting views. We find that no matter how you look at the ultimate value of category theory, at least three things can be said with certainty. Firstly, category theory is very systematic and orderly, secondly, it brings a new perspective on things, and thirdly, it is ubiquitous in modern mathematics.

Category theory brings order and unity to mathematical structures, for example, continuous function and homomorphism are both classified as morphisms in category theory. This is a much more consistent and logical way of naming things. The new perspective part of categories can be described as something of a more "external" view of structures, describing them mostly through their relationships with other structures, rather that the more "internal" viewpoint of traditional set based definitions. Thus category theory gives a reformulation of all elementary set theoretic concepts in the form of statements about arrows between objects.

Categorically minded people might describe a sense that the set-and-membership idea is getting in to close an object, and that the best approach is to take one step back and look at the bigger picture. Some might compare it to gaining a better perspective, like when going from the "old" differential geometry formulated in a coordinate based language compared to the "modern" coordinate independent language. Sets, like coordinates, are perhaps sometimes too down and dirty to catch the best view of a concept.

Category theory is certainly highly infused into modern mathematics. In textbooks and papers of very many fields, categorical language is used as matter of course. It is also becoming a central factor in mathematical physics. Especially in information theoretic approaches to quantum theory and in certain versions of quantum gravity.

B.2 Basic concepts

The definition of a category is not complicated, but contains some technical details that might obscure its basic simplicity. The following remark suggests the intuitive idea we are seeking.

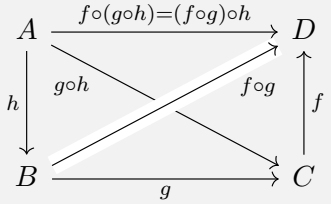
R A category is a collections of objects and morphisms. Morphisms can be imagined as "maps" between objects satisfying various properties.

Definition B.2.1 — Category. A category is a quadruple $\mathcal{C} = (Ob(\mathcal{C}), HOM_{\mathcal{C}}, Id_{\mathcal{C}}, \circ_{\mathcal{C}})$ where the elements are of the following type:

1. The element $Ob(\mathcal{C})$ is a collection of objects
2. The element $HOM_{\mathcal{C}}$ is the collection of morphisms (or arrows). We assign a collection $HOM_{\mathcal{C}}(A, B)$ to every pair of objects $A, B \in Ob(\mathcal{C})$
3. The map $Id_{\mathcal{C}}$ assigns to each object $A \in Ob(\mathcal{C})$ one corresponding element of $HOM_{\mathcal{C}}$ called the identity morphisms 1_A
4. The element $\circ_{\mathcal{C}}$ is a (partial) binary operator that maps every pair of compatible morphisms to a new morphism $\circ : HOM_{\mathcal{C}}(A, B) \times HOM_{\mathcal{C}}(B, C) \to HOM_{\mathcal{C}}(A, C)$

The morphism have the following properties

1. The operation $\circ_{\mathcal{C}}$ is an associative operation (the diagram below is commutative)



2. For each $f : A \to B \wedge g : B \to C$ we have that $1_B \circ f = f \wedge g \circ 1_B = g$
3. The sets of morphisms are pairwise disjoint

The morphisms form a collection with a (partial) binary operation that is associative and has identity. One can replace the elements of the category by their identity morphisms and in this manner the whole concept of a category is such a collection with partial binary operation. Morphisms are often also referred to as arrows. Many advocate this terminology as it emphasizes that morphisms are not necessarily maps.

Note carefully that we are not assuming the collections of objects, or the collection of morphisms, to be sets. The case when these collections actually are sets, forms an important special case, which we note in the two next definitions.

Definition B.2.2 — Small category. A small category means a category where the collection of all morphisms is a set (and not just a class).

Definition B.2.3 — Locally small category. If the class of all morphism between objects A and B (for all A and B) is a set, the category is called locally small.

To get some familiarity with the category concept it is wise to study a few example. We start with some simple but abstract examples that emphasis the structure found in the category concept. Afterwards we see how several common mathematics structures can be embedded and unified in the category concept.

■ **Example B.2.4 — Two object category.** This diagrams in figure B.1 shows a category of two objects. ■

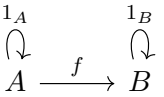


Figure B.1: A simple category of two objects.

■ **Example B.2.5 — Three object category.** In figure B.2 an example of a category with only three objects is shown. ■

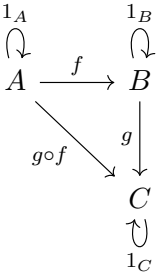


Figure B.2: Typical category diagram showing objects, morphisms, compositions, and identity morphisms

■ **Example B.2.6 — Set category.** The **Set** category, where $Ob(\mathcal{C})$ = the class of all set and for two set A and B $HOM_{\mathcal{C}}(A, B)$ = the set of all functions from A to B. The identity morphism is the identity map on the set and composition of morphisms is regular function composition. ■

■ **Example B.2.7 — Group category.** The **Grp** category, where $Ob(\mathcal{C})$ = the class of all groups and for two groups A and B $HOM_{\mathcal{C}}(A, B)$ = the set of all group homomorphisms from A to B. The identity morphism is the identity map on the group and composition of morphisms is regular function composition. ■

■ **Example B.2.8 — Module category.** The R-**Mod** category, where $Ob(\mathcal{C})$ = the class of all modules over the given ring R, and for two modules A and B $HOM_{\mathcal{C}}(A, B)$ = the set of all R-module homomorphisms from A to B. The identity morphism is the identity map on the module, and composition of morphisms is regular function composition. ■

■ **Example B.2.9 — Poset category.** An individual **PoSet** P forms a category, where $Ob(\mathcal{C})$ = the elements of P, and for any two element A and B of P such that $A \leq B$ we call this an arrow $f : A \to B$ (that is, two objects either have no morphisms between them or this unique morphism). The identity morphism is the "identity map" $A \leq A$, and composition of morphisms follows from the transitive property of the ordering. ■

Definition B.2.10 — Domain and codomain. The source and target of a morphism are called domain and codomain of a morphism.

Definition B.2.11 — Concrete categories. A concrete category is a a category where the objects are sets with extra structure on them, and the morphisms are structure preserving maps between the sets.

■ **Example B.2.12 — Concrete categories.** Top, Grp, Vect ■

Definition B.2.13 — Initial objects. An object 0 of a category $\mathcal{C}$ is an initial objects if for every object A of $\mathcal{C}$ there is exactly on arrow from 0 to A.

Definition B.2.14 — Terminal objects. An object 0 of a category $\mathcal{C}$ is a terminal objects if for every object A of $\mathcal{C}$ there is exactly on arrow from A to 0.

■ **Example B.2.15 — Initial and final objects.** In **Set** - the empty set is the unique initial objects and the singleton sets are the terminal objects. ■

B.3 Functors

Now that we have acquainted ourselves with the basic notion of a category, we want to present a simple example that clarifies the categorical way of thinking versus the traditional set based. Let us take the notion of a binary operation being associative. How can we formulate this notion in categories? Well it is obvious that we cannot talk about the elements of the objects, and that we must instead try to formulate this using properties of combinations of morphisms. Quite often we shall also need "special" objects to describe the specific nature of a category,but once again these "special" objects must be characterized by the morphisms they permit.

Functors are perhaps the most important concept in category theory. We shall have plenty use for them in later parts.

A functor between the categories $\mathcal{A}$ and $\mathcal{B}$ is a map objects to objects and morphisms to morphisms, that preserves identity and compositions.

Definition B.3.1 — Covariant functor. A covariant functor between categories $\mathcal{A}$ and $\mathcal{B}$ is two maps $F_1 : Ob(\mathcal{A}) \to Ob(\mathcal{B})$ and $F_2 : HOM_{\mathcal{A}} \to HOM_{\mathcal{B}}$ such that

1. $f : A_1 \to A_2$ goes to $F_2(f) : B_1 \to B_2$ where $B_1 = F_1(A_1), B_2 = F_1(A_2)$
2. $F(f \circ g) = F(f) \circ F(g)$ (functors preserve the product structure of morphism)
   (a) $F(id_A) = id_{F(A)}$ (functors preserve the identity morphisms)

Definition B.3.2 — Contravariant functor. A contravariant functor is a functor with morphisms in the target category reversed.

By using functors we can declare two categories to be equivalent. We first need a simple definition.

Definition B.3.3 — Identity functor. For a give category $\mathcal{A}$ the identity functor is the functor $ID_{\mathcal{A}} : \mathcal{A} \to \mathcal{A}$ that maps each object to itself, and maps each morphism to itself.

Definition B.3.4 — Equivalent categories. Two categories $\mathcal{A}$ and $\mathcal{B}$ are said to be equivalent

categories if there are two functors $F\colon \mathcal{A} \to \mathcal{B}$ and $G\colon \mathcal{B} \to \mathcal{A}$ such that $G \circ F \simeq ID_{\mathcal{A}}$ and $F \circ G \simeq ID_{\mathcal{B}}$ by equalities $G \circ F = ID_{\mathcal{A}}$ and $F \circ G = ID_{\mathcal{B}}$.

Even though we will not use it, it is perhaps helpful to know about the concept of isomorphic categories. The only difference between equivalent and isomorphic is that the isomorphisms $G \circ F \simeq ID_{\mathcal{A}}$ and $F \circ G \simeq ID_{\mathcal{B}}$ are replace by

Definition B.3.5 — Isomorphic categories. Two categories $\mathcal{A}$ and $\mathcal{B}$ are said to be isomorphic categories if there are two functors $F\colon \mathcal{A} \to \mathcal{B}$ and $G\colon \mathcal{B} \to \mathcal{A}$ such that $G \circ F = ID_{\mathcal{A}}$ and $F \circ G = ID_{\mathcal{B}}$.

B.4 Notes

The basic references for this appendix are Simmons [413] and Lawvere et al. [414].

C. Algebra

Algebra is the part of mathematics that concerns studying generalizations of operations like addition and multiplication.

C.1 Algebraic structures

The binary operation is at the center of almost all algebraic structures.

Definition C.1.1 — Binary operation. Let A be a set. A binary operations $\star$ is a map

$$\star : A \times A \to A \tag{C.1.1}$$

A set with a single binary operation is classified according to the following table. Note that we do not need to include completness or closedness as properties since they follow directly from C.1.1.

Type	Associative	Identity	Inverse	Commutative
Magma	-	-	-	-
Semigroup	+	-	-	-
Monoid	+	+	-	-
Group	+	+	+	-
Abelian group	+	+	+	+

Table C.1: Binary operation

A partial binary operation is a structure similar to a binary operation but the operation is not defined for all pairs in $A \times A$. The identity is not a global object in such a system, i.e. there can be separate identity elements for each element of A.

Definition C.1.2 — Groupoid. A groupoid is a small category where every morphism is invertible (is an isomorphism). On the set of all morphisms (which is a set since it is a small category) we can then define a group like binary operation. The binary operation has all the properties of a group (closed, associative, inverse, identity), except that it is not defined for all pairs of elements (it is a partial function),

Type	Associative	Identity	Inverse	Commutative
Semicategory	+	-	-	-
Category	+	+	-	-
Groupoid	+	+	+	-

Table C.2: Partial binary operation

We shall also use systems with two or more binary operations. In all cases relevant for us the first operation will be an Abelian group and be denoted by +. We also stipulate that the second binary operation will always be (left and right) distributive over the first. With the exception of Lie Algebras the second operation will be associative.

Definition C.1.3 — Distributive. Let S be an Abelian group with operation $+$, with a second binary operation given by $\star$. The second operation ($\star$) is said to be left and right distributive over $+$ if we have

$$x \star (y + z) = x \star y + x \star z \quad \text{and} \quad (y + z) \star x = y \star x + z \star x. \tag{C.1.2}$$

Type	Property	Other properties
Ring	Semigroup	-
Ring with unit	Monoid	-
Division ring	Group	+
Field	Abelian group	+

Table C.3: Two binary operations

A ring with unit will also be called a unital ring.

Definition C.1.4 — Idempotent. Given a binary operation $\star$ on S and and element $\in S$, then s is called idempotent if

$$s \star s = s.$$

If all elements of S are idempotent the operation $\star$ is called idempotent.

There is no consensus on whether a ring always should be unital-ring or what we call a ring (a ring without unit). This is a just a disagreement on terminology. It can cause confusion in other definitions. An example would be that if rings without unit are considered rings, an ideal is a subring, whereas otherwise (non-trivial) ideals are not subrings.

Definition C.1.5 R-Module A left (right) R-module is a triple $M = (G, R, \cdot)$ where
1. G is an Abelian group, $G = (G, +)$
2. R a ring, $R = (R, +, \times)$
3. The operation $\cdot$ is a left (right) scalar multiplication, i.e. a map $\cdot\colon R \times G \to G$ (for a right module $\cdot\colon G \times R \to G$)).

It satisfies the properties

$$\begin{aligned} r \cdot (a + b) &= ra + rb \quad \text{left distributive} \\ (r + s) \cdot a &= ra + sa \quad \text{right distributive} \\ r \cdot (s \cdot a) &= (rs) \cdot a \quad \text{generalized associative} \end{aligned}$$

Definition C.1.6 — R-Algebra. A left(right) R-algebra is an R-module M with an additional binary operation $*$(called multiplication) that is R-bilinear. Or in other words an R-module where M is also a ring, and such that scalar multiplication is R-bilinear. By R-bilinear we mean that in addition to being right and left distributive we have

$$(r \cdot a) * b = a * (r \cdot b) = r \cdot (a * b) \tag{C.1.3}$$

A unital R-algebra is an algebra where the ring is unital.

We have now defined a number of algebraic systems with a variety of operations defined on them. A homomorphism is a type of map between algebraic structures of the same that preserves the algebraic structure. the exact definition of a homomorphism depends on the algebraic structure in question, but the changes in the definition are very systematic. We therefore choose to present all of them in the form of a table. The only true complication is that for structures with a unit but without a full set of inverses (corresponding to this unit), we need to add the additional requirement that the map must preserve the unit element.
We can write the relations between the morphisms and these operations in the following manner. Assume we have a structures $(A, F, +, \times, \cdot, e_+, e_\times)$ and $(B, F, \oplus, \otimes, \odot, e_\oplus, e_\otimes)$. Here we intend $\oplus, \otimes$ to represent the corresponding binary operation in B and not directs sums and tensor products. For a morphism we have

$$\omega(a + b) = \omega(a) \oplus \omega(b) \tag{C.1.4}$$
$$\omega(a \times b) = \omega(a) \otimes \omega(b) \tag{C.1.5}$$
$$\omega(\alpha \cdot a) = \alpha \odot \omega(a) \tag{C.1.6}$$
$$\omega(e_+) = e_\oplus \tag{C.1.7}$$
$$\omega(e_\times) = e_\otimes \tag{C.1.8}$$

We refer to textbooks (see e.g. [415]) for the usual definition of cosets, normal subgroups and ideals, as well as the associated quotient structures and the isomorphism theorems. We will also use the definitions of maximal ideals and prime ideals from [415].

Definition C.1.7 — R-linear map. A R-module morphism is also called an R-linear map.

Definition C.1.8 — Generating set. A generating set $\{g_i\}$ of a module M over the ring R is a subset of M such that $\sum r_i g_i$ can express all elements in the module.

Morphism type	1st operation	2nd operation	Scalar mult.	Id1	Id2
Magma	+	-	-	-	-
Semigroup	+	-	-	-	-
Monoid	+	-	-	+	-
Group	+	-	-	-	-
Abelian group	+	-	-	-	-
Ring	+	+	-	-	-
Unital ring	+	+	-	-	+
Division ring	+	+	-	-	-
Field	+	+	-	-	-
Module	+	-	+	-	-
Algebra	+	+	+	-	-
Unital algebra	+	+	+	-	+

Table C.4: Properties of the various homomorphisms / morphisms. For structures with scalar multiplication the associated fields (or rings) are always assumed to be identical. The names Id1 and Id2 refers to the unit element of the 1st and 2nd operation. A plus in one of these columns implies that we require the unit to be preserved by the morphism. A minus in these columns implies that we do not add this requirement. Note however that in all cases considered here this requirement is still satisfield since if follows implicitly from the other axioms.

Definition C.1.9 — Minimal generating set. A minimal generating set is generating set G such that no proper subset of G is a generating set.

Definition C.1.10 — Finitely generated module. A finitely generated module has a finite generating set $\{g_i\}$.

Definition C.1.11 — Linearly independent set. A subset $S = \{s_i\}$ of an R-module is a linearly independent set if $\sum r_i s_i = 0$ implies $r_i = 0$ for all $i \in I$ for some index set I

Definition C.1.12 R-basis An R-basis of a module M, is a generating set for M that is linearly independent and minimal.

Definition C.1.13 — Free R-module. A free R-module is an R-module that is finitely generated and has an R-basis.

Theorem C.1.14 — Free R-modules. Free R-modules are isomorphic to $R^n = R \oplus R \oplus \dots$ where n is the number of elements of basis.

Theorem C.1.15 — Vector space. A vector space is a module over a field.

Definition C.1.16 — Direct sum and direct product of Abelian groups. The (finite) direct sum (or direct product) $(A \oplus B, \circ)$ of two Abelian groups $(A, *)$ and $(B, \cdot)$, is the set $A \times B$ with operation $\circ$ defined by $(a, b) \circ (a', b') = (a * a', b \cdot b')$. This extends to any finite list of Abelian groups.

The definitions of finite direct sum and finite direct product are identical. For infinite lists of groups they are different.

Definition C.1.17 — Inner semidirect product. Let G be group, and let N be a normal subgroup and H be a subgroup such that $H \cap N = \{e_G\}$. If every element of G can be expressed as a product hn (and nh) we say that

$$G = H \rtimes N \text{ or } G = N \ltimes H. \tag{C.1.9}$$

The semi-direct product is denoted by $\rtimes$ or $\ltimes$ with the "closed" end directed towards the normal subgroup. (Some publications use the opposite convention.)

Definition C.1.18 — Direct sum and direct product of modules. To extend the above definition to direct sum/product of two (left/right) modules over the the same ring R we need to add a definition for the (left/right) scalar multiplication. We define the (left) multiplication operation

$$R \times (A \oplus B) \to A \oplus B \tag{C.1.10}$$

by

$$r(a, b) \mapsto (ra, rb). \tag{C.1.11}$$

Definition C.1.19 — Free module over S. Let S be an arbitrary set and R a unital ring. Let $F(S)$ be the set of all maps $f\colon S \to R$ with finite support. The set $F(S)$ is a R-module by pointwise function addition and scalar multiplication. The functions defined by

$$\begin{cases} \delta_s : S \to R \\ \delta_s(t) = \begin{cases} 1 & t = s \\ 0 & t \neq s \end{cases} \end{cases} \tag{C.1.12}$$

form a basis for $F(S)$.

Definition C.1.20 — Tensor products of modules. The tensor products of two R-modules A and B is defined as follows. Let $F(A \times B)$ be the free group over $A \times B$. This set is an R-module. Define the following equivalence relation $\sim$ on $A \times B$.

$$\forall a, a_1, a_2 \in A, \forall b, b_1, b_2 \in B, \forall c \in R : \tag{C.1.13}$$
$$(a_1, b) + (a_2, b) \sim (a_1 + a_2, b), \tag{C.1.14}$$
$$(a, b_1) + (a, b_2) \sim (a, b_1 + b_2), \tag{C.1.15}$$
$$c(a, b) \sim (ca, b), \tag{C.1.16}$$
$$c(a, b) \sim (a, cb). \tag{C.1.17}$$

The tensor product of A and B is the R-module $A \otimes B = F(A \times B)/\sim$.

See also the balanced tensor product.

C.2 Graph theory

Graph theory started when Euler analyzed the problem of touring the bridges of Koningsberg. Graph theory is concerned with analyzing structures simply based on pairwise connections. We will not use any advanced features of this subject. Loop quantum gravity is based on embedded graphs. We therefore quickly review a few essentials.

Definition C.2.1 — Undirected graph. An undirected graph is a pair $G = (V, E)$ where V is a set of vertices and E is a set of edges. For every edge $v \in V$ we have $e = \{a, b\}$ with $a, b \in V$.

Definition C.2.2 — Directed graph. A directed graph is a pair $G = (V, E)$ where V is a set of vertices and E is a set of edges. For every edge $v \in V$ we have $e = (a, b) \in V \times V$.

Often we allow E to be a multiset. This means that there can be multiple edges between two vertices.

C.3 Clifford algebras and spinors

Clifford algebras play a major role in physics, differential geometry and particularly in noncommutative geometry. Clifford algebras represent a generalization of the algebra of complex numbers and the algebra of quaternions. To define Clifford algebras we need some introductory concepts. We first define bilinear forms which are maps that take two vectors as input to produce an element of the relevant field. (Note that this is basically the same as a (0,2)-tensor.)

Definition C.3.1 — Bilinear form. Let V be a vector space in K-Vect. A bilinear form ω on V is a map $\omega : V \times V \to K$ that is K-linear in both arguments. With $u, v, w \in V$ and $a \in K$ we require that

$$\omega(au, v) = \omega(u, av) = a\omega(u, v) \tag{C.3.1}$$
$$\omega(u + v, w) = \omega(u, w) + \omega(v, w) \tag{C.3.2}$$
$$\omega(u, v + w) = \omega(u, v) + \omega(u, w). \tag{C.3.3}$$

A simple special case of a bilinear form is a symmetric bilinear form.

Definition C.3.2 — Symmetric bilinear form. A symmetric bilinear form is a bilinear form ω that is symmetric in the two arguments. I.e. a bilinear form ω such that

$$\omega(v, w) = \omega(w, v). \tag{C.3.4}$$

A form (or one-form) takes a single vectors as an input to produce an element of the relevant field. A quadratic form is a form with special properties.

Definition C.3.3 — Quadratic form. Given a vector space V in K-Vect a quadratic form Q is a map $Q : V \to K$ such that:

1. The map Q satisfies $Q(kv) = k^2 Q(v)$.
2. The map $\omega : V \times V \to K$ defined by $\omega(v, w) = Q(v+w) - Q(v) - Q(w)$ is a symmmetric bilinear form.

Note that a symmetric bilinear form ω also defines quadratic form Q by $Q(v) = \omega(v, v)$. A Clifford algebra can alternatively be defined by the relation

$$u \cdot v + v \cdot u = 2\omega(u, v). \tag{C.3.5}$$

An inner product (or metric tensor) defines a quadratic form. For a Riemannian manifold the metric tensor defines a quadratic form in each tangent space.

As a vector space the Clifford algebra of a vector space V is isomorphic to the vector space part of the exterior algebra (this is independent of the quadratic form).

We give a simple theorem that helps in classifying quadratic forms. We first deal with the real case.

Theorem C.3.4 Any nondegenerate quadratic form on a real vector space V can be represented (by a suitable choice of basis on V) by a diagonal matrix with diagonal entries d_i, such that $d_i \in \{-1, 1\}$.

The complex case is even simpler.

Theorem C.3.5 Any nondegenerate quadratic form on a complex vector space V can be represented (by a suitable choice of basis on V) by a diagonal matrix with diagonal entries d_i, such that $\forall i d_i = 1$.

We are now ready to define a Clifford algebra. The thing to note about Clifford algebras is that the multiplication is determined by the associated quadratic form.

Definition C.3.6 — Clifford algebra. Given a vector space V in K-Vect with a quadratic form Q the Clifford algebra $Cl(V, Q)$ (associated to V and Q) is the algebra generated from the elements of V by the relation

$$v^2 = Q(v)1. \tag{C.3.6}$$

On n-dimensional vectors spaces we have a set of standard quadratic forms.

Definition C.3.7 — Standard quadratic form - positive. For a real or complex n-dimensional vectors spaces (which are always isomorphic to $\mathbb{R}^n$ or $\mathbb{C}^n$), with elements denoted by $x = (x_!, \ldots, x_n)$ we define the standard (strictly positive) quadratic form Q_n to be

$$Q_n(x) = x_1^2 + \ldots + x_n^2. \tag{C.3.7}$$

For all negative quadratic forms (on real vector spaces) we have the corresponding definition for Q_{-n}. For nondegenerate quadratic forms of mixed signatures we define

Definition C.3.8 — Standard quadratic form - mixed signature. For a real n-dimensional vectors spaces (which is isomorphic to $\mathbb{R}^n$), with elements denoted by

$$x = (x_1, \ldots, x_p, x_{p+1}, \ldots, x_{p+q}), \tag{C.3.8}$$

we define the standard (mixed signature) quadratic form $Q_{p,q}$ to be

$$Q_{p,q}(x) = x_1^2 + \ldots + x_p^2 - x_1^{p+1} - \ldots - x_{p+q}^2. \tag{C.3.9}$$

Associated with vector spaces equipped with one of the standard quadratic forms there are so called standard Clifford algebras.

Definition C.3.9 — Standard Clifford algebra. We define the standard positive Clifford algebras on $\mathbb{R}^n$, to be Clifford algebra on $\mathbb{R}^n$ with the standard positive quadratic form Q_n, and denote it by $Cl_n(\mathbb{R})$. We define the standard mixed Clifford algebra on $\mathbb{R}^{p+q}$, to be the Clifford algebra on $\mathbb{R}^{p+q}$ with the standard positive quadratic form $Q_{p,q}$, and denote it by $Cl_{p,q}(\mathbb{R})$. We define the standard complex Clifford algebra on $\mathbb{C}^n$, to be the Clifford algebra on $\mathbb{C}^n$ with the standard positive quadratic form Q_n, and denote it by $Cl_n(\mathbb{C})$.

Theorem C.3.10 — Classification of Clifford algebras. All Clifford algebras are isomorphic to $M_k(D)$ or $M_k(D) \oplus M_k(D)$ for some integer k and with $D = \mathbb{R}, \mathbb{C}$ or $\mathbb{H}$.

Definition C.3.11 — Spinors. Let $\mathcal{A}_{p,m}$ be a Clifford algebra and π a representation of $\mathcal{A}_{p,m}$ on $L(V)$ for some vector space V, then the elements of V are called spinors and V is called spinor space.

C.4 Graded algebras

Definition C.4.1 — Graded module. A graded vector space is module V with a specified decomposition $V = \bigoplus_{i \in I} V_i$. This decomposition is called a I-grading.

Definition C.4.2 — $\mathbb{Z}_2$-graded module. A $\mathbb{Z}_2$-graded module is a module V with a specified decomposition $V = \bigoplus_{i \in \mathbb{Z}_2} V_i = V_0 \oplus V_1$. This decomposition is called a $\mathbb{Z}_2$-grading.

A $\mathbb{Z}_2$-graded module is also called a super module. In case the module is a vector space it is often called a super vector space. Elements that are member of V_0 or V_1 are called homogeneous elements.

Definition C.4.3 — Parity. The parity of a homogeneous element v is defined by

$$|v| = \begin{cases} 0 & \text{if } v \in V_0 \\ 1 & \text{if } v \in V_1 \end{cases} \tag{C.4.1}$$

The subspace V_0 is called the even subspace, and V_1 is called the odd subspace.

Definition C.4.4 — Graded ring. A graded ring is a ring that has a given decomposition as an Abelian group $R = \bigoplus_{i \in \mathbb{N}} R_i = R_0 \oplus R_1 \oplus \ldots$. Multiplication in the ring must satisfy $R_i \cdot R_j \subset R_{i+j}$, where $i + j$ means $i + j$ mod 2.

A $\mathbb{Z}_2$-graded ring is graded ring where the grading just ranges over $\mathbb{Z}_2$.

Definition C.4.5 — Graded algebra. Version 1 - A graded R-algebra A is a graded ring A over an ungraded ring R.
Version 2 - A graded R-algebra A is a graded ring A over an graded ring R such that

$$A_i R_j \subset A_{i+j} \tag{C.4.2}$$
$$R_i A_j \subset A_{i+j}. \tag{C.4.3}$$

In the next definition we use graded algebra in the sense of the version one definition.

Definition C.4.6 — Differential graded algebra. A differential graded algebra $(A, d))$ is a graded algebra A with a map $d\colon A \to A$ or degree $+1$ or -1 such that

$$d \circ d = 0 \tag{C.4.4}$$

$$d(a \cdot b) = (da) \cdot b + (-1)^{deg(a)} a \cdot (db) \tag{C.4.5}$$

■ **Example C.4.7 — de Rham - DGA.** Let M be a smooth manifold and $\Omega^n(M)$ the set of n-form fields on M. The set $\Omega = \bigoplus_{i=0}^{n} \Omega^i(M)$ is a graded vector space. With the operation $\wedge$ is becomes a graded algebra. Let d be exterior derivation, then (Ω, d) is a DGA. ■

C.5 Notes

The basic references for this appendix are Adkins [415], Dummit [416] and Fraleigh [417]. For Boolean algebras see [418].

D. Measure theory and probability

Measure theory is an abstraction of the theory of integration to very general settings. A measure space is a space with a limited amount of structure. A measure space is not even required to have a topology. A measure on X is a non-negative function on a selected set of subsets of the set X. It is an important fact that, in general, we cannot use the full power set of X and still achieve a measure on X with the desired properties. Measure theory is foundation for both integration theory and the theory of probability. In probability theory we restrict ourselves to measures with total measure one.

D.1 Measure theory and integration

We first define some properties need to select a proper selection of subsets for the present purposes.

Definition D.1.1 — Algebra of sets. Let X be a nonempty set. An algebra of set on X is a collection A of subsets of X that is closed under complements and finite unions.

Corollary D.1.2 For an algebra A of sets on X we have $\emptyset, X \in A$.

Definition D.1.3 — σ-algebra. Let X be a nonempty set. A σ-algebra on X is a collection Σ of subsets of X that is closed under complements and countable unions (not just finite unions).

A σ-algebra is also known as a σ-field.

Corollary D.1.4 A σ-algebra is closed under complements, countable unions and countable intersections.

The space X is specified implicitly by the algebra but it is common to make this set explicit.

Definition D.1.5 — Measurable space. A measurable space is a pair (X, Σ) where X is a set

and Σ is a σ-algebra on X.

We can make measurable spaces into a category by defining their morphisms to be be what is know as measurable functions.

Definition D.1.6 — Measurable function. A function $f : X \to Y$ between two measure spaces (X, M) and (Y, N) is called measurable iff $f^{-1}(E) \in M$ for $\forall E \in N$.

If we require that such a function be a bijection, and that also its inverse is measurable, we have then defined an isomorphism between measurable spaces.

On the set Σ of a measurable space we can define a map called a measure.

Definition D.1.7 — Measure. Let (X, Σ) be a measurable space. A measure on X is a map $\mu : \Sigma \to [0, \infty]$ such that

1. $\mu(\emptyset) = 0$
2. If E_i is a sequence of disjoint sets then $\mu(\bigcup E_i) = \sum \mu(E_i)$.

We now combine all the elements so far defined into one structure.

Definition D.1.8 — Measure space. A measure space is a triple (X, Σ, μ) where (X, Σ) is a measurable space and μ is a measure.

We make measure space into a category by defining the morphisms between measure spaces as measure preserving measurable functions.

Definition D.1.9 — Measure space morphism. Let (X, Σ_X, μ_X) and (Y, Σ_Y, μ_Y) be measure spaces. A measure space morphism is a map $f\colon X \to Y$ that is measurable and for which $\mu_X(f^{-1}(E)) = \mu_Y(E)$ for $\forall E \in N$.

Integration theory can now be built up starting with integration of particularly simple functions, and then defining the general integral as a limit of such simple integrations (see e.g. Folland [21]).

The concept of the space of essentially bounded functions is important for our development of algebraic probability theory.

Definition D.1.10 — Bounded function. Let f be a function from a space X to either the reals or the complex numbers. We say that f is a bounded function if

$$(\exists M \in \mathbb{R})(\forall x \in X)|f(x)| < M \tag{D.1.1}$$

Definition D.1.11 — Essentially bounded function. Let (X, Σ, μ) be a measure space. A function $f\colon X \to \mathbb{R}$ is called an essentially bounded function if there exists a bounded measurable function $g\colon X \to \mathbb{R}$ (w.r.t. the Borel measure on $\mathbb{R}$) such that $f = g$ except for a set $M \in \Sigma$ which satisfies $\mu(M) = 0$ (i.e. except for set of measure zero).

A corresponding definition can be made with $\mathbb{C}$ instead of $\mathbb{R}$

D.2 Probability theory

Probability theory is very important for the physical sciences. It represents an axiomatization of the basic structures believed to be true about probability. From the restricted perspective of definitions, a probability space is nothing else than a measure space of measure one.

Definition D.2.1 — Probability space. A probability space is a measure space (X, Σ, μ) with $\mu(X) = 1$.

The morphisms for a probability space are the same as for a more general measure space (see definition D.1.9).

On a topological space one interesting way to construct a σ-algebra is by using the topology of a space. One way of doing this results in the Borel sets and the Borel measures.

Definition D.2.2 — Borel sets. Let X be a topological space. A Borel set of X is any set that is generated from its open sets by countable unions, countable intersections and complementation.

Lemma D.2.3 Let X be a topological space. The set of all Borel sets of X is the smallest σ-algebra containing all open sets.

We also denote this as the Borel σ-algebra.

Definition D.2.4 — Borel measure. Let X be a topological space. A Borel measure on X is any measure defined on the σ-algebra of Borel sets of X.

Definition D.2.5 — Radon measure. A Radon measure is a Borel measure on a Hausdorff space X that is locally finite and regular.

In probability the most important functions are referred to as random variables. In tune with their name (as variables) the domain of such functions are often suppressed.

Definition D.2.6 — Random variable. A real/complex random variable is a real/complex valued function (on a probability space), that is measurable with respect to the Borel measure on $\mathbb{R}$ (or $\mathbb{C}$).

D.3 Notes

The basis reference for this appendix on measure theory and probability theory is the book by Folland [21]. A more category related definition of probability spaces can be found in [22].

E. Topology

Topology is one of the major branches of mathematics. It is said to be the only major branch that was not anticipated (or considered) in ancient times [103]. Topology can be said very loosely to be a kind of shapeless putty-dough "geometry" that does not use neither distances nor angles. The only basic notion is an abstract idea of proximity encoded in a listing of special subsets called open sets. We put geometry in quotes since it is not really geometry at all, but we still feel tempted to use the word since it is probably impossible for humans to visualize something like a "topological sphere" without putting a geometry on it. We therefore occasionally need to remind ourselves that a sphere and a (hollow) cube have the same topological structure, and that the single handle coffee cup and the torus are also equivalent topologically. The proper "picture" of a pure topological space should be completely without shape or geometric form. From the technical point of view, topology deals with general definition of continuity, limits, and convergence.

E.1 Basic topology

The fundamental definition of topology consist of setting up a designated collection of subsets and demanding that they have certain closure properties under the operations of unions and intersections. These properties under these basic set operation is the abstract specification of "closeness" properties.

Definition E.1.1 — Topological space. A topological space is a pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T}$ is a collection of subsets of X (called the topology on X) such that

1. The sets $\emptyset$ and X are members of $\mathcal{T}$.
2. The set $\mathcal{T}$ is closed under finite intersections.
3. The set $\mathcal{T}$ is closed under arbitrary unions.

From the basic definition we have two immediate pieces of terminology that are very useful.

Definition E.1.2 — Open set. For a topological space $(X, \mathcal{T})$ an open set is a subset of X that is a member of $\mathcal{T}$.

Definition E.1.3 — Closed set. Let $(X, \mathcal{T})$ be a topological space. A closed set of $(X, \mathcal{T})$ is a subset of X that is the complement of an open set.

Note that the words open and closed can be misleading. In many other contexts these words are opposites, which is to say that not closed means the same as open. In topology a closed set is simply the complement of an open set. Its important to note that nothing prevents a set from being both open and closed, neither open nor closed, or just closed or just open.

After open (and closed) sets, neighborhoods are the second most important concept. Some topology texts use neighborhoods as the primary objects, and define open sets in terms of them. One reason for this could be that neighborhoods have perhaps a more intuitive relation to the "closeness" concept. In most modern topology open sets are considered fundamental and neighborhoods are a derived concept.

Definition E.1.4 — Neighborhood. A neighborhood N of a point p in a topological space $(X, \mathcal{T})$ is a subset of X containing an open set O containing p. We write $p \in O \subseteq N \subseteq X$.

Definition E.1.5 — Open neighborhood. A open neighborhood is a neighborhood that is also an open set.

Note that some authors define a (general) neighborhood to be what we will call an open neighborhood.

Before introducing further concepts let us look at some examples of topologies. The two simplest examples of topologies valid for arbitrary sets are the one given by the power set, and the one given by just the complete set along with the empty set.

■ **Example E.1.6 — Trivial topology.** Let X be any set and let $\mathcal{T} = \{\emptyset, X\}$, then $(X, \mathcal{T})$ is a topological space. This topology is called the trivial topology. ■

■ **Example E.1.7 — Discrete topology.** Let X be any set and let $\mathcal{T} = \mathcal{P}(X)$, then $(X, \mathcal{T})$ is a topological space. This topology is called the discrete topology. ■

To complete the category of topological spaces we need to define the morphisms between the objects. Morphisms between topological spaces are called continuous functions, while isomorphisms are called homeomorphisms. To enable a general definition of continuous functions can be said to be the raison d'Ã[a]tre for topology. A topology is the minimal structure that we must put on a set to enable us to define continuous functions.

The definition of a continuous function is quite simple. It is a function that preserves the structure of the open sets. The only thing to note is that this preservation of topologies works in the opposite direction of the function.

Definition E.1.8 — Continuous function. A continuous function f between the topological spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ is a map $f : X \to Y$ such that for all $O_Y \in \mathcal{T}_Y$ we have that (the inverse image) $f^{-1}(O_Y) \in \mathcal{T}_X$.

An immediate and important result is that continuity is preserved under composition of continuous functions.

Theorem E.1.9 Let $f : A \to B$ and $g : B \to C$ be two functions between the topological

spaces A and B, and B and C respectively. Let h be the functions $g \circ f : A \to C$. If f and g are continuous then h is continuous.

Proof. Let W denote arbitrary open set of C. We know by definition that $g^{-1}(W)$ is an open set W' of B. And also by definition we know that $f^{-1}(W')$ is an open set of A. Therefore $(g \circ f)^{-1}(W) = f^{-1}(g^{-1}(W)) = f^{-1}(W')$ is an open set of A. ■

A topological isomorphism is a bijection that preserves the topological structure in both directions.

Definition E.1.10 — Homeomorphism. A homeomorphism f between the topological spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ is a bijection $f : X \to Y$ that is bicontinuous. That is, a one-to-one onto map such that both f and f^{-1} are continuous.

In the above all the necessary ingredients to define a category of topological spaces have been developed.

Definition E.1.11 — Top. The collection **Top** is the pair $(\mathcal{O}_{Top}, \mathcal{M}_{Top})$ where $\mathcal{O}_{Top}$ are all possible topological spaces, and $\mathcal{M}_{Top}$ is the collection of all continuous functions between them. We define compositions of morphism between two morphism that have appropriately matching codomain and domain to be regular function composition. We define the identity operator on each set to be the ordinary identity map.

Corollary E.1.12 — Top. The pair **Top** $= (\mathcal{O}_{Top}, \mathcal{M}_{Top})$ is a category.

Proof. We have shown above that composition of continuous functions produces a new continuous function and hence composition of morphisms by function compostion produces a new morphism as required. Function composition is associative and hence morphism composition is also associative. The identity map on a topological space is obviously a continuous functions and hence is a morphism. The identity map when composed with other functions has the properties of a left (or right) identity. Thus **Top** is a category. ■

It is not always possible, convenient or necessary to give a topology on a space by listing all the (sub)sets that make up the topology. Instead we often specify a topology by given a smaller collection of sets from which it can be derived. Such a collection is called a basis. We first define what constitutes a basis for an existing topology.

Definition E.1.13 — Basis for an existing topology. A collection of subsets $\mathcal{B}$ is said to be a basis for the (existing) topology on a topological space $(X, \mathcal{T})$ iff every member of $\mathcal{T}$ is a union of members of $\mathcal{B}$.

Next we define what is required for a collection of subset to form a basis for some topology on a set.

Definition E.1.14 — Basis for a new topology. Let X be a set, and $\mathcal{B}$ a collection of subset of X. The collection $\mathcal{B}$ is said to be a basis for a (new) topology on X iff

1. For each $x \in X$ there is at least one $B \in \mathcal{B}$ such that $x \in B$.
2. If $x \in B_1 \cap B_2$ with $B_1, B_2 \in \mathcal{B}$ then $\exists B_3 \in \mathcal{B}\,(x \in B_3 \subset B_1 \cap B_2)$.

The interesting question now is to say what the open set generated by such a basis looks like. The answer is simple.

Lemma E.1.15 — Generated topology. The topology generated by $\mathcal{B}$ is the collection of all sets that can be expressed as unions of elements of $\mathcal{B}$.

It is useful to have some vocabulary for comparing topologies.

Definition E.1.16 — Weaker and stronger topologies. Let $(X, \mathcal{T})$ be a topological space and $\mathcal{S}$ be another topology on X. If $\mathcal{S} \subset \mathcal{T}$ we say that $\mathcal{S}$ is a weaker (or coarser or smaller) topology than $\mathcal{T}$. Correspondingly $\mathcal{T}$ is said to be a stronger (or finer or larger) topology than $\mathcal{S}$.

E.2 Further concepts

Having defined the basics of topology in the previous section we are now ready for further developments. The next step is the three related concepts of the interior, the closure and the boundary of a set.

Definition E.2.1 — Interior. The interior $Int(A)$ of a set A is the union of all open sets contained in A.

The interior is the largest open set contained in A. A point $p \in Int(A)$ is called an interior point of A .

Definition E.2.2 — Exterior. The exterior $Ext(A)$ is the union of all open sets that are disjoint from A.

Definition E.2.3 — Closure. The closure $Cl(A)$ of a set A is the intersection of all closed set containing A. The complement of the closure of A is equal to the exterior of A.

The closure is the smallest closed set containing A.

Definition E.2.4 — Boundary. The boundary ∂A of a set A is the set $\partial A = Cl(A) - Int(A)$.

The boundary is are elements that are in the closure of A but not in the interior of A. Any set A splits a topological space X into $X = Int(A) \cup \partial A \cup Ext(A)$.

The definition of topological spaces is very general. For this reason there are many spaces which are topological spaces but are very strange compared to what we normally consider a space to be. In most applications we usually restrict the possible topological spaces to spaces that comply with some properties that make them more "normal". This is of course significant in the sense that future applications in fundamental physics could choose to consider a less restricted set of topological spaces. One of the most basic restrictions used is the restriction to spaces that are Hausdorff.

Definition E.2.5 — Hausdorff topological space. A Hausdorff topological space is a topological space X such that for each pair of elements x and y of X there exists open sets O_1, O_2 such that $O_1 \cap O_2 = \emptyset$ and $x \in O_1$ and $y \in O_2$.

The next properties we want to define are compact and paracompact. Before we define these we need some terms describing covers and their properties. The definition of a cover is very simple.

Definition E.2.6 — Cover. A collection of sets $C = \{U_\alpha \mid \alpha \in A\}$ is a cover of the set X if

$$X \subset \bigcup_{\alpha \in A} U_\alpha$$

We shall only use the concept of cover in the context where X is a topological space. We need some further rather simple terminology.

Definition E.2.7 — Subcover. A subcover C' is a subset of a cover C of X that is also a cover of X.

Definition E.2.8 — Open cover. A cover C of a topological space $(X, \mathcal{T})$ is called an open cover of X if all the sets in C are open.

A refinement is a somewhat more subtle term. Note that it doesn't just add more open sets.

Definition E.2.9 — Refinement. A refinement of a cover C of a topological space $(X, \mathcal{T})$ is cover D of X such that each set of D is contained in some set of C.

We need this next definition to define paracompactness.

Definition E.2.10 — Locally finite. A cover C of a topological space $(X, \mathcal{T})$ is called a locally finite cover of X if every point has a neighbourhood N that intersects only finitely many set of the cover C.

We are now finally ready to give the definition of compact. Being (or not being) compact is an important concept when classifying spaces.

Definition E.2.11 — Compact. A topological space is called compact if every open cover has a finite subcover.

Loosely speaking a compact space means closed or finite space. Spaces like the sphere, or the closed intervals of the real line are both compact. Paracompact is a term we will use to limit the topological spaces that can be considered as manifolds. Manifolds are treated later.

Definition E.2.12 — Paracompact. A topological space is called paracompact if every open cover has a locally finite open refinement.

Equivalently we could have used that every cover contains a countable refinement. Paracompact spaces need not be finite or closed, but when they are infinite, they are infinite in a way that very loosely is "no worse" than $\mathbb{R}^n$. A compact space is paracompact, and every metric space is paracompact.

Connectedness is also an important topological property. It means more or less what we would expect, but the definition is still a bit abstract since we don't have many properties to refer to at this stage.

Definition E.2.13 — Connected. A topological space is connected iff it is not the union of two (or more) open disjoint nonempty subsets.

A connected space is more or less a space that cannot be "broken apart" into two or more separate topological spaces. A different but similar connectedness concept is that of being path connected.

Definition E.2.14 — Path connected. A topological space X is path connected iff it for all points x, y there is a continuous map $\gamma : [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = y$.

We can use properties of paths to give the definition of simply connected, which loosely mean something like that there are no holes around which loops can wind around and be able to get untangled from. This definition includes terminology from homotopy theory which is defined i a later chapter.

Definition E.2.15 — Simply connected. A topological space X is simply connected iff it is path connected and its fundamental group an each of its points is trivial (i.e. any continuous map $\gamma : S^1 \to X$ can be contracted to a point).

Definition E.2.16 — Limit point. Let S be a subset of a topological space X. A point x in X is a limit point (or accumulation point or cluster point) of S if every neighbourhood of x contains at least one point of S different from x itself.

Definition E.2.17 — Dense set. A subset A of a topological space X is called dense if every point x in X either belongs to A or is a limit point of A.

Proposition E.2.18 If a subset A of X is dense in X then $\overline{A} = X$.

Some authors use the above proposition as the definition of dense, and derive our definition as a proposition. We can see that intuitively dense means that all elements of X are either in A or can be expressed as sequences of elements of A.

The notion of a separating set is important both for algebraic geometry and for quantization.

Definition E.2.19 — Separating set. Let S be a subset of functions from the set X to the set A. The set S is called a separating set (of functions) if for every $x, y \in X$ with $x \neq y$ there is a functions $f \in S$ such that $f(x) \neq f(y)$.

We also say, we the same meaning, that S separate points (of X). The notion of a separating set of functions is related to the property of being dense.

Theorem E.2.20 — Stone-Weierstrass theorem. Let X be a compact Hausdorff space, and let $C(X, \mathbb{R})$ be the set of continuous real-valued functions on X. Let S is a subalgebra of $C(X, \mathbb{R})$ that contains a non-zero constant function. Then S is dense in $C(X, \mathbb{R})$ iff it separates points of X.

We include a useful characterization of closed set by their properties with respect to convergence of generalized sequences (called nets).

Theorem E.2.21 — Closed set. A subset A of a topological space X is closed in X if and only if every limit of every net of elements of A also belongs to A.

E.3 Derived topological spaces

In mathematics we are always interested in how we can make more stuff using the stuff we have. For topology this means combining topological spaces into new such spaces, as well as other maneuvers for making smaller or larger spaces from exiting ones. Such operations are not just

about creating new spaces but also about providing an understanding of where some spaces "come from".

The first procedure we define involves making a subset of a topological space into its own topological space with a topology based on the topology of the superset.

Definition E.3.1 — Relative topology. Let Y be a non-empty subset of the topological space $(X, \mathcal{T})$. Let $\mathcal{S}$ be be given by

$$\mathcal{S} = \{Y \cap O \mid O \in \mathcal{T}\}. \tag{E.3.1}$$

The pair $(Y, \mathcal{S})$ is a topological space. The topology $\mathcal{S}$ on Y is called the relative topology (or subspace topology, or induced topology).

Corollary E.3.2 The relative topology defined above is a topology on Y

Proof. We use the setup from definition E.3.1. 1) We first show that $\emptyset, Y \in \mathcal{S}$. Since $\emptyset \in \mathcal{T}$ then $\emptyset \cap Y = \emptyset$ is in $\mathcal{S}$. Since $Y \in \mathcal{T}$ then $Y \cap X = X$ is in $\mathcal{S}$. 2) Second, we show it to be closed under finite intersections. Any finite collection of open set of $\mathcal{S}$ can be written as $Y \cap O_1, \ldots, Y \cap O_n$ for some collection $O_1, \ldots, O_n \in \mathcal{T}$. The intersection of these set is $(Y \cap O_1) \cap \ldots \cap (Y \cap O_n)$ which is equal to $(Y \cap \ldots \cap Y) \cap (O_1 \cap \ldots \cap O_n) = Y \cap (O_1 \ldots \cap O_n)$ by the commutativity and associativity of intersections (also $Y \cap \ldots \cap Y = Y$). The set $(\cap O_1) \ldots \cap O_n)$ is a finite intersection of open set of X so it is an open set of X. Hence $Y \cap (O_1 \ldots \cap O_n)$ is an open set of Y. 3) Third, we show it to be closed under arbitrary unions. An arbitrary collection of open sets in Y can be written $(Y \cap O_i$ with $i \in I$ some index set. The union of these set is $\bigcup_{i \in I} (Y \cap O_i)$. This is equal to $\bigcup_{i \in I} (O_i \cap Y) = \bigcup_{i \in I} (O_i) \cap Y$. The set $= \bigcup_{i \in I} (O_i)$ is an arbitrary union of open sets in X so it is an open set in X. Hence $\bigcup_{i \in I} (O_i) \cap Y$ is an open set of Y. ■

Definition E.3.3 — Initial topology. Let Y be a topological space and let $\{f_i\}$ be a set of functions $f_i : X \to Y$. The initial topology on a set X is the weakest topology that makes all the functions $\{f_i\}$ continuous.

Sometimes the initial topology and weak topology is treated as synonyms, but we will treat the weak topology as a special case of the initial topology.

Definition E.3.4 — Weak topology. Let X be a vector space. The weak topology on X is the initial topology on X with respect to the functions defined by the dual space $X^\star$.

E.4 Notes

The basic references for this appendix are Munkres [419], Kelly [420], and an online book by Morris [104].

F. Differential geometry

Differential geometry is based on spaces that are locally Euclidean. Basically such a space is a topological space where every open set looks topologically like $\mathbb{R}^n$ (or equivalently looks like an open set of $\mathbb{R}^n$). These spaces are called (topological) manifolds. Topological manifolds are topological spaces that can be "charted" by coordinate charts. For a topological manifolds the charts are only related by continuity. To enable the use of differentiation and calculus we need to demand coordinate charts that are "mutually compatible" with regards to differentiation. Spaces in this category are called differentiable manifolds. To use concepts from geometry, some measure of distance and angles need to be added. This is done by adding a so called metric tensor to the manifold. Metric tensors are properly inner products on each tangent space. Such tensors are called metrics tensors since one can derive a proper metric on the manifold from them. Manifolds with a metric tensor are called Riemannian manifolds. Another set of spaces which are of interest for us are bundles. Bundles can be purely topological, they can be differentiable, and they can also be geometrical. Bundles are a generalization of the Cartesian product of two spaces. Bundles look locally like a product of two spaces, but can have very different structures globally.

F.1 Differential topology

The starting point is the most general manifold called a topological manifold. As mentioned we only require a certain local structure for such a manifold, and even though there are "charts", differentiability is not meaningful (since it is not a chart independent notion).

Definition F.1.1 — Topological manifold. A n-dimensional topological manifold is a Hausdorff topological space $(X, \mathcal{T})$ such that each open set in $O \in \mathcal{T}$ is homeomorphic to an open set $V \subset \mathbb{R}^n$.

One often requires the topological space to be paracompact or second-countable. A topological space is already sufficient to define continuous functions. Therefore the locally Euclidean criteria is just an extra constraint on the topology. To define differentiable functions and do calculus we need to specify a collection of charts (or coordinate systems) on the space that enables the definition of

differentiation on manifolds.

Definition F.1.2 — Chart. A chart on an topological space X is a pair (U, φ), where U is an open set of X, and φ is a homeomorphism

$$\varphi \colon U \to V, \tag{F.1.1}$$

where V is an open set $V \subset R^n$ for some n.

To support differentiation the charts must be compatible in the following sense.

Definition F.1.3 — Compatible charts. Two charts (U_1, φ_1) and (U_2, φ_2) on an n-dimensional topological manifold M are called C^k-compatible iff the map

$$\varphi_1 \circ \varphi_2^{-1} : \varphi_2(U_1 \cap U_2) \to \varphi_1(U_1 \cap U_2) \tag{F.1.2}$$

is C^k.

The homeomorphism included in the definition of the topological manifolds are charts. Such charts are invertible but are only guaranteed to be C^0-compatible on overlapping domains. We will usually assume that charts are C^∞-compatible. We now define collections of charts such that they together cover the manifold.

Definition F.1.4 — Atlas. A collection of charts $\{(U_i, \varphi_i)\}$ on and n-dimensional topological manifold M are called an atlas if $\bigcup U_i = M$.

To make progress towards differentiability we must again require that all pairs of charts in an atlas are compatible.

Definition F.1.5 — C^k-atlas. An atlas $\mathcal{A} = \{(U_i, \varphi_i)\}$ is called a C^k-atlas iff for any two charts in $\mathcal{A}$ the two charts are C^k compatible.

To avoid defining manifolds to be different because of simply adding or removing a chart we prefer to define them with respect to a maximal set of charts.

Definition F.1.6 — Maximal C^k-atlas. A C^k-atlas $\mathcal{A}$ is called a maximal C^k atlas iff for any chart (U, φ) that is C^k compatible with each chart in $\mathcal{A}$ then $(U, \varphi) \in \mathcal{A}$.

We are now ready to put a differentiable structure of the manifold.

Definition F.1.7 — C^k differentiable manifold. A n-dimensional C^k differentiable manifold is pair $(\mathcal{M}, \mathcal{A})$ where $\mathcal{M}$ is a topological manifold and $\mathcal{A}$ is a maximal C^k-atlas.

Note that the definition of a maximal atlas does not say that this maximal atlas is unique. It may be the case that there exists different maximal atlases available for a given manifold. This means that in many cases there are different possible differentiable structures that we can put on a given manifold. We will mostly consider C^∞-manifolds.

Definition F.1.8 — C^k-function. A function $f \colon M \to \mathbb{R}$ with maximal C^l-atlas $\mathcal{A} = \{(U_i, \varphi_i)\}$ is a C^k-function iff $k \leq l$ and for every φ_i the function

$$f \circ \varphi_i^{-1} \colon \mathbb{R}^n \supset V_i \to \mathbb{R} \tag{F.1.3}$$

is C^k-differentiable.

We denote the set of C^k-differentiable functions on M by $C^k(M)$. Specifically we have the space $C^\infty(M)$ of smooth functions on a smooth manifold M. Note that the set $C^k(M)$ for any k forms a commutative **R**-algebra under pointwise addition and multiplication along with scalar multiplication with real scalars.

Definition F.1.9 — Closed manifold. A closed manifold is a compact manifold without boundary.

Definition F.1.10 — Orientable manifold. An n-dimensional manifold M is orientable if there exists a differential n-form field $\omega(x)$ which is everywhere non-zero.

F.2 Vectors

Definition F.2.1 — Tangent vector. Let M be a C^∞-manifold and let $p \in M$. A tangent vector v_p at p is a map

$$v_p \colon C^\infty(M) \to \mathbb{R} \tag{F.2.1}$$

such that v_p is a generalized $\mathbb{R}$-derivation.

The derivation in this definition is sometimes called a derivation at a point but it fits nicely into the definition H.9.4. For C^k manifold or for C^k-mappings one would have to introduce the notion of germs (see [33] or [45]).

Definition F.2.2 — Tangent space. Let M be a smooth manifold and let $p \in M$. The tangent space T_pM at p is the set of all tangent vectors at p.

Proposition F.2.3 — Tangent space basis. Let (U, φ) be a chart on M with $p \in U$. The mappings

$$\partial_{x_i} \colon C^\infty(M) \to \mathbb{R} \tag{F.2.2}$$

$$\partial_{x_i} \colon f(x) \mapsto \left.\frac{\partial(f \circ \varphi^{-1})}{\partial x_i}\right|_{\varphi(p)} \tag{F.2.3}$$

is a basis for the tangent space T_pM.

Proof. ■

We refer to the basis in proposition F.2.3 as the coordinate basis.

Definition F.2.4 — Cotangent space. Let M be a smooth manifold T_pM be the tangent space at p. The cotangent space T_p^*M at p is the dual of T_pM.

Proposition F.2.5 — Cotangent space basis. Let ∂_{x_i} be the coordinate basis for T_pM, then

the maps

$$dx^i \colon T_pM \to \mathbb{R} \tag{F.2.4}$$
$$dx^i \colon \partial_{x_j} \mapsto \delta^i_j \tag{F.2.5}$$
$$dx^i \colon \sum V^j \partial_{x_j} \mapsto V^i \tag{F.2.6}$$

is a basis for the cotangent space T^*_pM.

Proof. ■

Definition F.2.6 — Covector. A covector (or one-form) ω at $p \in M$ is an element $\omega \in T^*_pM$.

Definition F.2.7 — Tangent bundle. The tangent bundle of a manifold M is the set

$$TM = \bigsqcup_{p \in M} T_pM = \bigcup_{p \in M} \{((p, v_p) \mid v_p \in T_pM\}\,. \tag{F.2.7}$$

Definition F.2.8 — Vector field. A vector field on M is a global section of TM.

We denote the set of all smooth vectors fields on M by $\mathfrak{X}(M)$ or $\Gamma(TM)$.

Definition F.2.9 — Commutator. The commutator $[X, Y]$ it the vector field

$$[X, Y]\, f = X(Y(f)) - Y(X(f)) \tag{F.2.8}$$

Definition F.2.10 — Integral curve. Let X be a section of the tangent bundle TM of M. An integral curve γ is a map

$$\gamma \colon I \to M\gamma \colon t \mapsto \gamma(t) \tag{F.2.9}$$

such that for all $t \in I$ we have

$$T_{\gamma(t)}M \ni \gamma'(t) = X(\gamma(t)). \tag{F.2.10}$$

A flow is basically an infinite family of (maximal) integral curves associated to a smooth vector field.

Definition F.2.11 — Flow. Let $V \in \Gamma(TM)$ be a smooth vector field on a smooth manifold M. The flow generated by V is the map

$$\sigma \colon \mathbb{R} \times M \to M \tag{F.2.11}$$

such that for each $x \in M$ the $\sigma_t \colon \mathbb{R} \to M$ is a (maximal) integral curve of V that passes through $x \in M$ when $t = 0$.

Proposition F.2.12 For the maps $\sigma_t \colon M \to M$ of a flow σ on M we have that

$$\sigma_0 = id \tag{F.2.12}$$
$$\sigma_s \circ \sigma_t = \sigma_t \circ \sigma_s = \sigma_{s+t} \tag{F.2.13}$$
$$\sigma_s^{-1} = \sigma_{-s} \tag{F.2.14}$$

Proposition F.2.13 The maps $\sigma_t \colon M \to M$ of a flow σ on M are diffeomorphisms.

Definition F.2.14 — Oriented atlas. An oriented atlas is an atlas where all transition functions are orientation preserving (the Jacobian determinant is positive).

Definition F.2.15 — Oriented manifold. An oriented manifold is a manifold with an oriented atlas.

Definition F.2.16 — Volume form. A volume form on a manifold M is a nowhere vanishing section of $\bigwedge T^*M$.

Lemma F.2.17 — Orientable manifold. A manifold that admits a volume form is orientable.

F.3 Differential operations

Lie derivative is basically the derivative along a diffeomorphism induced by a flow.

Definition F.3.1 — Lie derivative. The Lie derivative $(\mathcal{L}_X T)_p$ of a tensor T with respect to a vector field X with associated flow $\sigma(t,p)$ and a point $p \in M$ of a manifold M is defined as

$$(\mathcal{L}_X T)_{x \in M} = \lim_{\epsilon \to 0} \frac{T|_x - (\sigma_\epsilon)_* \, T|_{\sigma_{-\epsilon}(x)}}{\epsilon} = \frac{T|_{\sigma_\epsilon(x)} - (\sigma_\epsilon)_* \, T|_x}{\epsilon}. \tag{F.3.1}$$

Proposition F.3.2 The Lie derivative of a tensor is a tensor of the same type. The Lie derivative also satisfies

$$\mathcal{L}_X aS + bT = a\mathcal{L}_X S + b\mathcal{L}_X T \tag{F.3.2}$$
$$\mathcal{L}_X S \otimes T = \mathcal{L}_X S \otimes T + S \otimes \mathcal{L}_X T \tag{F.3.3}$$

Definition F.3.3 — Killing vector field. A Killing vector field is a vector field X such that the Lie derivative of the metric with respect to X is zero. We write this as $\mathcal{L}_X g = 0$.

The flow of the Killing fields generate continuous isometries of the manifold.

Definition F.3.4 — Exterior derivative. Let M be a manifold and leg $\Lambda^p(M)$ be the vector space of p-form fields on M. The exterior derivative is a map

$$d \colon \Lambda^p(M) \to \Lambda^{p+1}(M) \tag{F.3.4}$$

1. Linear
2. Leiniz rule
3. idempotent
4. $d\colon f \mapsto df$ where $f \in C^\infty(M)$

Definition F.3.5 — Affine connection. Affine connection (covariant derivative) is a map

$$\nabla\colon \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}(M) \tag{F.3.5}$$

1. Linear
2. Leibniz

F.4 Symplectic manifolds

Symplectic manifolds are essential for defining the Hamiltonian framework in physics. We start by defining the symplectic structures for vector spaces before we put the same structure on the tangent space of a manifold. This section is based on [33].

Definition F.4.1 — Symplectic vector space. A symplectic vector space is vector space V where there is a map ω that satisfies

$$\omega\colon V \times V \to \mathbb{R} \tag{F.4.1}$$
$$\omega(v, v) = 0 \tag{F.4.2}$$
$$\forall (v \in V)\omega(u, v) = 0 \implies u = 0. \tag{F.4.3}$$

The two-form ω is called a symplectic form. The symplectic form induces isomorphisms between vectors and covectors.These isomorphisms are called musical isomorphism and we denote them as $\omega^\flat\colon V \to V^*$ and $\omega^\sharp\colon V^* \to V$.

Definition F.4.2 — Musical isomorphism (symplectic). The musical isomorphism (symplectic) are defined by

$$\omega^\flat\colon V \ni v \mapsto \omega(v, \,) \in V^* \tag{F.4.4}$$
$$\omega^\sharp = (\omega^\flat)^{-1} \tag{F.4.5}$$
$$\omega^\sharp\colon V^* \ni \omega(v, \,) \mapsto v \in V \tag{F.4.6}$$

The same concept that applies for an non-degenerate (0,2)-tensor, most frequently it is used when the tensor in question is the metric tensor (or inner product).

Proposition F.4.3 — Canonical basis. Given a symplectic vector space there is a basis $\{e_i\}$ such that

$$\omega = \sum_{i=1}^{n} e_i^* \wedge e_{i+n}^*. \tag{F.4.7}$$

The symbol e_i^* denotes the dual of e_i.

Definition F.4.4 — Symplectic manifold. A symplectic manifold (M, ω) is a manifold M with

a closed and non-degenerate two-form ω.

In the Hamiltonian framework any function defines a flow on the manifolds. The mechanics to define this flow are the Hamiltonian vector fields.

Theorem F.4.5 — Darboux's theorem. Let (M, ω) be a 2n-dimensional symplectic manifold. For any $p \in M$ there is a chart (U, φ) such that

$$\omega = \mathrm{d}x^1 \wedge \mathrm{d}x^{1+n} \ldots \mathrm{d}x^n \wedge \mathrm{d}x^{n+n} \,. \tag{F.4.8}$$

The coordinates of such a chart is called Darboux coordinates.

Definition F.4.6 — Hamiltonian vector field. Let (M, ω) be a symplectic manifold and let $f \in C^\infty(M)$. The Hamiltonian vector field generated by f is

$$X_f = -\omega^\sharp(df). \tag{F.4.9}$$

In Darboux coordinates we have

$$X_f = \frac{\partial f}{\partial x^{i+n}} \frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i} \frac{\partial}{\partial x^{i+n}}, \tag{F.4.10}$$

or in more common notation

$$X_f = \frac{\partial f}{\partial p_i} \frac{\partial}{\partial q^i} - \frac{\partial f}{\partial q^i} \frac{\partial}{\partial p_i}. \tag{F.4.11}$$

Some define the Hamiltonian vector field by the slightly more awkward statement: The Hamiltonian vector field corresponding to f is the vector field X_f such that

$$df = \omega(\ , X_f). \tag{F.4.12}$$

A Poisson algebra is an algebra with two multiplication operations that interacts in a special way.

Definition F.4.7 — Poisson algebra. A Poisson algebra is a triple $(\mathcal{P}, \cdot, \{\cdot,\cdot\})$ such that
1. $(\mathcal{P}, \cdot,)$ is an associative algebra
2. $(\mathcal{P}, \{\cdot,\cdot\})$ is a Lie algebra
3. $\{\cdot,\cdot\}$ is a derivation on $(\mathcal{P}, \cdot,)$, i.e. $\{a, b \cdot c\} = \{a, b\} \cdot c + b \cdot \{a, c\}$

Definition F.4.8 — Poisson structure. A Poisson structure on a manifold M is an extra binary operation $\{,\}$ on the commutative $\mathbb{R}$-algebra $C^\infty(M)$ such that $C^\infty(M)$ becomes a commutative Poisson algebra.

Any symplectic form ω defines a Poisson structure by

$$\{f, g\} = \omega^{\mu\nu} \frac{\partial f}{\partial x^\mu} \frac{\partial g}{\partial x^\nu}. \tag{F.4.13}$$

In Darboux coordinates this becomes

$$\{f, g\} = \frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial g}{\partial q^i} \frac{\partial f}{\partial p_i}. \tag{F.4.14}$$

Definition F.4.9 — Poisson manifold. A Poisson manifold is a manifold with a Poisson structure.

F.5 Bundles

Bundles are a generalization of the Cartesian product of two spaces. Bundles look locally like a product of two spaces, but they can have a very different structure globally. Seen in a slightly different light, bundles provide a concept of a family of spaces parameterized (in a continuous manner) by another space. We can also think of bundles as providing an arena for generalized fields on X, where the fields in question are not maps $\phi : X \to V$ into a single space V but instead are maps $\phi : X \to V_x$ into spaces V_x smoothly parameterized by elements $x \in X$. Such fields are called sections of bundles.

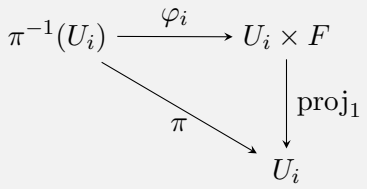
Definition F.5.1 — Bundle. A bundle ξ is a triple (E, B, π) where E and B are sets and π is a map $\pi : E \to B$ called the projection. The set E is called the total space and B is called the base space,

This definition of bundle is very broad with no restrictions on the spaces and maps involved in the definition. To be more useful one will usually restrict the spaces to be members of a certain category and the projection map to be an epimorphism within that category. A typical example would be the category of topological spaces with π a continuous surjection.

Definition F.5.2 — Fiber. Given a bundle $\xi = (E, B, \pi)$, and a point $x \in B$, the fiber at x, written F_x, is given by $F_x = \pi^{-1}(x)$.

For the fiber at x we obviously have $F_x \subset E$. In addition we see that, since the domain of π is E, $E = \bigcup_{x \in B} F_x$, which means that the total space is the union of all the fibers above the points of the base space. In a general bundle fibers at different points can be very different, but in many cases the fibers at different points are all homeomorphic to the same space F called the typical fiber, which leads us to the definition of a fiber bundle.

Definition F.5.3 — Fiber bundle. A fiber bundle ξ is a structure (E, B, π, F, G) where E, B and F are topological spaces, and π is a continuous map $\pi : E \to B$. For each $x \in B$ there is an open set U_i containing x and a homeomorphism $\varphi_i : \pi^{-1}(U_i) \to (U_i \times F)$ such that the following diagram commutes.



The set of all $\{U_i, \varphi_i\}$ is called the local trivialization of the bundle. For two overlapping open set $U_i \cap U_j \neq \emptyset$, both containing x, we have the corresponding two maps $\varphi_i : \pi^{-1}(U_i) \to (U_i \cap U_j \times F)$ and $\varphi_j : \pi^{-1}(U_j) \to (U_j \times F)$. These functions map a point p_x on the fiber F_x to points in $(U_j \times F)$ by $\varphi_i : p_x \mapsto (x, \psi_{ix}(p_x))$ and $\varphi_j : p_x \mapsto (x, \psi_{jx}(p_x))$. the maps $\psi_j \circ \psi_j^{-1} : F \to F$ are homeomorphisms of F. The set of all maps $\omega_{ijx} = \psi_j \circ \psi_j^{-1}$ form a group, called the structure group, and we demand that this group is isomorphic with G.

The bundle definition above should perhaps have been written $\xi = (E, B, \pi, F, G, \{U_i, \varphi_i\})$ but the local trivialization is often suppressed in the notation. To make the structure of the local trivialization clearer we could define the fibers over the open set U_i by $F_{U_i} = \bigcup_{x \in U_i} F_x$. We would then have $\varphi_i : F_{U_i} \to U_i \times F$. In other words the union of fibers of U_i can be expressed as a product of the open set of the base space with the typical fiber. An interesting generalization of the fiber bundles is the fibration. We will not go into detail about fibrations, we just mention that it is quite similar but now the fibres are only homotopically equivalent and not homeomorphic.

Definition F.5.4 — Bundle morphism. A bundle morphism between two bundles $\xi_1 = (E_1, B_1, \pi_1)$ and $\xi_2 = (E_2, B_2, \pi_2)$ is a pair of maps $(f_1), f_2)$ where $f_1 : E_1 \to E_2$ and $f_2 : B_1 \to B_2$ are both continuous functions, and the diagram below commutes.

$$\begin{array}{ccc} E_1 & \xrightarrow{f_1} & E_2 \\ \downarrow{\scriptstyle \pi_1} & & \downarrow{\scriptstyle \pi_2} \\ B_1 & \xrightarrow[f_2]{} & B_2 \end{array}$$

Definition F.5.5 — Local section. A local section ω on a fiber bundle $\xi = (E, B, \pi)$ is a map $\omega : U \to E$ defined on some open set $U \subset B$ such that $\pi \circ \omega = id_U$.

In other words a local section is map from $U \subseteq B$ such that at each point the map goes into the fiber above x.

Definition F.5.6 — Global section. A global section is a local section with $U = B$.

Not all bundles allow global sections. We denote the space of local sections over U by $C(U, E)$. The space of global sections is written as $\Gamma(E)$ or $\Gamma(B, E)$.

R The are two points to talking about a section of the tangent bundle as opposed to a vector field. One is that with a vector field on usually implies the existence of global canonical identification between "the fibers". In a section the map is an assignment of an element of a space at each point but there is no implied canonical identification. I.e. it is a "different" space at each point. (Even though they are usually isomorphic, there is no canonical isomorphism.) The second point is that sections cannot always be defined globally. When we talk about field a global definition is usually implied.

A vector bundle is a essentially a fiber bundle where the the fiber is a vector space. We also set some demands on the projection mappings.

Definition F.5.7 — Vector bundle. A vector bundle is a bundle $\xi = (E, B, \pi, F, G)$ where F is a vector space and G is a linear group.

Definition F.5.8 — Principal bundle. A principal bundle is a fiber bundle $\xi = (E, B, \pi, F, G)$ such that F is equal to the structure group G. The group G also defines a right action on E that preserves fibers.

F.6 Connections

The tangent spaces T_pM of a manifold M are all isomorphic. There is however no canonical isomorphism. This means there is no canonical way to compare vectors from different tangent spaces. A connection is an extra structure on a manifold that allows such a comparison. A connection also allows us to extend derivation on manifolds from scalar fields to vector and tensor fields.

Definition F.6.1 — Connection. $\xi = (E, B, \pi, F, G)$ be a smooth vector bundle. A connection on E is a linear map

$$\nabla \colon \Gamma(E) \to \Gamma(E \otimes T^*M) \tag{F.6.1}$$

such that

$$\nabla(f\sigma) = f\nabla(\sigma) + \sigma \otimes \mathrm{d}f \tag{F.6.2}$$

Note that for a connection could also be written as

$$\nabla \colon \Gamma(TM) \times \Gamma(E) \to \Gamma(E). \tag{F.6.3}$$

For the special case of $\Gamma(E) = \Gamma(TM)$ this becomes

$$\nabla \colon \Gamma(TM) \times \Gamma(TM) \to \Gamma(E). \tag{F.6.4}$$

This connects definition F.6.1 to the more well known picture of a connection mapping two vector fields (X, Y) to a new vector field $\nabla_X Y$. Various terms like covariant derivative and covariant derivative in the direction of X are used for the same or closely related objects.

F.7 Riemannian manifolds

Definition F.7.1 — Riemannian metric. A Riemannian metric g on a manifold M is a $(0, 2)$ tensor field on M such that for all points $p \in M$ with $U, V \in T_p(M)$ we have

$$g_p(U, V) = g_p(V, U) \tag{F.7.1}$$
$$g_p(U, U) \geq 0 \quad \text{with} \quad g_p(U, U) = 0 \implies U = 0. \tag{F.7.2}$$

Definition F.7.2 — Pseudo-Riemannian metric. A pseudo-Riemannian metric g on a manifold M is a $(0, 2)$ tensor field on M such that for all points $p \in M$ with $U, V \in T_p(M)$ we have

$$g_p(U, V) = g_p(V, U) \tag{F.7.3}$$
$$g_p(U, V) = 0 \implies U = 0. \tag{F.7.4}$$

Definition F.7.3 — Riemannian manifold. A (smooth) Riemannian manifold (M, g) is a (smooth) manifold M with a Riemannian metric g defined on it.

Definition F.7.4 — Pseudo-Riemannian manifold. A (smooth) Pseudo-Riemannian manifold (M, g) is a (smooth) manifold M with a Pseudo-Riemannian metric g defined on it.

We assume the regular definition of: signature of the metric, Euclidean metric, Lorentzian metric, spacelike timelike, null-like for vectors and curves.

Definition F.7.5 — Curvature. Let $\pi\colon E \to X$ be a vector bundle with connection ∇. The curvature of the connection the endomorphism valued 2-form

$$F^{\nabla}(v, w,)s = \nabla_v \nabla_w s - \nabla_w \nabla_v s + \nabla_{[v,w]} s. \tag{F.7.5}$$

F.8 Complex manifolds

In this section we generalize some notions of differential geometry to the complex case.

Definition F.8.1 — Holomorphic function. A holomorphic function is a map $f\colon \mathbb{C}^n \to \mathbb{C}$ such that for $f = f_1 + i f_2$ and $(x^\mu), (y^\mu \in \mathbb{C}^n$ we have

$$\frac{\partial f_1}{\partial x^\mu} = \frac{\partial f_2}{\partial y^\mu} \quad \frac{\partial f_2}{\partial x^\mu} = -\frac{\partial f_1}{\partial y^\mu}. \tag{F.8.1}$$

For a function $f\colon \mathbb{C}^n \to \mathbb{C}^m$ to be holomorphic we require all of the functions $(f^1, \dots, f^m)$ to be holomorphic as defined above.

Definition F.8.2 — Complex charts. Let M be a topological space. A complex chart (U, φ) on M is an open set $U \subset M$ and a map $\varphi\colon U \to \mathbb{C}^n$ such that φ is homeomorphism from U to an open set of $\mathbb{C}^n$ for some n.

Definition F.8.3 — Holomorphic atlas. labeldef:Holomorphic atlas A holomorphic atlas for a topological space M is a family of charts (U_i, φ_i) is a collection of complex charts of M such that U_i cover M and the transition maps are holomorphic.

Definition F.8.4 — Complex manifold. A Complex manifold is a tuple $(M, (U_i, \varphi_i))$ such that M is a connected Hausdorff topological space and (U_i, φ_i) is a maximal holomorphic atlas.

Definition F.8.5 — Riemann surface. A Riemann surface is one-dimensional complex manifold.

Definition F.8.6 — Holomorphic function. A function $f\colon M \to N$ between two complex manifolds M and N is a called a holomorphic function (between manifolds) if for each chart the induces map is holomorphic.

Definition F.8.7 — Biholomorphic. Two complex manifolds are said to be biholomorphic if there is a holomorphic map between them which has a holomorphic inverse.

Biholomorphic is the isomorphism between complex manifolds.

F.9 Various topics

In this section we collect various topic that we use in the text.

Definition F.9.1 — Hypersurface. Let M be an n-dimensional manifold. A hypersurface in M is a (n-1)-dimensional submanifold of M.

The difference in dimensions between a manifold and a submanifold is sometimes called the codimension . A hypersurface is a submanifold of codimension one.

Definition F.9.2 — Causal curve. A causal curve is a curve that is always timelike or spacelike.

Definition F.9.3 — Cauchy surface. A Cauchy surface C of a spacetime M is a spatial hypersurface such that all causal curves of M intersect C once and only once.

Definition F.9.4 — Globally hyperbolic. A spacetime is globally hyperbolic if it has a Cauchy surface.

Definition F.9.5 — Embedding. Let X and Y be topological spaces. An embedding of X into Y is map $\phi\colon X \to Y$ such that X and $\phi(X)$ (with the subspace/relative topology) are homeomorphic.

More generally an embedding is an injective structure preserving map.

Definition F.9.6 — Immersion. Let X and Y be topological spaces. An immersion of X into Y is map $\phi\colon X \to Y$ such that for all $x \in X$ there is a neighborhood O of x such that $\phi|_O : X \to Y$ (or $\phi\colon O \to Y$) is an embedding.

An immersion need not be injective. (fix this)

Definition F.9.7 — Foliation. Let M be an (m+n)-dimensional manifold. A codimension p foliation of M is a partition of M into q-dimensional manifolds (leaves) such that locally M is homeomorphic to $\mathbb{R}^p \times \mathbb{R}^q$.

The subsets defined by a foliation are called leaves or slices. We will also use the word spatial slice for slice of a Lorentzian manifold that is everywhere spacelike.

F.10 Notes

The basic references for this appendix are the books by Martin [90], Nakahara [118] , Spivak [421] and Lee [422].

G. Lie groups and Lie algebras

G.1 Lie groups and Lie algebras

A Lie group combines the properties of a manifold and a group into one. This produces a rich theory with an abundance of applications within mathematics and physics. Lie groups can in a sense be considered to be the simplest examples of manifolds. The structure of a Lie group is almost completely determined by its linear structure close to the identity. The theory of Lie algebras provides a way to analyze and use this structure.

Definition G.1.1 — Lie group. A (real) Lie group is a group that is also a finite-dimensional real smooth manifold where the operations of multiplications and taking inverses are smooth.

To every Lie group there is an associated Lie algebra. Lie algebras can also be defined abstractly without reference to the Lie group. We first provide the abstract theory before going into the connection with Lie groups.

Definition G.1.2 — Lie algebra. A Lie algebra is a vectors space V of a field F with an additional multiplication operation (often denoted by brackets) that satisfies

$$[ax + by, z] = a[x, z] + b[y, z] \tag{G.1.1}$$

$$[x, y] = -[y, x] \tag{G.1.2}$$

$$[x, [y, z]] + [x, [y, z]] + [x, [y, z]] = 0 \tag{G.1.3}$$

Proposition G.1.3 Let A be an associative algebra. On A we define the bracket operation in terms of the associative multiplication with

$$[a, b] = ab - ba. \tag{G.1.4}$$

We denote the algebra with this new structure by A_L. The algebra A_L is a Lie algebra.

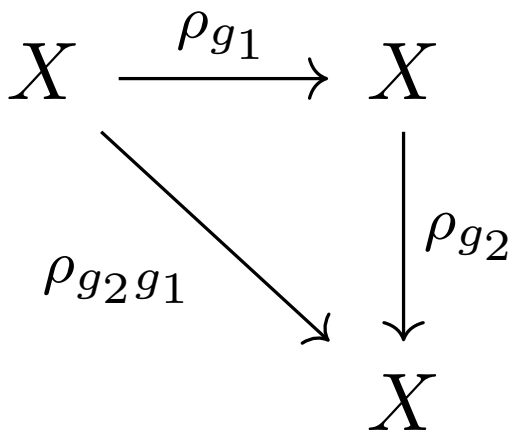


Figure G.1: Group action commutative diagram.

Definition G.1.4 — Enveloping algebra. Enveloping algebra Universal enveloping algebra

Definition G.1.5 — Generators. Generators

Definition G.1.6 — Casimir element. Casimir element

G.2 Representation theory

Definition G.2.1 — Lie algebra representation. Let V be vector space over the field F, and let A be a Lie algebra. A Lie algebra representation is a Lie algebra homomorphism $\rho : A \to End_L(V)$.

Definition G.2.2 — Group action. Let G be a group and X a set. A (left) group action of G on X is a group morphism $G \to S(X)$ into the permutation group of X. Or alternatively a map $\rho : G \times X \to X : (g, x) \mapsto g.x$ such that

1. For all $x \in X$ we have $e.x = x$.
2. For all $g, h \in G$ we have $(gh).x = g.(h.x)$.

The corresponding definitions for right group actions are obvious and we consider them to be defined by their left action counterparts. If we write the action corresponding to $g \in G$ as ρ_g the criteria can be written as $\rho_{g1} \circ \rho_{g2} = \rho_{g1g2}$. This corresponds to saying that the diagram in figure **??** commutes. If G is a Lie group and X is a smooth manifold one usually requires that ρ_g is a diffeomorphism.

Theorem G.2.3 Let ρ be a group action of G on X. The following statements are true

$$\rho_e = Id_X \tag{G.2.1}$$

$$\rho_{g^{-1}} = \rho_g^{-1} \tag{G.2.2}$$

$$\tag{G.2.3}$$

Definition G.2.4 — G-module. A (left) G-module is an Abelian group A, with a group G that has a left group action $\rho : G \times A \to : (g, a) \mapsto (g.a)$ such that $g.(a + b) = g.a + g.b$

Definition G.2.5 — Orbit of an element. The orbit of an element $x \in X$ is the image of x by the all the elements of G. We write $\mathrm{Orb_x} = G.x = \{gx \mid g \in G\} \subset X$.

Two points of X being in the same orbit defines an equivalence relation on X. As for any equivalence relation this defines a partition of X. The set of all orbits of X under the action of G is called the orbit space X/G.

Definition G.2.6 — Stabilizer of an element. The stabilizer of an element $x \in X$ is the set of all $g \in G$ such that g maps x to x. We write $\mathrm{Stab_x} = \{g \in G \mid g.x = x\} \subset G$

Definition G.2.7 — Transitive action. A group action G on X is transitive iff

$$(\forall x, y)\,(\exists g)\; g.y = x \tag{G.2.4}$$

In a transitive action each point has only one orbit and all points have the same orbit. The one and only orbit is equal to the whole space X, and every point can be reach by every other point by the group action.

Definition G.2.8 — Free action. A group action G on X is free iff

$$g.x = x \iff g = e \tag{G.2.5}$$

In a free action there are no non-trivial fixed points.

G.3 Harmonic analysis

Definition G.3.1 — Left translate. Let G be locally compact Hausdorff topological group, and let S be a subset of G. The left translate of S is give by

$$gS \coloneqq \{gs \mid s \in S\} \tag{G.3.1}$$

The right translate is defined correspondingly.

Definition G.3.2 — Left-invariant measure. Let μ be a measure on the Borel sets of G. Such a measure if called a left-invariant measure if for all Borel set of G and for all $g \in G$ we have that $\mu(S) = \mu(gS)$.

Theorem G.3.3 — Haar's theorem. Let G be locally compact Hausdorff topological group. There is a unique measure on G satisfying the following properties:

1. The measure is left-invariant.
2. The measure is finite on every compact set.
3. The measure is outer regular.
4. The measure is inner regular.

This measure on a group G is called the de Haar measure and denote by μ_H.

Theorem G.3.4 — Peter-Weyl theorem. The set of maps from G to matrix coefficients of representations of G is dense in the set $C^0(G, \mathbb{C})$.

G.4 Specific groups and algebras

Definition G.4.1 — GL(n,C). The group GL(n,C) is defined as

$$\mathrm{GL}(n, C) = \{\Lambda \mid \Lambda \in \mathrm{M}_n(\mathbb{C}) \wedge \det(\Lambda) \neq 0\} \tag{G.4.1}$$

Definition G.4.2 — SU(n). The group SU(n) is defined as

$$\mathrm{SU}(n) = \{\Lambda \mid \Lambda \in \mathrm{GL}(n, C) \wedge \Lambda\Lambda^* = \Lambda^*\Lambda = \mathbb{I} \wedge \det(\Lambda) = 1\} \tag{G.4.2}$$

Definition G.4.3 — su(n). The Lie algebra $\mathfrak{su}(n)$ is defined as

$$\mathfrak{su}(n) = \{\Lambda \mid \Lambda \in \mathrm{M}_n(\mathbb{C}) \wedge \Lambda^* = -\Lambda \wedge \mathrm{tr}(\Lambda) = 0\} \tag{G.4.3}$$

For SU(2) we can give an explicit form of the matrices by

$$\mathrm{SU}(2) = \left\{ \begin{pmatrix} \alpha & -\overline{\beta} \\ \beta & \overline{\alpha} \end{pmatrix} : \quad \alpha, \beta \in \mathbf{C}, |\alpha|^2 + |\beta|^2 = 1 \right\} \tag{G.4.4}$$

The Pauli matrices are defined as

$$\sigma_1 = \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{G.4.5}$$

$$\sigma_2 = \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{G.4.6}$$

$$\sigma_3 = \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{G.4.7}$$

By using the Pauli matrices we can express the Lie algebra su(2) as the real algebra given by

$$\mathfrak{su}(2) = \mathrm{span}\,\{i\sigma_1, i\sigma_2, i\sigma_3\} \tag{G.4.8}$$

For NCG we shall often need $i\mathfrak{su}(2)$. We express this as

$$i\mathfrak{su}(2) = \mathrm{span}\,\{\sigma_1, \sigma_2, \sigma_3\} \tag{G.4.9}$$

We define the algebra of quaternions $\mathbb{H}$ the real algebra given by

$$\mathbb{H} = \mathrm{span}\,\{\mathbb{I}, i\sigma_1, i\sigma_2, i\sigma_3\} \tag{G.4.10}$$

G.5 Notes

The basic references for this appendix are de Kerf [423][424], Cornwell [246] and Fuchs [425].

H. Functional analysis

Functional analysis and operator algebras involves analysis and linear algebra over infinite dimensional spaces. In such cases topological considerations are important. In physics infinite dimensional linear algebra is often treated just like if it was finite dimensional.

H.1 Basic concepts

A metric captures the minimal set of properties that a distance function must have for the distance concept to make sense. Metrics can be defined on a space without any additional structures.

Definition H.1.1 — Metric. Let X be a set, and let $x, y, z \in X$. A metric on X is a map $d\colon X \times X \to \mathbb{R}$ satisfying

1. $d(x, y) = d(y, x)$ (symmetric)
2. $d(x, y) \geq 0$ (non-negative)
3. $d(x, y) = 0 \iff x = y$ (non-degenerate)
4. $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality, subadditivity)

Definition H.1.2 — Metric space. A metric space is a pair (M, d) where M is a set and d is a metric on M.

A norm defines a concept of item length. It is only defined for vector spaces (modules). The definition of a norm specifies how it interacts or relates to vector addition and scalar multiplication.

Definition H.1.3 — Norm. Let V be a vector space over the field F, where F is either $\mathbb{R}$ or $\mathbb{C}$, and let $v, w \in V$ and $\alpha \in F$. A norm on V is a map $\|\| : V \to \mathbb{R}$ satisfying

1. $\|v + w\| \leq \|v\| + \|w\|$ (triangle inequality, subadditivity)
2. $\|\alpha v\| = |\alpha|\|v\|$ (scaling)

3. $\|v\| = 0$ iff $v = 0$ (non-degenerate)

Note that $|\alpha|$ is a real number such that $|\alpha|\|v\|$ is defined. The norm is not defined for vector spaces over more general fields than $\mathbb{R}$ or $\mathbb{C}$.

Definition H.1.4 — Normed vector space. A normed space is a pair $(V, \|\|\|)$ where V is a real/complex vector space and $\|\|\|$ is a norm on V.

For operator spaces the concept of a seminorm is important.

Definition H.1.5 — Seminorm. A seminorm is map on a vector space that satisfies only property 1 and 2 of definition H.1.3.

Definition H.1.6 — Inner product. Let V be a vector space over the field F, where F is either $\mathbb{R}$ or $\mathbb{C}$, and let $v, w \in V$ and $\alpha \in F$. A inner product on V is a map $(\ ,\ ) \colon V \times V \to F$ satisfying

1. $(v, w) = \overline{(w, v)}$
2. $(\alpha v, w) = \alpha(v, w)$
3. $(v, v) \geq 0$
4. $(v, v) = 0 \iff v = 0$

Note that for $F = \mathbb{R}$ we define $\overline{x} = x$. It follows from property 1 that $(v, v) = \overline{(v, v)}$, and thus (v, v) is a real number and therefore the inequality in property 3 is well defined even for complex inner product spaces.

Definition H.1.7 — Inner product space. A inner product space is a pair $(V, \|\|\|)$ where V is a real/complex vector space and $\|\|\|$ is an inner product on V.

H.2 Sequences and completeness

Complete spaces are central for many areas of mathematics. We refer to [105, section 1.10] for proofs of the statements in this section.

Definition H.2.1 — Sequence. Let (M, d) be a metric space. A sequence s is a map

$$s \colon \mathbb{N} \to M \tag{H.2.1}$$

$$s \colon n \mapsto s_n \tag{H.2.2}$$

We could define convergence of sequences using only the topological structure but we will only need the metric version.

Definition H.2.2 — Convergent sequence. A sequence s_n on a metric space (M, d) is called a convergent sequence with limit L, if there for any real number $\epsilon > 0$ there is a natural number N such that for all $n > N$ we have that $d(s_n, L) < \epsilon$.

We say that the sequence s_n is convergent and converges to L.

Definition H.2.3 — Cauchy sequence. A Cauchy sequence is a sequence $\{x_i\}$ of elements in a metric space with metric d, such that for any real number $\epsilon > 0$ there exists an integer N such that $(\forall m, n > N)\ d(x_m, x_n) < \epsilon$.

Note that the Cauchy criterion is stronger than just requiring $(\forall m > N)\, d(x_m, x_{m-1}) < \epsilon$. Every convergent sequence is a Cauchy sequence but in a general metric space a Cauchy sequence is not necessarily convergent. We now define spaces where Cauchy sequences are always convergent.

Definition H.2.4 — Complete metric space. A complete metric space is a metric space where every Cauchy sequence converges.

If we have an "incomplete" space where some Cauchy sequences does not converge, we can extend this space (in a unique way) so that it becomes complete.

Definition H.2.5 — Completion. The completion of a metric space M is the space $\overline{M}$. The space $\overline{M}$ is constructed as a set of equivalence classes of Cauchy sequences in M.

The space M is a dense subset of $\overline{M}$. The dense embedding $M \to \overline{M}$ is provided by mapping $m \in M$ to equivalence classes of constant Cauchy sequences $s\colon n \mapsto m$. We used this concept when we completed the rational numbers to get the real numbers in appendix A.

H.3 Infinite dimensional vector spaces

In this section we define various vector spaces that are complete with respect to a metric induces by inner products or norms. The definitions in this section also apply to finite dimensional vector spaces, but they are usually automatically true for such spaces, and they are therefore most interesting for infinite dimensional vector spaces.

Definition H.3.1 — Banach space. A real/complex Banach space is a real/complex vector space with a norm, such that the space is complete with respect to the metric induced by the norm.

The set $C(X)$ of continuous complex valued functions on a compact space X is a Banach algebra with the supremum norm. The set $B(X)$ of bounded functions $X \to \mathbb{C}$ is a Banach space with the supremum norm. All finite dimensional real/complex vector space with a norm are Banach spaces.

Definition H.3.2 — Hilbert space. A Hilbert space is a real/complex vector space with an inner product, such that the space is complete with respect to the metric induced by the inner product.

Any finite dimensional real or complex inner product space is a Hilbert space. Banach spaces are more general than Hilbert spaces in the sense that all Hilbert space are Banach spaces. Banach spaces are the foundation for defining $W^\star$- and $C^\star$-algebras.

We can combine norms with algebras to give normed algebras.

Definition H.3.3 — Normed algebra. A normed algebra is an algebra A with a norm such that for all $a, b \in A$ we have

$$\|ab\| \leq \|a\|\|b\|. \tag{H.3.1}$$

Banach algebras are complete normed algebras.

Definition H.3.4 — Banach algebra. A Banach algebra is a real/complex Banach space A that is also an associative algebra. The algebra multiplication is continuous with respect to the topology induced by the norm in the sense that the following is satisfied

$$\forall a, b \in A \quad \|ab\| \leq \|a\|\,\|b\| \tag{H.3.2}$$

Structure	Corresponding complete structure
normed vector spaces	Banach space
normed algebra	Banach algebra
inner product vector space	Hilbert space
inner product algebra	Hilbert algebra
*-algebra	C*-algebra

Table H.1: Algebraic topological spaces

H.4 Operators and norms

It is well known from quantum mechanics and mathematics that the set of bounded operators on a Hilbert space is a very important subset of the set of all operators. We will define the concept of a bounded operator in a slightly more general setting.

Definition H.4.1 — Bounded operator. Let T be a linear operator between two normed vector spaces X and Y. The operator $T : X \to Y$ is called a bounded operators if there exists a real number $M > 0$ such that

$$\|Tv\|_Y = M\|v\|_X. \tag{H.4.1}$$

All linear operators between finite dimensional vector spaces are bounded. The set of bounded operators on a Hilbert space $\mathcal{H}$ is denoted $B(\mathcal{H})$. The space $B(\mathcal{H})$ is a Banach space with respect to the operator norm. The operator norm of a bounded operator L is the smallest positive real number M that satisfies equation H.4.1.

Definition H.4.2 — Operator norm. Let $B(X, Y)$ be the set of bounded linear operator between two normed vector spaces X and Y. Let $T \in B/X, Y)$. The operator norm on $B(X, Y)$ is given by the following equivalent expressions

$$\begin{aligned}
\|T\|_{op} &= \inf\{M \geq 0 : \|Tv\|_Y \leq M\|v\|_X \text{ for all } v \in V\} \\
&= \sup\{\|Tv\|_Y : v \in V \text{ with } \|v\|_X \leq 1\} \\
&= \sup\{\|Tv\|_Y : v \in V \text{ with } \|v\|_X = 1\} \\
&= \sup\left\{\frac{\|Tv\|_Y}{\|v\|_X} : v \in V \text{ with } v \neq 0\right\}.
\end{aligned}$$

Note that the operator norm is on the space of operators $B(X, Y)$ but depends on (or is defined by) the norms on X and Y.

Definition H.4.3 — Positive operator. A positive operator is an operator $T\colon \mathcal{H} \to \mathcal{H}$ such that

$$\langle u|Tu\rangle \geq 0 \quad \forall u \in \mathcal{H}. \tag{H.4.2}$$

Definition H.4.4 — Adjoint of an operator. Let V be a real/complex Hilbert space with inner product $(\ ,\ )$, and let $u, v \in V$. The adjoint of a bounded linear operator $T\colon V \to V$ is the

bounded linear operator $T^*\colon V \to V$ such that

$$(T^*u, v) = (u, Tv). \tag{H.4.3}$$

The uniqueness of T^* follows from the Riez representation theorem.

Definition H.4.5 — Self-adjoint operator. Let V be a real/complex Hilbert space. A self-adjoint operator T is a bounded linear operator $T\colon V \to V$ such that $T^* = T$.

Definition H.4.6 — Unitary operator. Let V be a real/complex Hilbert space with inner product $(\ ,\ )$, and let $v, w \in V$. A unitary operator $U\colon V \to V$ is an operator such that

$$\langle Uv|Uw\rangle = \langle v|w\rangle \tag{H.4.4}$$

|

Definition H.4.7 — Projection operator. Let V be a vector space. A projection operator on V is an operator that idempotent. That is and operator P such that $P^2 = P$.

Definition H.4.8 — Orthogonal projection operator. Let V be a complete inner product vector space. An orthogonal projection operator on V is an operator that idempotent and self-adjoint. That is and operator P such that $P^2 = P$ and $P^* = P$.

An orthogonal projection operators is also called a Hermitian projection operator.

Definition H.4.9 — Resolvent set. The resolvent set $\rho(a)$ of an element $a \in A$ of a complex *-algebra A with unit $1 \in A$ is the set of complex numbers $\lambda \subset \mathbb{C}$ such that the element $(a - \lambda 1) \in A$ is invertible.

Definition H.4.10 — Resolvent. The resolvent is the map $R(\lambda, a) : \rho(a) \to (a - \lambda 1)^{-1}$.

An invertible element of an algebra A is an element a such that there is an element $b \in A$ and $ab = ba = 1$, where $1 \in A$ is the multiplicative unit element of A.

Definition H.4.11 — Spectrum of a C*-algebra. The spectrum $\sigma(a) \subset \mathbb{C}$ of an element $a \in A$ of a complex *-algebra A with unit $1 \in A$ is the set of complex numbers $\lambda \in \mathbb{C}$ such that the element $(a - \lambda 1) \in A$ is non-invertible (does not have a multiplicative inverse).

Corollary H.4.12 $\sigma(a) = \mathbb{C} \setminus \rho(a)$

Definition H.4.13 — Spectral radius. The spectral radius $\nu(a)$ of an element a of a complex *-algebra with unit is defined as

$$\nu(a) \coloneqq \sup\{|\lambda| \mid \lambda \in \sigma(a)\} \tag{H.4.5}$$

H.5 Algebras with involutions

The start operation is a generalization of complex conjugation. It is an operation that is both an involution and an anti-automorphism.

Definition H.5.1 — Involution. An involution on the set X is a map $f : X \to X$ that is its own inverse. That is

$$Id_X = f \circ f : X \to X \tag{H.5.1}$$

Any involution is (obviously) a bijection. Simple examples of involutions are (informally described) the identity map, multiplying by -1, taking the reciprocal, complex conjugation and taking the transpose or the conjugate transpose.

Definition H.5.2 — Anti-automorphism. Let S be a set with a binary operation denoted as multiplication. An anti-automorphism is a bijection of S to itself that reverses the order of multiplication.

Definition H.5.3 — $*$-ring. A $*$-ring is a ring R with a map that is an involution and an anti-automorphism, $* : R \to R$ defined by $* : r \mapsto r*$.

$$(a+b)^* = a^* + b^* \tag{H.5.2}$$
$$(ab)^* = b^* a^* \tag{H.5.3}$$
$$(x^\star)^\star = x \tag{H.5.4}$$
$$1^\star = 1 \quad \text{(for unital rings)} \tag{H.5.5}$$
$$\tag{H.5.6}$$

Definition H.5.4 — $*$-algebra. A $*$-algebra (over a ring R) is an involutive R-algebra A. This is a complex algebra with an involution over a $*$-ring R, that is compatible with all the operations in the algebra in the sense that

$$(a+b)^* = a^* + b^* \tag{H.5.7}$$
$$(ab)^* = b^* a^* \tag{H.5.8}$$
$$(\lambda a)^* = \bar{\lambda} a^* \tag{H.5.9}$$
$$(x^\star)^\star = x \tag{H.5.10}$$
$$1^\star = 1 \quad \text{(for unital algebras)} \tag{H.5.11}$$

Norms can be combined with the involution operation. We then get a normed *-algebra.

Definition H.5.5 — Normed *-algebra. A normed *-algebra is a *-algebra with a norm that satisfies:

$$\|a^\star\| = \|a\| \text{ (the isometric property).} \tag{H.5.12}$$

An algebra that satisfies definitions H.5.4 and H.5.5, as well as being complete, is called a Banach *-algebra.

Definition H.5.6 — Banach *-algebra. A Banach *-algebra is a Banach algebra that is also a normed *-algebra.

Definition H.5.7 — *-morphism. A *-morphism is a map $\phi : A \to B$ between two *-algebras

A and B such that

$$\phi(a_1 + a_2) = \phi(a_1) + \phi(a_2) \tag{H.5.13}$$
$$\phi(a_1 \cdot a_2) = \phi(a_1) \cdot \phi(a_2) \tag{H.5.14}$$
$$\phi(a_1^*) = \phi(a_1)^* \tag{H.5.15}$$

Definition H.5.8 — *-automorphism. A *-automorphism is a map $\phi : A \to A$ where A is a *-algebras such that

$$\phi(a_1 + a_2) = \phi(a_1) + \phi(a_2) \tag{H.5.16}$$
$$\phi(a_1 \cdot a_2) = \phi(a_1) \cdot \phi(a_2) \tag{H.5.17}$$
$$\phi(a_1^*) = \phi(a_1)^* . \tag{H.5.18}$$

Definition H.5.9 — Unit elements. The unit elements $\mathrm{Inv}(A)$ of a ring A are the multiplicatively invertible elements of a unital ring. That is all $u \in A$ such that u^{-1} exists and $uu^{-1} = u^{-1}u = 1$.

Corollary H.5.10 The set $\mathrm{Inv}(A) = \{u \in A | uu^{-1} = u^{-1}u = 1\}$ is a group under multiplication and is called the *group of units*.

Definition H.5.11 — Unitary elements. Unitary elements of a unital algebra with involution $u \in A$ $uu^* = u^*u = 1$. The set $\mathcal{U}(A) = \{u \in A | uu^* = u^*u = 1\}$ is a group under multiplication and is called the *group of unitaries*.

Definition H.5.12 — Self-adjoint elements. A self-adjoint element of an algebra $\mathcal{A}$ with involution is an element $a \in \mathcal{A}$ such that

$$a^\star = a \tag{H.5.19}$$

Definition H.5.13 — C*-algebra. A C*-algebra is a Banach *-algebra where the the following is satisfied

$$\|x^*x\| = \|x\|^2 . \tag{H.5.20}$$

The C*-algebra criteria is very strong. As an example of this we mention that in a C*-algebra the norm is uniquely determined by the algebraic structure.

■ **Example H.5.14 — Commutative C*-algebra.** The algebra of complex valued function on a compact Hausdorff topological space with the supremum norm and with involution as complex conjugation, is a commutative C*-algebra. ■

■ **Example H.5.15 — Noncommutative C*-algebra.** Let $\mathcal{H}$ be a complex Hilbert space and let $B(\mathcal{H})$ be the set of bounded operator on $\mathcal{H}$. Let addition and scalar multiplication on $B(\mathcal{H})$ be defined by pointwise addition and scalar multiplication, and binary multiplication by composition. The Hermitian adjoint operation is an involution on this algebra. The algebra $B(\mathcal{H})$ with the operator norm (derived from the inner product norm on $\mathcal{H}$) is a C*-algebra. If $\mathcal{H}$ is a finite dimensional vector space of dimension n, then $B(\mathcal{H}) = M_n(\mathbb{C})$ is a C*-algebra. Any subalgebra

of $B(\mathcal{H})$, which is a closed set in the norm topology (on $B(\mathcal{H})$), is also a C*-algebra. ∎

Proposition H.5.16 The C* criteria implies the *-norm criteria. That is we have the following implication

$$\|x^*x\| = \|x\|^2 \implies \|x^*\| = \|x\|\,. \tag{H.5.21}$$

We list three equivalent definitions of a von Neumann algebra since they are all important. The first two define as a specific type of sub-algebra of the algebra of bounded operators. The third definition is purely abstract without any reference to a specific realization.

Definition H.5.17 — Commutant. The commutant (or the centralizer) S' of a subset S of a group G is the set of all element of G that commutes with all elements of S.

The double commutant is the commutant of the commutant (still with respect to the original full set G).

Definition H.5.18 — von Neumann algebra (1). A von Neumann algebra $\mathcal{A}$ is a subalgebra of the algebra of bounded operators $B(\mathcal{H})$ on a Hilbert space $\mathcal{H}$ that satisfies the following

1. The algebra $\mathcal{A}$ is a *-algebra.
2. The set $\mathcal{A} \subseteq B(\mathcal{H})$ is a closed set in the weak operator topology.
3. The algebra $\mathcal{A}$ contains the identity.

Definition H.5.19 — von Neumann algebra (2). A von Neumann algebra $\mathcal{A}$ is a subalgebra of the algebra of bounded operators $B(\mathcal{H})$ on a Hilbert space $\mathcal{H}$ that satisfies the following

1. The algebra $\mathcal{A}$ is closed under the * operation.
2. The algebra $\mathcal{A}$ is is equal to its double commutant.

Definition H.5.20 — von Neumann algebra (3). A von Neumann algebra $\mathcal{A}$ is an algebra that has a predual.

Definition H.5.21 — Linear functional. A Linear functional ω is a K-linear map from a K-vector space to K.

Definition H.5.22 — Ordered vector space. An ordered vector space is a pair $(V, \leq)$ such that V is a vector space and $\leq$ is a partial order on V. In addition the following must hold:

1. For all z if $x \leq y$ then $x + z \leq y + z$
2. For all $\lambda \geq 0$ if $x \leq y$ then$\lambda x \leq \lambda y$

Definition H.5.23 — Positive linear functional. A positive linear functional is a linear functional ω on and ordered vector space $(V, \leq)$ such that

$$(\forall v \geq 0)\omega(v) \geq 0 \tag{H.5.22}$$

Definition H.5.24 — State. A state on an algebra is positive linear functional ω with $\omega(I) = 1$.

Definition H.5.25 — Faithful state. A state ω is a faithful state if for all $A \in \mathcal{A}$ if $0 \leq A$ and $\omega(A) = 0$ then $A = 0$.

Definition H.5.26 — Normal state. A states is called normal if for every increasing sequence $A_1, A_2, A_3, ...$ in $\mathcal{A}$ with strong limit equal to A then we have $lim_{n\to\infty}(A_n) = A$

Definition H.5.27 — Cyclic vector. Let $\mathcal{A}$ be a unital *-algebra and $\mathcal{H}$ a Hilbert space, and let $\pi : \mathcal{A} \to B(\mathcal{H})$ be *-algebra morphism. A vector ψ is cyclic (for π) if

$$cl\left\{\pi(a)\psi \mid a \in \mathcal{A}\right\}, \tag{H.5.23}$$

where cl means the closure with respect to the inner-product (norm) topology of $\mathcal{H}$. This is equivalent to requiring that

$$\{\pi(a)\psi \mid a \in \mathcal{A}\} \tag{H.5.24}$$

be a dense set in $\mathcal{H}$.

The GNS construction assigns to any C*-algebra $\mathcal{A}$ and state ω a Hilbert space $\mathcal{H}_\omega$ and a representation π_ω. We state it in the theorem below. The GNS construction is what allows us to map the abstract purely algebraic formulation presented in chapter 2 to the common formulation using Hilbert spaces and operator algebras. More precisely, it is the Gelfand-Naimark theorem that allows such a mapping, but this theorem uses (multiple copies of) the GNS construction to set up a map that is not just a representation but is actually an isomorphism.

Theorem H.5.28 — GNS theorem. Let $\mathcal{A}$ be a unital C*-algebra and ω a state on $\mathcal{A}$, then there is a Hilbert space H_ω and a representation (*-algebra morphism) $\pi_\omega : \mathcal{A} \to B(\mathcal{H}_\omega)$ determined by ω. Let $(\ ,\ )$ be the inner product on H_ω and let $A \in \mathcal{A}$. The representation has the following properties

1. There exists a cyclic vector $\psi_\omega \in \mathcal{H}_\omega$.
2. $\omega(A) = (\psi_\omega, \pi_\omega(A)\psi_\omega)$
3. Every other representation π with a Hilbert space $\mathcal{H}_\pi$, and with a cyclic vector $\psi_\pi \in \mathcal{H}_\pi$, and such that $\omega(A) = (\psi_\pi, \pi(A)\psi_\pi)$, is unitarily equivalent to π_ω.

Proof. See [13] for a proof. ■

Theorem H.5.29 — Gelfand-Naimark theorem. A C*-algebra is isomorphic to an algebra of bounded operators on a Hilbert space.

Proof. The proof uses a set of separating states to create a faithfull representations as a direct product of corresponding GNS representations. See [13] for a full proof. ■

H.6 Differential operators

We will not treat abstract differential operators in any detail. Still, much of the treatment of NCG is based on Laplace type and Dirac type operators, and it is therefore relevant with some definitions specifying the proper context for such operators.

Definition H.6.1 — Differential operator (1). A differential operator is a map between to function spaces that can be expressed as a sum of derivatives on the first function space.

Definition H.6.2 — Differential operator (2). A differential operator is a map between the spaces of smooth sections of two bundles that can be expressed as a sum of derivatives on the first bundle.

H.7 Various results

Important operators in QM are unbounded. Unbounded operators pose delicate domain issues as they are not defined on the whole Hilbert space. The method of rigged Hilbert spaces using a Gelfand triple was invented to handle this and related issues.

Definition H.7.1 — Gelfand triple. A Gelfand triple is a sequence of three set with two injective maps

$$B \to H \to B^* \tag{H.7.1}$$

The set B is a dense subspace of H, and B^* is the dual (as a topological vector space) of B.

Gelfand triple are also know as rigged Hilbert space. Rigged in this context means something like "fully equipped".

■ **Example H.7.2 — Name.** The classic example of a Gelfand triple is given by the Hilbert space of one-dimensional QM. With $\mathcal{H} = L^2(\mathbb{R})$ and $B = S(\mathbb{R})$ the Schwartz space, and $B^* = S'(\mathbb{R})$ the space of tempered distributions. ■

Theorem H.7.3 — Weierstrass extreme value theorem. A continuous function from a non-empty compact space to a subset of the real numbers attains a maximum and a minimum.

Definition H.7.4 — Supremum norm. The supremum norm for a real or complex bounded functions on a domain S the supremum norm (or uniform norm, or Chebyshev norm, or infinity norm) is given by for a function f as

$$\|f\|_\infty = \|f\|_{\infty,S} \coloneqq \sup\{|f| \mid x \in S\} \tag{H.7.2}$$

H.8 KK-theory

We will not go into the depths of KK-theory, but we need several concepts from this area.

Definition H.8.1 — A-module. Given a C*-algebra A, an A-module is a module with scalars in A (in the same sense as for a ring R and an R-module).

Definition H.8.2 — (A,B)-bimodule. An (A, B)-bimodule is a left A-module E, that is also a right B-module, and where left and right multiplication commute.

$$a \cdot (e \cdot b) = (a \cdot e) \cdot b \tag{H.8.1}$$

As a special case an A-bimodule will denote an (A, A)-bimodule .

Definition H.8.3 — Inner product on (A, B)-bimodule. An inner product (A, B) -bimodule is an (A,B)-bimodule with a inner product that takes values in B.

Definition H.8.4 — Hilbert (A,B)-bimodule. A complete (completed) inner product (A,B)-module is called a Hilbert (A,B)-bimodule.

Definition H.8.5 — Hilbert bimodule. A Hilbert bimodule (A,B) is a right A-module and a (left) Hilbert B-module.

The tensor product of left or right modules over the same ring can be generalized.

Definition H.8.6 — Balanced tensor product. The balanced tensor products of a left R-modules A and right R-module B is defined as follows. Let $F(A \times B)$ be the free group over $A \times B$. This set is an Abelian group. Define the following equivalence relation $\sim$ on $A \times B$.

$$\forall a, a_1, a_2 \in A, \forall b, b_1, b_2 \in B, \forall c \in R : \tag{H.8.2}$$
$$(a_1, b) + (a_2, b) \sim (a_1 + a_2, b), \tag{H.8.3}$$
$$(a, b_1) + (a, b_2) \sim (a, b_1 + b_2), \tag{H.8.4}$$
$$(ac, b) \sim (a, cb), \tag{H.8.5}$$

The balanced tensor product of A and B is the R-module $A \otimes B = F(A \times B)/\sim$. If A and B are (X, R)- and (R, Y)-bimodules then with same definition the balanced tensor product is an (X, Y)-bimodule.

Rings are often studied by studying their modules.

Definition H.8.7 — Morita equivalence. Two rings R and S are Morita equivalent if the corresponding categories of modules R-mod and S-mod are equivalent as categories.

For two commutative rings Morita equivalence implies isomorphisms between R and S. But a commutative ring R can be Morita equivalent to a (obviously) non-isomorphic noncommutative ring S.

H.9 Differential algebras

A derivation is linear map that interacts with the product of the structure is manner described by the ordinary product rule of differentiation. The derivations are additive morphisms but they are not true ring morphisms since they replace the normal morphism demand of

$$D(ab) = D(a)D(b) \tag{H.9.1}$$

with

$$D(ab) = D(a)b + aD(b) \tag{H.9.2}$$

Definition H.9.1 — K-derivation. A K-derivation on a K-algebra A, is a K-linear map $D : A \to A$ such that $D(ab) = (Da)b + a(Db)$

Definition H.9.2 — Inner derivation. Let A be a K-algebra. Any element $b \in A$ defines a map $D : A \to A$ by $b(a) : a \to (ba - ab)$. This map is a derivation, and any such derivation is called an inner derivation.

Since an inner derivation is basically the commutator of two elements there are no (non-trivial) inner derivations on a commutative algebra.

Definition H.9.3 — Outer derivation. An outer derivation is a derivation that is not a inner derivation.

In a commutative algebra all derivations are outer derivations. A more general concept of derivation is defined by a map from an algebra into an algebra bimodule.

Definition H.9.4 — Generalized K-derivation. Let A be a K-algebra and let M be an A-bimodule, and D. A generalized K-derivation is a K-linear map $D : A \to M$ such that $D(ab) = (Da)b + a(Db)$. Note that $D(ab), D(a), D(b) \in M$ and $a, b \in A$, and that the defining property requires the A-bimodule structure to be well defined.

Definition H.9.5 — The K-module of derivations. The K-module of K-derivations of A is denoted $Der_K(A)$ The K-module of derivations of A into the A-bimodule M is denote $Der_K(A, M)$.

We have corresponding adaptions of the terms inner and outer derivations for the generalized derivations.

Definition H.9.6 — Inner (general) derivation. Let A be a K-algebra and let M be an A-bimodule. Any element $m \in M$ defines a map $D : A \to M$ by $ad(m) : a \to (ma - am)$. This map is a (general) derivation, and any such (general) derivation is called an inner (general) derivation.

Definition H.9.7 — Outer (general) derivation. An outer (general) derivation is a (general) derivation that is not a inner (general) derivation.

Definition H.9.8 — Differential rings, fields and algebras. A ring, field or algebra with one or more defined maps that are derivations. (Also applies to Lie algebras.)

■ **Example H.9.9 — Derivations.** The set of differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$ is a an $\mathbb{R}$-algebra. Partial derivative of those are $\mathbb{R}$-derivations. ■

Definition H.9.10 — Differential algebra. A differential algebra is an pair (A, d) where A is an algebra and d is a derivation $d\colon A \to A$.

Definition H.9.11 — Differential graded algebra. Let $\mathcal{A}$ be an associative K-algebra with unit 1. Let $\Omega(\mathcal{A})$ be a graded associative K-algebra which we write as

$$\Omega(\mathcal{A}) = \bigoplus_{r=0}^{\infty} \mathcal{A}^r \tag{H.9.3}$$

A differential graded algebra (DGA) is pair (A, d) where d is a linear operator

$$d : \Omega^r(\mathcal{A}) \to \mathcal{A}^{r+1} \tag{H.9.4}$$

which satisfies $d^2 = 0$, $d(1) = 0$ and

$$d(\omega\eta) = d(\omega)\eta + (-1)^k \omega(d\eta) \quad \text{for}\omega \in \tag{H.9.5}$$

$$mathcalA^k \tag{H.9.6}$$

Definition H.9.12 — Differential calculus. A Differential calculus is a differential graded algebra $\Omega^\bullet\mathcal{A}$ such that

$$\Omega^0 = \mathcal{A} \tag{H.9.7}$$

Definition H.9.13 — First order differential calculus. $\left(\Omega^1(\mathcal{A}), d\right)$ is a first order differential calculus

H.10 Functional analysis

We will on several occasions use special techniques related to functions on a set of functions. Such functions are often called functional, but we will mostly avoid this term as it has some imprecise connotations (like suggesting that functionals are not functions, which they are). On occasion we use the notation $[f]$ to suggest that f is to be understood as a variable on a function space. We will not make any attempt at completeness and we simply state a few concept and results that we need.

Let M be a space of functions $f : X \to \mathbb{R}$. There is a functional F_x that for each function f returns the value of f at the point $x \in X$. We can write this functional as

$$F_x[f] = \int f(x')\delta(x - x')dx' = [f](x) \tag{H.10.1}$$

Such functions will be important. They form the elementary functions out of which we formulate the Poisson brackets of the classical algebra of observables. We will just stipulate that functional derivatives follow the rules of linearity, the product rules and the chain rule. The functional derivative of a function of functions $F[\phi]$ will be written as

$$\frac{\delta F[\phi]}{\delta\phi} \tag{H.10.2}$$

Generalized functions, distributions

Many "functions" encountered in mathematical physics are not actually functions, the prime example being the Dirac function (the Dirac distribution). Often in physics these are treated just like ordinary functions, but this can lead to problems. Indeed much of renormalization can be seen as way of dealing with the distributional nature of the field. Since we are dealing quite a bit with some more detailed aspects of functions spaces we will have occasion to use this concept and we need to be more precise. We first define a suitable set of test functions.

Definition H.10.1 — Test functions. The set of test functions $D(\mathbb{R})$ is defined as the set of functions $f : \mathbb{R} \to \mathbb{R}$ such that the following criteria are fulfilled.

1. The function f is infinitely differentiable.
2. The function f has compact support.

The set of test functions forms a vector space.

Definition H.10.2 — Distribution. A distribution is linear map from the set of test functions $D(\mathbb{R})$ to $\mathbb{R}$.

The space of all distributions on $D(\mathbb{R})$ is denoted $D'(\mathbb{R})$. One can view $D'(\mathbb{R})$ as the dual space of $D(\mathbb{R})$. The set of distributions is also a vector space.

Sometimes a more general set of test functions are needed. Schwartz space is an example of a space of generalized test functions. Schwartz space is the space of all infinitely differentiable functions that are rapidly decreasing at infinity, and where all partial derivatives are also rapidly decreasing at infinity.

Definition H.10.3 — Schwartz space. Schwartz space $S(\mathbb{R}^n)$ is defined by

$$S\left(\mathbb{R}^n\right) = \{f \in C^\infty(\mathbb{R}^n) : \|f\|_{\alpha,\beta}; \infty \quad \forall \alpha, \beta\}$$

The set of distributions on Schwartz space (the dual of Schwartz space) is called the set of tempered distributions and is denoted $S'(\mathbb{R}^n)$.

Definition H.10.4 — Projection. A projection is a map $P\colon X \to Y$ such that $Y \subset X$ and that $P \circ P = P$

H.11 Notes

A treatement of functional derivatives appropriate for our use can be found in [27]. General references for functional analysis are Reed and Simmons [426], Ha [105] and Rudin [427]. For the material on C*-algebras and W*-algebras see the book by Blackadar [107]. A brief treatement of Morita equivalence and KK-theory can be found in [101]. Information on differential algebras can be found in [110] [111] [428] [429] [109] [430].

I. Algebraic topology

Algebraic topology is about establishing a correspondence between a given topological space and a set of groups or rings (sometimes modules or algebras). We can then answer questions about topology by reformulating the question to be about the corresponding algebraic structures. The most fundamental conclusion that can be made in homology is that if two spaces have different homology they are not equivalent. The reverse statement does not hold in general. Homology and cohomology can be defined for other spaces of the geometric variety as well as for purely algebraic spaces. Homological algebra takes an abstracted version of homology and employs it for purposes of analyzing and classifying algebraic structures. Techniques from algebraic topology and homology have an ever increasing place in mathematical physics. We therefore give a brief introduction to the subject.

More explicitly this correspondence takes the form of a covariant or contravariant functor between the category of topological spaces and the relevant algebraic category, typically the category of Abelian groups. A (covariant) functor means that for topological spaces A and B with morphism $f : A \to B$ there are corresponding Abelian groups G and H and a morphism $k : G \to H$. The three major such functors are called homotopy, homology and cohomology.

I.1 Homology, cohomology, homological algebra concepts

Homology and cohomology are dual theories. Even though they both can be defined in terms of the other they have somewhat different emphasis and strengths. In addition it depends on the subject matter which theory is easiest to define (or invent), and as such they are both important as theories in themselves (i.e. cohomology are not always just defined as the co-theory of homology). (co)Homology a topological space starts with associating it space to a (co-)chain complex. We will return to the issue of how this association is made. There are different possible (co)homologies that one can define on a topological space and these will define different chain complexes.

Definition I.1.1 — Chain complex. A chain complex $(A_\bullet d_\bullet)$ is a sequence (A_n, d_n) of Abelian

groups A_n with group morphisms $d_n : A_n \to A_{n-1}$ such that $d_n \circ d_{n+1} = 0$. We write this as

$$\cdots \to A_{n+1} \xrightarrow{d_{n+1}} A_n \xrightarrow{d_n} A_{n-1} \xrightarrow{d_{n-1}} A_{n-2} \to \cdots \xrightarrow{d_2} A_1 \xrightarrow{d_1} A_0 \xrightarrow{d_0} A_{-1} \xrightarrow{d_{-1}} A_{-2} \xrightarrow{d_{-2}} \cdots \tag{I.1.1}$$

Definition I.1.2 — Cochain complex. A cochain complex $(A^\bullet, d^\bullet)$ is a sequence (A^n, d^n) of Abelian groups A^n with group morphisms $d_n : A^n \to A^{n+1}$ such that $d_{n+1} \circ d_n = 0$. We write this as

$$\cdots \to A^{-2} \xrightarrow{d^{-2}} A^{-1} \xrightarrow{d^{-1}} A^0 \xrightarrow{d^0} A^1 \xrightarrow{d^1} A^2 \to \cdots \to A^{n-1} \xrightarrow{d^{n-1}} A^n \xrightarrow{d^n} A^{n+1} \to \cdots . \tag{I.1.2}$$

■ **Example I.1.3 — Chain complex.** The classical examples of chain complexes are those defined by simplical or singular homology. They essentially associate chains to triangulations of the topological spaces. We will look at this in more detail in later sections. ■

■ **Example I.1.4 — Cochain complex.** The classical example of a cochain complex is the one defined by de Rham cohomology. This complex is defined for topological spaces that are manifolds and is essentially defined by the various groups of n-form fields on a given manifold. We will look at this in more detail in later sections. ■

The (co)chain complexes are an intermediate step in defining homology. We now process the chains to extract the interesting information that is found in them. We not introduce three (or six) essential pieces of terminology. They are(co)chain, (co)boundary, (co)cycle. The (co)chain is the groups listed in (co)chain complex. The (co)boundaries, the (co)cycles are subgroups of the chain groups.

Definition I.1.5 — Chain group. The Abelian groups A_n listed in a chain complex are called chain groups. The elements are called chains.

Definition I.1.6 — Group of boundaries. The group of boundaries B_n of the chain group A_n is the image of d_{n-1}.

Definition I.1.7 — Group of cycles. The group of cycles Z_n of the chain group A_n is the kernel of d_n.

The equivalent definitions for cochains are

Definition I.1.8 — Cochain group. The Abelian groups A^n listed in a chain complex are called cochain groups. The elements are called cochains.

Definition I.1.9 — Group of coboundaries. The group of coboundaries B^n of cochain group A^n is the image of d^{n+1}.

Definition I.1.10 — Group of cocycles. The group of cocycles Z^n of a cochain group A^n is the kernel of d^n.

Theorem I.1.11 The sets B_n and Z_n are normal subgroups of A_n, and B_n is a normal subgroup of Z_n. The sets B^n and Z^n are normal subgroups of A^n, and B^n is a normal subgroup of Z^n.

Definition I.1.12 — Homology group. The homology group of a chain group A_n, is the quotient group $H_p = Z_n/B_n$.

Definition I.1.13 — Homology classes. The equivalence classes of cycles $[z_p] \in H_p$ are called homology classes.

Definition I.1.14 — Cohomology group. The cohomology group of a cochain group A^n, is the quotient group $H^p = Z^n/B^n$.

Definition I.1.15 — Cohomology classes. The equivalence classes of cocycles $[z^p] \in H^p$ are called cohomology classes.

Thus homology is a covariant functor from the category of chain complexes to the category of R-modules, and one can similarly show that cohomology is a covariant functor from the category of cochain complexes to the category of R-modules.

Definition I.1.16 — Exact sequence. An exact sequence is a (co)chain complex whose (co)homology groups are all trivial.

An chain complex that is an exact sequence is also called an acyclic chain complex. The triviality of the homology groups translates to $\mathrm{Im}(d_n) = \mathrm{Ker}(d_{n+1})$.

Definition I.1.17 — Short exact sequence. A short exact sequence is an exact sequence of 5 elements of the form

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0, \tag{I.1.3}$$

or alternatively

$$1 \to A \xrightarrow{f} B \xrightarrow{g} C \to 1. \tag{I.1.4}$$

The existence of a short exact sequence means that $C \cong B/f(A)$.

Definition I.1.18 — Split exact sequence. An exact sequence

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0, \tag{I.1.5}$$

is said to split if it is isomorphic to

$$0 \to A \xrightarrow{f} (A \oplus C) \xrightarrow{g} C \to 0. \tag{I.1.6}$$

Simplical homology

Defining simplicial homology comes down to defining how to triangulate a toplogical space so that it is represented a simplical complex. A simplical complex is a particular kind of sum of simplices. A triangulation means to express a (geometric) set in terms of some simpler geometric building blocks (like triangles and tetrahedra). We first need the definition of an n-simplex which is basic set of building blocks for our triangulations.

Definition I.1.19 — n-simplex. An n-simplex is a n-dimensional polytope which is the convex hull of its n + 1 vertices.

The basic examples of this are points (0-simplex), line-segments (1-simplex), triangles (2-simplex) and tetrahedron (3-simplex). Beyond three dimension they are just called n- simplices. We shall require that each simplex is equipped with an orientation, which is simply an ordering of its vertices. We represent a simplex as an order list of its vertices $\sigma_i = (v_1, ..., v_n)$. Next we need to define how we allow these simplices to be put together into a simplicial complex, but first we need the concept of a face of a simplex. A face is basically a subsimplex of a simplex. Formally we have

Definition I.1.20 — Face. The convex hull of any nonempty subset of the n+1 points that define an n-simplex is called a face of the simplex.

A face is obviously a simplex (of lower dimensions) in itself. The ordering of a face is inherited from its parent simplex.

Definition I.1.21 — Simplicial complex. A simplicial complex K is a set of simplicies that satisfies the following two conditions.

1. Any face of a simplex from K is also in K.
2. The intersection of two simplices $\sigma_1, \sigma_2 \in K$ is either empty or a face of both σ_1 and σ_2.

Various generalization of the simplicial complex exist, the most common are delta-set and simplicial set. We know define how we use the simplicial complex for discretizing a space.

Definition I.1.22 — Triangulation. A triangulation of a topological space X is a simplicial complex K, homeomorphic to X, together with a homeomorphism $h : K \to X$.

We have now define how to associate a simplicial complex with a topological space. We now proceed to define the chain complex of a simplicial complex. The A_n chain group is defined to be the group of all possible Z-linear combination of the n-simplices of the simplicial complex. We write this as

$$\sum z_i \sigma_i \tag{I.1.7}$$

We define the negative of a n-simplex to the n-simplex with orientation reversed. The homorphism between the chain groups is called the boundary map and is defined as follows.

Definition I.1.23 — Boundary map. Let $\sigma = (v_1, ..., v_n)$ be an n-simplex belonging to simplicial complex, then its boundary map is given by

$$d_n(\sigma) = \sum (-1)^n (v_1, \ldots, \hat{v_i}, \ldots, v_n), \tag{I.1.8}$$

where the hat denotes that this element has been left out. The map is extended by linearity to all linear sums over Z of the n-simplices of the singular complex.

Theorem I.1.24 The boundary of a boundary is zero. That is, for any n, we have that $d_{n+1} d_n = 0$.

Lemma I.1.25 The set of chain groups and the associated boundary maps defined above define a chain complex.

As we have defined before, such a chain complex defines a set of homology groups. The corresponding homology groups are called the n-th homology group of the topological spaces that is

homeorphic to the simplicial complex. To make these definitions meaningful we need to confirm that the arbitrariness in the choice of simplicial complex does not affect the actual homology groups.

Theorem I.1.26 The homology groups of a topological space is independent of the choice of triangulation.

Singular homology

Singular homology is similar to simplicial homology. In this case the n-chain groups of a space X are defined a bit differently. The n-chain group is the set of all finite Z-linear combinations $\sum z_i\sigma_i$ of all continuous maps $\sigma_i : \Delta^n \to X$. An individual map $\sigma_i : \Delta^n \to X$ is called a singular n-simplex. The boundary map is again defined by sum

$$d_n(\sigma) = \sum(-1)^n(v_1, \ldots, \hat{v}_i, \ldots, v_n). \tag{I.1.9}$$

This result of this map is again a singular (n-1)-simplex. The rest of the setup is defined analogously to simplicial homology. For spaces where both theories apply the homology groups are isomorphic.

Cohomology from homology

Let a chain-complex be given by

$$\cdots \to A_{n+1} \xrightarrow{d_{n+1}} A_n \xrightarrow{d_n} A_{n-1} \xrightarrow{d_{n-1}} A_{n-2} \to \cdots \tag{I.1.10}$$

From this chain complex we can define a new related cochain complex by taking the dual of all groups and all group homomorphisms. To define the dual of groups we need to assume they are modules. If no explicit module structure is given we will see the groups as modules over $\mathbb{Z}$. Any Abelian group can bee viewed as module over $\mathbb{Z}$. The dual of group A_n(seen as a module over a ring R) is then the group (module) $Hom(A_n, R)$. The dual of the boundary map d_n^* is given by the transpose of d_n. These mappings define a new cochain complex, which then defines the corresponding cohomology theory.

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & A_n & \xrightarrow{d_n} & A_{n+1} & \xrightarrow{d_n} & A_{n+2} & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle \eta_1} & & \downarrow{\scriptstyle \eta_2} & & \downarrow{\scriptstyle \eta_3} & & \\
\cdots & \longleftarrow & A_n^* & \xleftarrow{d_{n+1}^*} & A_{n+1}^* & \xleftarrow{d_n^*} & A_{n+2}^* & \longleftarrow & \cdots
\end{array}$$

de Rham cohomology

The theory of de Rham cohomology is a cohomology defined for a smooth manifold M. This cohomology theory is defined directly, and not as the dual of a homology theory. (It still is the dual of a homology theory, but that is not how it is defined.) The n-chain groups of the cochain complex defining de Rham cohomology are the sets of n-forms $\Omega^n(M)$ on M, which form a group under ordinary form addition. The boundary map is the ordinary exterior derivative denoted by d. We write this cochain complex as

$$0 \to \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \Omega^2(M) \xrightarrow{d} \Omega^3(M) \to \cdots . \tag{I.1.11}$$

As is well know from differential geometry, for a given n-form α, we have that $d(d\alpha) = 0$. Since d is a linear operator (i.e. a module morphism) these maps are group homomorphisms. We have

thus defined a cochain complex. Cocycles of this cochain complex are called closed forms , and coboundaries are called exact forms. The cohomology groups are called de Rham cohomology groups, and the n-th de Rham cohomology group is the group of equivalence classes of closed n-forms modulo exact n-forms.

Theorem I.1.27 — Poincare duality theorem. Let M be a closed oriented n-dimensional manifold, R be a ring, and k an integer, then there is a canonical isomorphism between the k-th cohomology group $H^k(M, R)$ and the (n-k)th homology group $H_{n-k}(M, R)$

Dolbeault cohomology

Dolbeault cohomology is the analog of de Rham cohomology for complex manifolds. In this case one operates with (p, q) differential forms. The cohomology groups are denoted as $H^{p,q}(M, \mathbb{C})$. The dimensionality (as a vector space) for each (p, q) cohomology group is called a Hodge number, and it is denoted as $h^{p,q}(M)$.

Hochschild cohomology

Homology and cohomology and also be defined for algebraic spaces. Hochschild cohomology (and homology) is a cohomology theory for associative algebras over rings. We define Hochschild cohomology by defining its cochain complex. Let k be a ring, let A be an associative k-algebra, and M and A-bimodule. In the Hochschild cochain complex the cochain groups are defined by

$$C^0 = M \quad C^n(A, M) := \mathrm{Hom}(A^{\otimes n}, M). \tag{I.1.12}$$

The cochain complex is

$$C^0(A, M) \xrightarrow{\delta} C^1(A, M) \xrightarrow{\delta} C^2(A, M) \xrightarrow{\delta} C^3(A, M) \to \cdots, \tag{I.1.13}$$

where the boundary map δ is defined by

$$\begin{aligned} (\delta m)(a) &= ma - am \\ (\delta f)(a_1, \ldots, a_{n+1}) &= a_1 f(a_2, \ldots, a_{n+1}) \\ &\quad + \sum_{i=1}^{n} (-1)^{i+1} f(a_1, \ldots, a_i a_{i+1}, \ldots, a_{n+1}) \qquad \text{(I.1.14)} \\ &\quad + (-1)^{n+1} f(a_1, \ldots, a_n) a_{n+1}. \qquad \text{(I.1.15)} \end{aligned}$$

We can specialize to the case where $M = A^*$ and $k = \mathbb{C}$. Doing this we get (note the number of factors in the tensor product has increased by one)

$$C^0 = \mathbb{C} \quad C^n(A, M) := \mathrm{Hom}(A^{\otimes n+1}, \mathbb{C}). \tag{I.1.16}$$

The boundary map b is defined by

$$\begin{aligned} (bf)(a_0, \ldots, a_{n+1}) &= \sum_{i=0}^{n} (-1)^{i+1} f(a_0, \ldots, a_i a_{i+1}, \ldots, a_{n+1}) \qquad \text{(I.1.17)} \\ &\quad + (-1)^{n+1} f(a_{n+1} a_0, \ldots, a_n). \qquad \text{(I.1.18)} \end{aligned}$$

The pairs $(C^n(A, M), b_n)$ define a cochain complex. The cohomology groups of this cochain complex are called Hochschild cohomology groups, and are denoted by $HH^n(A)$.

Cyclic cohomology

Cyclic cohomology is very closely related to Hochschild cohomology. We consider only the case where $M = A^*$ and $k = \mathbb{C}$. The cyclic cochain complex can most easily be defined as a subcomplex of the Hochschild cochain complex. The essential difference is that we operate on subset of chains, namely the subset of cyclic cochains. Cyclic cochains are explained by the following definition.

Definition I.1.28 — Cyclic cochain. An n-cochain of the Hochschild complex is called cyclic if

$$f(a_n, a_0, \ldots, a_{n-1}) = (-1)^n f(a_0, \ldots, a_n) \tag{I.1.19}$$

The set of all cyclic n-cochains is denoted $C^n_{lambda}(A)$. We need the following lemma

Lemma I.1.29 — Cyclic cohains. The boundary operator b maps $C^n_{lambda}(A)$ to a subset of $C^{n+1}_{lambda}(A)$.

Proof. See [100]. ■

We can therefore define the cyclic cochain complex as

$$C^0_\lambda(A, M) \xrightarrow{\delta} C^1_\lambda(A, M) \xrightarrow{\delta} C^2_\lambda(A, M) \xrightarrow{\delta} C^3_\lambda(A, M) \to \cdots \tag{I.1.20}$$

The cyclic cohomology of this complex (or of A) is defined as usual and is denoted by $HC^n(A)$.

Homotopy

Homotopy defines groups based on maps of intervals into a space. If the paths are loops with a fixed base point we can define the structure of a group on the equivalence classes of such loops. This group can tells us about the number of "holes" in X that can be "caught" by one-dimensional loops. Higher order homotopy groups, define by using n-spheres, instead of loops can tell us about higher dimensional "holes". We first define the basic elements of a path and a loop.

Definition I.1.30 — Path. A path in a space X is a map $f : [0, 1] \to X$.

Definition I.1.31 — Loop. A loop is a path f where $f(0) = f(1)$.

A loop is of course nothing else that a path that ends at the same point as it starts. The start (and end) point of a loop is called its basepoint.

Definition I.1.32 — Homoptopy of paths. A homoptopy of paths is a continuous map $F(s, t) : [0, 1] \times [0, 1] \to X$ such that

1. The start and end points $F(0, t)$ and $F(1, t)$ are the same for all t.

Two paths f_1, f_2 in $X$are said to be homotopic if there exists a homotopy of paths $F(s, t)$ in X, such that $f_1 = F(s, 0)$ and $f_2 = F(s, 1)$. If we apply the definition of being homotopic to the set of all loops with the same basepoint we obtain an equivalence relation. We have the slightly more general theorem

Theorem I.1.33 The relation of homotopy on paths with fixed endpoints is an equivalence relation.

Let f and g be two loops on X. We define their product by

$$f * g = \begin{cases} f(2t) & \text{if } t \in [0, 1/2] \\ g(2t-1), & \text{if } t \in [1/2, 1] \end{cases} \tag{I.1.21}$$

We can define an inverse of a loop $f(t)$ as $f(t-1)$, we define the identity map as the constant loop where $f(t)$ is equal to the basepoint for all t.

Theorem I.1.34 The equivalence classes of loops in a space X at a fixed basepoint x_0 forms a group under the product, inverse and identity defined above. It is called the fundamental group and denoted by $\pi_1(X, x_0)$

■ **Example I.1.35 — Euclidean space.** The fundamental group of $\mathbb{R}^n$ is the trivial group. ■

■ **Example I.1.36 — Circle.** The fundamental group of S^1 is Z. ■

A loop in $_X$is essentially a map $f : S^1 \to X$. The fundamental group is the lowest order homotopy group, it can be generalized by the same constructions as just shown by using maps $f : S^n \to X$. This will define higher-order homotopy groups. Higher-order homotopy groups can distinguish higher dimensional holes but are difficult to compute. Homotopy groups of order higher than the dimensions of the space it caracterized are not necessarily zero. This phenomena is considered to an anomalous (not meaningfull) feature of homotopy theory.

I.2 Notes

The basic reference for this chapter is Hatcher [431].

J. Other math

J.1 Various

Definition J.1.1 — Support of a function. The support of a function is the subset of the domain where the function is non-zero.

$$supp(f) = \{x \mid f(x) \neq 0\} \tag{J.1.1}$$

For functions on a topological space the support is defined to be the closure of the set on which it is non-zero. A function on a topological space which support is a compact set is said to have compact support. A function is said to have finite support if the support is a finite set.

J.2 Almost periodic functions

Almost periodic functions are important for loop quantum cosmology. They will form the basis functions for the Hilbert space of states in Friedmann cosmology. After defining the almost periodic functions on $\mathcal{R}$ we will define a compactification called the Bohr compactification of $\mathcal{R}$. The Bohr compactification allows us to trade almost periodic function on $\mathcal{R}$ with continuous functions on the Bohr Compactification.

Definition J.2.1 — Relatively dense. A set $E \subset \mathbb{R}$ is called relatively dense if there exists a number $l > 0$ such that all intervals of the type $\langle a, a + l\rangle \subset \mathbb{R}$ contains at least one number from E.

Definition J.2.2 — ϵ-almost period. Let X be a complete metric space with metric $d(x, y)$. A number τ is called an ϵ-almost period of a function $f : \mathbb{R} \to X$ if $d\left(f(x), f(x + \tau)\right) \leq \epsilon$ for all x.

Definition J.2.3 — Almost periodic function. A continuous function $f : \mathbb{R} \to X$ is called almost periodic if it has a relatively dense set of ϵ-almost periods for any $\epsilon > 0$.

Lemma J.2.4 Every periodic function is an almost periodic function.

Proof. Let f be a periodic function with period T. Then any integer multiple of T is also a period of f. Any such integer multiple is also an ϵ-almost period of f for any $\epsilon > 0$. Choose some $l > T$ then any interval $\langle a, a + l\rangle$ contains an ϵ-almost period of f. Hence the set of ϵ-almost period is relatively dense for any ϵ, and hence the function f is almost periodic. ■

Definition J.2.5 — Bohr compactification. The Bohr compactification of a topological group G is compact Hausdorff topological group and a continuous group homomorphism $\phi : G \to H$ which is universal.

Theorem J.2.6 The Bohr compactification exists and is unique up to isomorphisms.

Theorem J.2.7 Let H be the Bohr compactification of G with homomorphism ϕ. A bounded continuous complex-valued function f on G is almost periodic iff there is a continuous function g on H such that $f = g \cdot \phi$.

J.3 Fredholm operators

Definition J.3.1 — Fredholm operator. A Fredholm operator is a bounded linear operator between to Banach spaces, with finite dim kernel and cokernel, (and closed range).

Definition J.3.2 — Fredholm module. Let A be a separable C*-algebra, and let $\mathcal{H}$ be a Hilbert space. A Fredholm module $(\rho, \mathcal{H}, F)$ where ρ is a representation of A on $\mathcal{H}$ and F is an operator of square 1, with the property that

$$[F, a] \tag{J.3.1}$$

is compact for all $x \in A$.

Definition J.3.3 — Index of a Fredholm operator. The index of a Fredholm operator is given by

$$Index = Dim(ker) - Dim(coker) \tag{J.3.2}$$

J.4 Real and complex structure

It is often useful to define a real vector space from a complex one and vice versa.

Definition J.4.1 — Real structure. Let V be a complex vector space. A real structure on V is an anti-linear involution $J \colon V \to V$.

Using the real structure J one defines two real vector spaces

$$V^+ = \{v \in V \mid Jv = v\} \tag{J.4.1}$$
$$V^- = \{v \in V \mid Jv = -v\}, \tag{J.4.2}$$

and proves that $V \simeq V^+ \oplus V^-$. Here V is a vector space over $\mathbb{C}$ and $V^+ \oplus V^-$ is a vector space over $\mathbb{R} \times \mathbb{R}$. Component-wise complex conjugation on $\mathbb{C}^n$ is an example of a real structure. By selecting all vectors that are invariant under complex conjugation one picks out the subspace V^+ of "real" vectors.

Definition J.4.2 — Pseudo real structure. Let V be a complex vector space. A pseudoreal structure on V is an anti-linear $J\colon V \to V$ such that $J^2 = -Id_V$.

A pseudoreal structure is also called a quaternionic structure. Note that the next definition is not the same thing as it defines a linear map instead of an anti-linear map.

Definition J.4.3 — Complex structure. Let V be a real vector space. A complex structure on V is an linear map $J\colon V \to V$ such that $J^2 = -Id_V$.

Using the complex structure J one defines complex scalar multiplication on V by setting

$$(x + iy)v = xv + yJ(v). \tag{J.4.3}$$

This creates a new complex vector space V_J from the real vector space V.

Definition J.4.4 — Complexification (1). Let V be a real vector space. The complexification of V is the complex vector space

$$V^{\mathbb{C}} = V \otimes \mathbb{C}, \tag{J.4.4}$$

where multiplication by complex numbers is defined by

$$\alpha(v \otimes \beta) = v \otimes \alpha\beta \quad v \in V, \alpha, \beta \in \mathbb{C}. \tag{J.4.5}$$

Definition J.4.5 — Complexification (2). Let V be a real vector space. The complexification of V is the complex vector space

$$V^{\mathbb{C}} = V \oplus V, \tag{J.4.6}$$

where multiplication by complex numbers is defined by

$$(x + iy)(v \otimes w) = (xv - yw, yv + xw) \quad v, w \in V, x, y \in \mathbb{R}. \tag{J.4.7}$$

There is also an alternative way to specify the last construction. Let $W = V \oplus V$ be a real vector space. The map $J\colon (v, w) \mapsto (-w, v)$ defines a complex structure on W and the complex multiplication defined by J is equivalent to the one in definition J.4.5.

J.5 Notes

Almost periodic functions are treated in[432].

Quotes

Preface

Husserl [433, quoted on page 254]

Chapter 1

Hossenfelder [434]
Misner et al. [76, page 1212]
Wigner [435]
Baez [436, page 182]
Polchinski [291, page 2]

Chapter 2

Misner et al. [76, page 1208]
Frolich and Schubnel [437, page 2]
Aronson [394, page 110]

Chapter 3

Weinberg [44, page XX]
Weinberg [64, page 8]
Schwartz [51, page 508]
Miraman [50]
Bekaert et al.[438]

Chapter 4

Sotiriou et al. [439]
Hertz, 1884
Casola and Liberati [82, page 6]
Arkani-Hamed[440, page 2]
Aldrovani and Pereira [84, page 191]

Chapter 5

Baez [436, p177]

Chapter 6

Kastler [441, p132]

Chapter 8

Baez [436, p182]

Chapter9

Tong [250, p209]
Horowitz and Polchinski [442, p169]
Balasubramanian [443, p399]
Linde [444, p399]

Chapter 10

Rovelli [445, page 1]

References

References for Chapter 1: Epistemology

[1] Gorini, Edoardo. "Supersymmetry Searches in _ATLAS". Sept. 2016. Web-link (cited on page 1).

[2] Cern. *Atlas Experiment - Public Results*. Web-link (cited on page 1).

[3] organizers, Conference. *Is SUSY Alive and Well? - IftWorkShops - Survey*. Web-link (cited on page 1).

[4] Flori, Cecilia. *A First Course in Topos Quantum Theory*. Springer, Mar. 27, 2013 (cited on pages 4, 5).

[5] David, Francois. *The Formalisms of Quantum Mechanics: An Introduction*. Volume 893. Springer, 2014. Web-link (cited on pages 4, 21).

[6] Stevens, S. S. "On the Theory of Scales of Measurement". In: *Science* 103.2684 (June 7, 1946), pages 677–680. DOI: 10.1126/science.103.2684.677. pmid: 17750512. Web-link (cited on page 5).

[7] Bostock, David. *Philosophy of Mathematics: An Introduction*. John Wiley & Sons, Mar. 9, 2009 (cited on page 6).

[8] Bancerek, Grzegorz et al. "Mizar: State-of-the-Art and Beyond". In: *Intelligent Computer Mathematics*. Edited by Kerber, Manfred et al. Lecture Notes in Computer Science 9150. Springer International Publishing, July 13, 2015, pages 261–279. DOI: 10.1007/978-3-319-20615-8_17. Web-link (cited on page 6).

[9] Kerber, Manfred et al., editors. *Intelligent Computer Mathematics*. Volume 9150. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015. Web-link (cited on page 6).

[10] Takeuti, Gaisi and Zaring, Wilson M. *Introduction to Axiomatic Set Theory*. Springer-Verlag, 1982 (cited on pages 7, 300).

[11] Carr, Bernard. *Universe Or Multiverse?* Cambridge University Press, June 21, 2007 (cited on pages 9, 240, 241).

References for Chapter 2: Physical theories in general

[5] David, Francois. *The Formalisms of Quantum Mechanics: An Introduction*. Volume 893. Springer, 2014. Web-link (cited on pages 4, 21).

[12] Bongaarts, Peter. *Quantum Theory*. Springer, 2015. Web-link (cited on pages 13, 20–26, 148, 149).

[13] Strocchi, F. *An Introduction to the Mathematical Structure of Quantum Mechanics: A Short Course for Mathematicians*. World Scientific, 2008 (cited on pages 13, 14, 20–23, 26, 27, 149, 353).

[14] Faddeev, L. D. and Yakubovsky, O. A. *Lectures in Quantum Mechanics for Mathematician Students*. St.-Petersburg University Press, 1980 (cited on pages 13, 20, 22, 26, 27, 149).

[15] Rédei, Miklós and Summers, Stephen Jeffrey. "Quantum Probability Theory". In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 38.2 (2007), pages 390–417. Web-link (cited on pages 14, 17).

[16] Maassen, Hans. "Quantum Probability Theory". In: *Lecture Notes. Caput College Kans-theorie, Wiskunde, KuN, najaar, the Netherlands* (2003). Web-link (cited on pages 14, 17).

[17] Connes, Alain. *Noncommutative Geometry*. Gulf Professional Publishing, 1994 (cited on pages 14, 61, 67, 73).

[18] Reichl, L. E. *A Modern Course in Statistical Physics*. Wiley, Aug. 3, 2009 (cited on page 14).

[19] *Hilbert's Problems*. In: *Wikipedia, the Free Encyclopedia*. June 28, 2016. Web-link (cited on page 16).

[20] Andrei, Kolmogorov. *Grundbegriffe Der Wahscheinlichkeitsrechnung*. 1933 (cited on page 16).

[21] Folland, Gerald B. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, 2013. Web-link (cited on pages 16, 20, 318, 319).

[22] Tao, Terrence. *Algebraic Probability Spaces*. 2014-06-29T06:22:29+00:00. Web-link (cited on pages 17, 319).

[23] Nica, Alexandru and Speicher, Roland. *Lectures on the Combinatorics of Free Probability*. Cambridge University Press, Sept. 7, 2006 (cited on page 19).

[24] Parthasarathy, K. R. *An Introduction to Quantum Stochastic Calculus*. Springer Science & Business Media, Dec. 13, 2012 (cited on page 19).

[25] Thiemann, Thomas. *Modern Canonical Quantum General Relativity*. Cambridge University Press, 2007. Web-link (cited on pages 20, 134, 135, 149, 150, 152, 153, 155, 158–160, 162–167, 169–175, 181, 185, 194, 195, 271).

[26] Takhtadzhian, Leon Armenovich. *Quantum Mechanics for Mathematicians*. American Mathematical Soc., 2008 (cited on pages 20, 27, 135).

[27] Greiner, Walter and Reinhardt, Joachim. *Field Quantization*. Springer Science & Business Media, June 29, 2013 (cited on pages 20, 135, 358).

[28] Rothe, Heinz J. and Rothe, Klaus D. *Classical and Quantum Dynamics of Constrained Hamiltonian Systems*. Volume 81. World Scientific, 2010. Web-link (cited on pages 20, 135).

[29] Henneaux, Marc and Teitelboim, Claudio. *Quantization of Gauge Systems*. Princeton university press, 1992. Web-link (cited on pages 20, 135).

[30] Gleason, Jonathan James. *The C-Algebraic Formalism of Quantum Mechanics*. 2009. Web-link (cited on pages 21, 23, 27).

[31] Davidson, Kenneth R. *C*-Algebras by Example*. American Mathematical Soc., 1996 (cited on page 23).

[32] nLab. *Classical Mechanics in nLab*. Web-link (cited on page 24).

[33] Rudolph, Gerd and Schmidt, Matthias. *Differential Geometry and Mathematical Physics: Part I. Manifolds, Lie Groups and Hamiltonian Systems*. Springer Science & Business Media, Nov. 9, 2012 (cited on pages 24, 331, 334).

[34] nLab. *Jordan-Lie-Banach Algebra in nLab*. Web-link (cited on page 25).

[35] Nielsen, Michael A. and Chuang, Isaac L. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Dec. 9, 2010 (cited on pages 26, 254, 256).

[36] Kanmani, S. "Entangled States : Classical versus Quantum". In: (July 10, 2009). arXiv: 0907.1784. Web-link (cited on page 26).

[37] Majewski, W. A. "Quantum Correlations; Quantum Probability Approach". In: (July 17, 2014). arXiv: 1407.4754. Web-link (cited on page 26).

[38] Ohya, Masanori and Volovich, I. *Mathematical Foundations of Quantum Information and Computation and Its Applications to Nano- and Bio-Systems*. Springer Science & Business Media, Jan. 15, 2011 (cited on pages 27, 67).

[39] Woit, Peter. "Quantum Theory, Groups and Representations: An Introduction (under Construction)". In: (2015). Web-link (cited on page 27).

[40] Duvenhage, Rocco. "Recurrence in Quantum Mechanics". In: (Feb. 4, 2002). arXiv: quant-ph/0202023. Web-link (cited on page 27).

[41] Duvenhage, Rocco. "The Nature of Information in Quantum Mechanics". In: (Mar. 15, 2002). arXiv: quant-ph/0203070. Web-link (cited on page 27).

[42] Duvenhage, Rocco. "The General Structure and Ergodic Properties of Quantum and Classical Mechanics: A Unified C*-Algebraic Approach". In: (Dec. 5, 2003). arXiv: quant-ph/0312051. Web-link (cited on page 27).

References for Chapter 3: Fixed background theories - QFT

[43] Mandl, Franz and Shaw, Graham. *Quantum Field Theory*. John Wiley & Sons, June 6, 2013 (cited on pages 29, 31, 43, 246).

[44] Weinberg, Steven. *The Quantum Theory of Fields*. Volume 1. Cambridge University Press, Aug. 13, 1996 (cited on pages 30, 31, 38, 40–42, 44, 204, 242, 371).

[45] Wald, Robert M. *General Relativity*. University of Chicago Press, May 15, 2010 (cited on pages 30, 33, 36–38, 44, 53, 68, 134, 244, 331).

[46] McCabe, Gordon. *The Structure and Interpretation of the Standard Model*. Volume 2. Elsevier, 2011. Web-link (cited on pages 30, 31, 37).

[47] Arkani-Hamed, Nima and Trnka, Jaroslav. "The Amplituhedron". In: *arXiv preprint arXiv:1312.2007* (2013). Web-link (cited on pages 30, 40).

[48] Wigner, E. "On Unitary Representations of the Inhomogeneous Lorentz Group". In: *Annals of Mathematics*. Second Series 40.1 (Jan. 1, 1939), pages 149–204. DOI: 10.2307/1968551. JSTOR: 1968551 (cited on pages 30, 38, 44).

[49] Bargmann, V. and Wigner, E. P. "Group Theoretical Discussion of Relativistic Wave Equations". In: *Proceedings of the National Academy of Sciences of the United States of America* 34.5 (May 1948), pages 211–223. pmid: 16578292. Web-link (cited on pages 30, 38, 44).

[50] Mirman, R. *Massless Representations of the Poincaré Group: Electromagnetism, Gravitation, Quantum Mechanics, Geometry*. iUniverse, May 2005 (cited on pages 31, 371).

[51] Schwartz, Matthew D. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2013. Web-link (cited on pages 31, 38, 41, 43, 371).

[52] Gruber, Bruno. *Symmetries in Science III*. Springer Science & Business Media, Dec. 6, 2012 (cited on page 33).

[53] Ballentine, Leslie E. *Quantum Mechanics: A Modern Development*. World Scientific, Jan. 1, 1998 (cited on page 33).

[54] Mouchet, Amaury. "An Alternative Proof of Wigner Theorem on Quantum Transformations Based on Elementary Complex Analysis". In: *Physics Letters A* 377.39 (Nov. 2013), pages 2709–2711. DOI: 10.1016/j.physleta.2013.08.017. arXiv: 1304.1376. Web-link (cited on page 36).

[55] Pade, Jochen. *Quantum Mechanics for Pedestrians 2: Applications and Extensions*. Springer Science & Business Media, Nov. 8, 2013 (cited on page 36).

[56] Cheng, Ta-Pei. *Einstein's Physics: Atoms, Quanta, and Relativity - Derived, Explained, and Appraised*. OUP Oxford, Jan. 31, 2013 (cited on pages 37, 46).

[57] Qian, Lei. "Constraining Photon Mass by Energy-Dependent Gravitational Light Bending". In: *Science China Physics, Mechanics and Astronomy* 55.3 (Mar. 2012), pages 523–526. DOI: 10.1007/s11433-012-4648-2. arXiv: 1406.1592. Web-link (cited on page 37).

[58] Cornwell, John F. *Group Theory in Physics: An Introduction*. Academic Press, July 11, 1997 (cited on page 38).

[59] Mackey, George W. "Induced Representations of Locally Compact Groups I". In: *Annals of Mathematics*. Second Series 55.1 (Jan. 1, 1952), pages 101–139. DOI: 10.2307/1969423. JSTOR: 1969423 (cited on pages 38, 44).

[60] Mackey, George W. "Induced Representations of Locally Compact Groups II. The Frobenius Reciprocity Theorem". In: *Annals of Mathematics*. Second Series 58.2 (Sept. 1, 1953), pages 193–221. DOI: 10.2307/1969786. JSTOR: 1969786 (cited on pages 38, 44).

[61] Varadarajan, Veeravalli Seshadri. *Geometry of Quantum Theory*. Springer Science & Business Media, Dec. 3, 2007 (cited on pages 38, 44).

[62] Weinberg, Steven. "Photons and Gravitons in S-Matrix Theory: Derivation of Charge Conservation and Equality of Gravitational and Inertial Mass". In: *Physical Review* 135 (4B Aug. 24, 1964), B1049–B1056. DOI: 10.1103/PhysRev.135.B1049. Web-link (cited on pages 38, 44, 54).

[63] Weinberg, Steven. "Photons and Gravitons in Perturbation Theory: Derivation of Maxwell's and Einstein's Equations". In: *Physical Review* 138 (4B 1965), B988. Web-link (cited on pages 38, 44, 54).

[64] Weinberg, Steven. "What Is Quantum Field Theory, and What Did We Think It Is?" In: *arXiv preprint hep-th/9702027* (1997). Web-link (cited on pages 38, 40, 42, 44, 371).

[65] Pokorski, Stefan. *Gauge Field Theories*. Cambridge University Press, Mar. 13, 2000 (cited on page 41).

[66] Jenkins, Alejandro. “Composite Mediators and Lorentz Violation”. In: (Sept. 20, 2004). arXiv: hep-th/0409189. Web-link (cited on page 41).

[67] Huang, Kerson. “A Critical History of Renormalization”. In: (Oct. 21, 2013). DOI: 10.1142/S021775X13300500. arXiv: 1310.5533. Web-link (cited on page 42).

[68] ’tHooft, G. “Renormalization of Massless Yang-Mills Fields”. In: *Nuclear Physics B* 33.1 (Oct. 1, 1971), pages 173–199. DOI: 10.1016/0550-3213(71)90395-6. Web-link (cited on page 42).

[69] ’t Hooft, G. “Renormalizable Lagrangians for Massive Yang-Mills Fields”. In: *Nuclear Physics B* 35.1 (Dec. 1, 1971), pages 167–188. DOI: 10.1016/0550-3213(71)90139-8. Web-link (cited on page 42).

[70] Wilson, Kenneth G. “Renormalization Group and Critical Phenomena. II. Phase-Space Cell Analysis of Critical Behavior”. In: *Physical Review B* 4.9 (Nov. 1, 1971), pages 3184–3205. DOI: 10.1103/PhysRevB.4.3184. Web-link (cited on pages 42, 44).

[71] Wilson, Kenneth G. “Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture”. In: *Physical Review B* 4.9 (Nov. 1, 1971), pages 3174–3183. DOI: 10.1103/PhysRevB.4.3174. Web-link (cited on pages 42, 44).

[72] Hollowood, Timothy J. *Renormalization Group and Fixed Points: In Quantum Field Theory*. Springer Berlin Heidelberg, Apr. 20, 2013 (cited on page 43).

[73] Heisenberg, W. “Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen.” In: *Zeitschrift für Physik* 33.1 (Dec. 1925), pages 879–893. DOI: 10.1007/BF01328377. Web-link (cited on page 44).

[74] Born, M., Heisenberg, W., and Jordan, P. “Zur Quantenmechanik. II.” In: *Zeitschrift für Physik* 35 (8-9 Aug. 1926), pages 557–615. DOI: 10.1007/BF01379806. Web-link (cited on page 44).

[75] Yang, C. N. and Mills, R. L. “Conservation of Isotopic Spin and Isotopic Gauge Invariance”. In: *Physical Review* 96.1 (Oct. 1, 1954), pages 191–195. DOI: 10.1103/PhysRev.96.191. Web-link (cited on page 44).

References for Chapter 4: Background free theories - GR

[45] Wald, Robert M. *General Relativity*. University of Chicago Press, May 15, 2010 (cited on pages 30, 33, 36–38, 44, 53, 68, 134, 244, 331).

[56] Cheng, Ta-Pei. *Einstein’s Physics: Atoms, Quanta, and Relativity - Derived, Explained, and Appraised*. OUP Oxford, Jan. 31, 2013 (cited on pages 37, 46).

[62] Weinberg, Steven. “Photons and Gravitons in S-Matrix Theory: Derivation of Charge Conservation and Equality of Gravitational and Inertial Mass”. In: *Physical Review* 135 (4B Aug. 24, 1964), B1049–B1056. DOI: 10.1103/PhysRev.135.B1049. Web-link (cited on pages 38, 44, 54).

[63] Weinberg, Steven. “Photons and Gravitons in Perturbation Theory: Derivation of Maxwell’s and Einstein’s Equations”. In: *Physical Review* 138 (4B 1965), B988. Web-link (cited on pages 38, 44, 54).

[76] Misner, Charles W., Thorne, Kip S., and Wheeler, John Archibald. *Gravitation*. W. H. Freeman, Sept. 15, 1973 (cited on pages 46, 53, 371).

[77] Rovelli, Carlo. *Quantum Gravity*. Cambridge university press, 2004. Web-link (cited on pages 46, 52, 53, 134, 153, 168, 181, 185, 194).

[78] Lovelock, David. “The Uniqueness of the Einstein Field Equations in a Four-Dimensional Space”. In: *Archive for Rational Mechanics and Analysis* 33.1 (1969), pages 54–70. DOI: 10.1007/BF00248156. Web-link (cited on page 46).

[79] Galilei, Galileo. *Dialogues Concerning Two New Sciences*. Dover, 1914 (cited on page 46).

[80] Casola, Eolo Di, Liberati, Stefano, and Sonego, Sebastiano. “Weak Equivalence Principle for Self-Gravitating Bodies: A Sieve for Purely Metric Theories of Gravity”. In: *Physical Review D* 89.8 (2014), page 084053. Web-link (cited on page 48).

[81] Will, Clifford M. *Theory and Experiment in Gravitational Physics*. Cambridge University Press, Mar. 11, 1993 (cited on page 49).

[82] Casola, Eolo Di, Liberati, Stefano, and Sonego, Sebastiano. “Nonequivalence of Equivalence Principles”. In: *American Journal of Physics* 83.1 (Jan. 1, 2015), pages 39–46. DOI: 10.1119/1.4895342. Web-link (cited on pages 49, 372).

[83] Pereira, José G. “Teleparallelism: A New Insight into Gravity”. In: *Springer Handbook of Spacetime*. Springer, 2014, pages 197–212. Web-link (cited on page 50).

[84] Aldrovandi, Ruben and Pereira, Jose G. *Teleparallel Gravity: An Introduction*. Volume 173. Springer Science & Business Media, 2012. Web-link (cited on pages 50, 372).

[85] Baez, John C. and Wise, Derek K. “Teleparallel Gravity as a Higher Gauge Theory”. In: *Communications in Mathematical Physics* 333.1 (2015), pages 153–186. Web-link (cited on page 50).

[86] Aalbers, J. “Conformal Symmetry in Classical Gravity”. In: (Aug. 16, 2013). Web-link (cited on page 51).

[87] Bärenz, Manuel. “General Covariance and Background Independence in Quantum Gravity”. In: (July 2, 2012). arXiv: 1207.0340. Web-link (cited on page 51).

[88] Giulini, Domenico. “Remarks on the Notions of General Covariance and Background Independence”. In: *Approaches to Fundamental Physics*. Springer, 2007, pages 105–120. Web-link (cited on page 51).

[89] Rickles, Dean. “Who’s Afraid of Background Independence?” In: *Philosophy and Foundations of Physics* 4 (2008), pages 133–152. Web-link (cited on page 51).

[90] Martin, Daniel. *Manifold Theory: An Introduction for Mathematical Physicists*. Elsevier, 2002. Web-link (cited on pages 52, 68, 340).

[91] Burgess, Cliff P. “Quantum Gravity in Everyday Life: General Relativity as an Effective Field Theory”. In: *Living Reviews in Relativity* 7 (2004). DOI: 10.12942/lrr-2004-5. Web-link (cited on pages 52, 53).

[92] Donoghue, John F. “Introduction to the Effective Field Theory Description of Gravity”. In: (Dec. 11, 1995). arXiv: gr-qc/9512024. Web-link (cited on pages 52, 53).

[93] Rovelli, Carlo. “Forget Time”. In: *Foundations of Physics* 41.9 (2011), pages 1475–1490. Web-link (cited on pages 53, 185).

[94] Feynman, Richard. *Feynman Lectures On Gravitation*. Boulder, Colo.: Westview Press, June 20, 2002 (cited on page 54).

[95] Barceló, Carlos, Carballo-Rubio, Raúl, and Garay, Luis J. “Unimodular Gravity and General Relativity from Graviton Self-Interactions”. In: *Physical Review D* 89.12 (June 16, 2014). DOI: 10.1103/PhysRevD.89.124019. arXiv: 1401.2941. Web-link (cited on pages 54, 55).

[96] Padmanabhan, Thanu. “From Gravitons to Gravity: Myths and Reality”. In: *International Journal of Modern Physics D* 17 (03n04 2008), pages 367–398. Web-link (cited on pages 54, 55).

[97] Ohanian, Hans C. and Ruffini, Remo. *Gravitation and Spacetime*. Cambridge University Press, Apr. 8, 2013 (cited on page 55).

References for Chapter 5: Noncommutative geometry - mathematics

[17] Connes, Alain. *Noncommutative Geometry*. Gulf Professional Publishing, 1994 (cited on pages 14, 61, 67, 73).

[38] Ohya, Masanori and Volovich, I. *Mathematical Foundations of Quantum Information and Computation and Its Applications to Nano- and Bio-Systems*. Springer Science & Business Media, Jan. 15, 2011 (cited on pages 27, 67).

[45] Wald, Robert M. *General Relativity*. University of Chicago Press, May 15, 2010 (cited on pages 30, 33, 36–38, 44, 53, 68, 134, 244, 331).

[90] Martin, Daniel. *Manifold Theory: An Introduction for Mathematical Physicists*. Elsevier, 2002. Web-link (cited on pages 52, 68, 340).

[98] Gracia-Bondia, José M., Varilly, Joseph C., and Figueroa, Hector. *Elements of Noncommutative Geometry*. Springer Science & Business Media, 2001 (cited on pages 61, 65–69, 71–74).

[99] Várilly, Joseph C. *An Introduction to Noncommutative Geometry*. European Mathematical Society, 2006 (cited on page 61).

[100] Khalkhali, Masoud. *Basic Noncommutative Geometry*. European Mathematical Society, 2009 (cited on pages 61, 63, 65, 68, 365).

[101] Suijlekom, Walter van. *Noncommutative Geometry and Particle Physics*. Springer, July 21, 2014 (cited on pages 61, 65, 73, 74, 78, 81, 84, 89–93, 95, 358).

[102] Beenakker, Wim, van den Broek, Thijs, and van Suijlekom, Walter D. *Supersymmetry and Noncommutative Geometry*. Springer, 2015. Web-link (cited on pages 61, 95).

[103] Tabak, John. *Beyond Geometry: A New Mathematics of Space and Form*. Facts on File, 2011 (cited on pages 61, 321).

[104] Morris, Sidney A. *Topology without Tears*. University of New England, 1989. Web-link (cited on pages 61, 170, 327).

[105] Ha, Dzung Minh. *Functional Analysis*. Matrix Editions, Jan. 1, 2006 (cited on pages 62, 346, 358).

[106] Aupetit, Bernard. *A Primer on Spectral Theory*. Springer-Verlag, 1991 (cited on pages 63, 65).

[107] Blackadar, Bruce. *Operator Algebras: Theory of C*-Algebras and Von Neumann Algebras*. Taylor & Francis, 2006 (cited on pages 65, 66, 358).

[108] Brodzki, Jacek. “An Introduction to K-Theory and Cyclic Cohomology”. In: (June 3, 1996). arXiv: `funct-an/9606001`. Web-link (cited on page 68).

[109] Sitarz, A. “3 1/2 Lectures on Noncommutative Geometry”. In: *Acta Polytechnica* vol. 48 (2008), pages 34–55 (cited on pages 68–71, 358).

[110] Sontz, Stephen Bruce. *Principal Bundles*. Universitext. Cham: Springer International Publishing, 2015. Web-link (cited on pages 69, 70, 358).

[111] Baehr, H. C., Dimakis, A., and Müller-Hoissen, F. "Differential Calculi on Commutative Algebras". In: *Journal of Physics A: Mathematical and General* 28.11 (June 7, 1995), pages 3197–3222. DOI: 10.1088/0305-4470/28/11/019. arXiv: hep-th/9412069. Web-link (cited on pages 69, 70, 358).

[112] Bourbaki, N. *Algebra I: Chapters 1-3*. 1st ed. 1989. 2nd printing 1998 edition. Springer, May 23, 2008 (cited on page 70).

[113] Landi, Giovanni. *An Introduction to Noncommutative Spaces and Their Geometries*. Springer Science & Business Media, July 1, 2003 (cited on pages 70, 71).

[114] Milnor, John and Stasheff, James D. *Characteristic Classes.* First Edition edition. Princeton, N.J: Princeton University Press, Aug. 1, 1974 (cited on page 72).

[115] Lawson, H. Blaine and Michelsohn, Marie-Louise. *Spin Geometry*. Princeton, N.J: Princeton University Press, Feb. 1, 1990 (cited on page 72).

[116] Esposito, Giampiero. *Dirac Operators and Spectral Geometry*. 1 edition. Cambridge ; New York: Cambridge University Press, Sept. 13, 1998 (cited on page 72).

[117] Connes, Alain. "On the Spectral Characterization of Manifolds". In: *arXiv preprint arXiv:0810.2088* (2008). Web-link (cited on page 73).

References for Chapter 6: Noncommutative geometry - physics overview

[101] Suijlekom, Walter van. *Noncommutative Geometry and Particle Physics*. Springer, July 21, 2014 (cited on pages 61, 65, 73, 74, 78, 81, 84, 89–93, 95, 358).

[118] Nakahara, Mikio. *Geometry, Topology and Physics*. CRC Press, 2003. Web-link (cited on pages 78, 340).

[119] Connes, Alain and Marcolli, Matilde. *Noncommutative Geometry, Quantum Fields and Motives*. American Mathematical Soc. (cited on pages 78, 81, 84, 85, 87, 93, 95).

[120] Chamseddine, Ali H. and Connes, Alain. "The Spectral Action Principle". In: *Communications in Mathematical Physics* 186.3 (1997), pages 731–750. Web-link (cited on pages 79, 88).

[121] Chamseddine, Ali H., Connes, Alain, and Mukhanov, Viatcheslav. "Geometry and the Quantum: Basics". In: *Journal of High Energy Physics* 2014.12 (2014), pages 1–25. Web-link (cited on pages 79, 94, 268).

[122] Chamseddine, Ali H., Connes, Alain, and Mukhanov, Viatcheslav. "Quanta of Geometry: Noncommutative Aspects". In: *Physical review letters* 114.9 (2015), page 091302. Web-link (cited on pages 79, 94, 268).

[123] Chamseddine, Ali H., Connes, Alain, and van Suijlekom, Walter D. "Grand Unification in the Spectral Pati-Salam Model". In: (July 29, 2015). arXiv: 1507.08161. Web-link (cited on pages 79, 93, 94, 129, 268).

[124] Chamseddine, A. H., Felder, G., and Frohlich, J. "Unified Gauge Theories in Non-Commutative Geometry". In: *Physics Letters B* 296.1 (1992), pages 109–116. Web-link (cited on page 81).

[125] Masson, Thierry. "Gauge Theories in Noncommutative Geometry". In: (2012), pages 73–98. DOI: 10.1063/1.4727990. arXiv: 1201.3345. Web-link (cited on page 87).

[126] Buck, Michel, Fairbairn, Malcolm, and Sakellariadou, Mairi. "Inflation in Models with Conformally Coupled Scalar Fields: An Application to the Noncommutative Spectral Action". In: *Physical Review D* 82.4 (Aug. 6, 2010). DOI: 10.1103/PhysRevD.82.043509. arXiv: 1005.1188. Web-link (cited on page 91).

[127] Chamseddine, Ali H. and Connes, Alain. “Why the Standard Model”. In: *Journal of Geometry and Physics* 58.1 (2008), pages 38–47. Web-link (cited on page 92).

[128] Chamseddine, Ali H., Connes, Alain, and van Suijlekom, Walter D. “Inner Fluctuations in Noncommutative Geometry without the First Order Condition”. In: *Journal of Geometry and Physics* 73 (2013), pages 222–234. Web-link (cited on page 93).

[129] Pati, Jogesh C. and Salam, Abdus. “Lepton Number as the Fourth "color"”. In: *Physical Review D* 10.1 (July 1, 1974), pages 275–289. DOI: 10.1103/PhysRevD.10.275. Web-link (cited on pages 93, 129).

[130] Chamseddine, Ali H., Connes, Alain, and van Suijlekom, Walter D. “Beyond the Spectral Standard Model: Emergence of Pati-Salam Unification”. In: *arXiv preprint arXiv:1304.8050* (2013). Web-link (cited on pages 93, 129).

[131] Devastato, Agostino, Lizzi, Fedele, and Martinetti, Pierre. “Grand Symmetry, Spectral Action, and the Higgs Mass”. In: *arXiv preprint arXiv:1304.0415* (2013). Web-link (cited on page 93).

[132] Devastato, Agostino. “Noncommutative Geometry, Grand Symmetry and Twisted Spectral Triple”. In: *Journal of Physics: Conference Series*. Volume 634. IOP Publishing, 2015, page 012008. Web-link (cited on page 93).

[133] Boyle, Latham and Farnsworth, Shane. “Non-Commutative Geometry, Non-Associative Geometry and the Standard Model of Particle Physics”. In: *New Journal of Physics* 16.12 (2014), page 123027. Web-link (cited on page 93).

[134] Boyle, Latham and Farnsworth, Shane. “A New Algebraic Structure in the Standard Model of Particle Physics”. In: *arXiv preprint arXiv:1604.00847* (2016). Web-link (cited on page 93).

[135] Farnsworth, Shane and Boyle, Latham. “Non-Associative Geometry and the Spectral Action Principle”. In: *arXiv preprint arXiv:1303.1782* (2013). Web-link (cited on page 93).

[136] Farnsworth, Shane and Boyle, Latham. “Rethinking Connes’ Approach to the Standard Model of Particle Physics via Non-Commutative Geometry”. In: *New Journal of Physics* 17.2 (2015), page 023021. Web-link (cited on page 93).

References for Chapter 7: Noncommutative geometry - physics details

[101] Suijlekom, Walter van. *Noncommutative Geometry and Particle Physics*. Springer, July 21, 2014 (cited on pages 61, 65, 73, 74, 78, 81, 84, 89–93, 95, 358).

[102] Beenakker, Wim, van den Broek, Thijs, and van Suijlekom, Walter D. *Supersymmetry and Noncommutative Geometry*. Springer, 2015. Web-link (cited on pages 61, 95).

[119] Connes, Alain and Marcolli, Matilde. *Noncommutative Geometry, Quantum Fields and Motives*. American Mathematical Soc. (cited on pages 78, 81, 84, 85, 87, 93, 95).

[123] Chamseddine, Ali H., Connes, Alain, and van Suijlekom, Walter D. “Grand Unification in the Spectral Pati-Salam Model”. In: (July 29, 2015). arXiv: 1507.08161. Web-link (cited on pages 79, 93, 94, 129, 268).

[129] Pati, Jogesh C. and Salam, Abdus. “Lepton Number as the Fourth "color"”. In: *Physical Review D* 10.1 (July 1, 1974), pages 275–289. DOI: 10.1103/PhysRevD.10.275. Web-link (cited on pages 93, 129).

[130] Chamseddine, Ali H., Connes, Alain, and van Suijlekom, Walter D. “Beyond the Spectral Standard Model: Emergence of Pati-Salam Unification”. In: *arXiv preprint arXiv:1304.8050* (2013). Web-link (cited on pages 93, 129).

[137] Jord, Boeijink. “Noncommutative Geometry of Yang-Mills Fields”. 2009. Web-link (cited on page 95).

[138] Dungen, Koen van den. “The Structure of Gauge Theory in Almost Commutative Geometries”. 2011. Web-link (cited on page 95).

[139] Chamseddine, Ali H., Connes, Alain, Marcolli, Matilde, et al. “Gravity and the Standard Model with Neutrino Mixing”. In: *Advances in Theoretical and Mathematical Physics* 11.6 (2007), pages 991–1089. Web-link (cited on page 95).

[140] Chamseddine, Ali H. and Connes, Alain. “Resilience of the Spectral Standard Model”. In: *Journal of High Energy Physics* 2012.9 (2012), pages 1–11. Web-link (cited on page 128).

References for Chapter 8: Loop quantum gravity

[12] Bongaarts, Peter. *Quantum Theory*. Springer, 2015. Web-link (cited on pages 13, 20–26, 148, 149).

[13] Strocchi, F. *An Introduction to the Mathematical Structure of Quantum Mechanics: A Short Course for Mathematicians*. World Scientific, 2008 (cited on pages 13, 14, 20–23, 26, 27, 149, 353).

[14] Faddeev, L. D. and Yakubovsky, O. A. *Lectures in Quantum Mechanics for Mathematician Students*. St.-Petersburg University Press, 1980 (cited on pages 13, 20, 22, 26, 27, 149).

[25] Thiemann, Thomas. *Modern Canonical Quantum General Relativity*. Cambridge University Press, 2007. Web-link (cited on pages 20, 134, 135, 149, 150, 152, 153, 155, 158–160, 162–167, 169–175, 181, 185, 194, 195, 271).

[26] Takhtadzhian, Leon Armenovich. *Quantum Mechanics for Mathematicians*. American Mathematical Soc., 2008 (cited on pages 20, 27, 135).

[27] Greiner, Walter and Reinhardt, Joachim. *Field Quantization*. Springer Science & Business Media, June 29, 2013 (cited on pages 20, 135, 358).

[28] Rothe, Heinz J. and Rothe, Klaus D. *Classical and Quantum Dynamics of Constrained Hamiltonian Systems*. Volume 81. World Scientific, 2010. Web-link (cited on pages 20, 135).

[29] Henneaux, Marc and Teitelboim, Claudio. *Quantization of Gauge Systems*. Princeton university press, 1992. Web-link (cited on pages 20, 135).

[45] Wald, Robert M. *General Relativity*. University of Chicago Press, May 15, 2010 (cited on pages 30, 33, 36–38, 44, 53, 68, 134, 244, 331).

[77] Rovelli, Carlo. *Quantum Gravity*. Cambridge university press, 2004. Web-link (cited on pages 46, 52, 53, 134, 153, 168, 181, 185, 194).

[93] Rovelli, Carlo. “Forget Time”. In: *Foundations of Physics* 41.9 (2011), pages 1475–1490. Web-link (cited on pages 53, 185).

[104] Morris, Sidney A. *Topology without Tears*. University of New England, 1989. Web-link (cited on pages 61, 170, 327).

[141] Freedman, Daniel Z. and Proeyen, Antoine Van. *Supergravity*. Cambridge University Press, Apr. 5, 2012 (cited on pages 134, 205).

[142] Green, Michael B., Schwarz, John H., and Witten, Edward. *Superstring Theory: Volume 1, Introduction*. Cambridge University Press, July 29, 1988 (cited on pages 134, 200, 206, 211, 222, 226, 229, 242, 243, 259).

[143] Polchinski, Joseph. *String Theory: Volume 1, An Introduction to the Bosonic String*. Cambridge University Press, Oct. 13, 1998 (cited on pages 134, 223, 224).

[144] Bodendorfer, Norbert. "A Note on Quantum Supergravity and AdS/CFT". In: *arXiv preprint arXiv:1509.02036* (2015). Web-link (cited on pages 134, 194, 252, 274, 281, 290).

[145] Date, Ghanashyam and Hossain, Golam Mortuza. "Matter in Loop Quantum Gravity". In: *Symmetry, Integrability and Geometry: Methods and Applications* 8 (2012), pages 10–26. Web-link (cited on pages 134, 273).

[146] Rovelli, Carlo and Vidotto, Francesca. *Covariant Loop Quantum Gravity: An Elementary Introduction to Quantum Gravity and Spinfoam Theory*. Cambridge University Press, Nov. 13, 2014 (cited on pages 134, 171, 273).

[147] Cianfrani, Francesco et al. *Canonical Quantum Gravity: Fundamentals and Recent Developments*. World Scientific Publishing Company, 2014. Web-link (cited on pages 134, 150, 158, 187).

[148] Ashtekar, Abhay and Singh, Parampreet. "Loop Quantum Cosmology: A Status Report". In: *Classical and Quantum Gravity* 28.21 (2011), page 213001. Web-link (cited on pages 135, 187, 188).

[149] Perez, Alejandro. "The Spin Foam Approach to Quantum Gravity". In: *Living Rev. Rel* 16.3 (2013), pages 1205–2019. Web-link (cited on pages 145, 272).

[150] Giulini, Domenico. "That Strange Procedure Called Quantisation". In: 631 (2003), pages 17–40. DOI: 10.1007/978-3-540-45230-0_2. arXiv: quant-ph/0304202. Web-link (cited on pages 148, 149).

[151] Giachetta, G., Mangiarotti, L., and Sardanashvili, Gennadi Aleksandrovich. *Geometric and Algebraic Topological Methods in Quantum Mechanics*. World Scientific, Jan. 2005 (cited on page 149).

[152] Giesel, Kristina and Sahlmann, Hanno. "From Classical to Quantum Gravity: Introduction to Loop Quantum Gravity". In: *arXiv preprint arXiv:1203.2733* (2012). Web-link (cited on pages 152, 177).

[153] Ashtekar, Abhay. "New Variables for Classical and Quantum Gravity". In: *Physical Review Letters* 57.18 (1986), page 2244. Web-link (cited on pages 153, 195).

[154] Ashtekar, Abhay. "New Hamiltonian Formulation of General Relativity". In: *Physical Review D* 36.6 (1987), page 1587. Web-link (cited on pages 153, 195).

[155] Jacobson, Ted and Smolin, Lee. "Nonperturbative Quantum Geometries". In: *Nuclear Physics B* 299.2 (1988), pages 295–345. Web-link (cited on pages 153, 195).

[156] Rovelli, Carlo and Smolin, Lee. "Loop Space Representation of Quantum General Relativity". In: *Nuclear Physics B* 331.1 (1990), pages 80–152. Web-link (cited on pages 153, 195).

[157] Perez, Alejandro. "Introduction to Loop Quantum Gravity and Spin Foams". In: *arXiv preprint gr-qc/0409061* (2004). Web-link (cited on pages 153, 160, 163, 165).

[158] Chiou, Dah-Wei. "Loop Quantum Gravity". In: *International Journal of Modern Physics D* 24 (01 Dec. 28, 2014), page 1530005. DOI: 10.1142/S0218271815300050. Web-link (cited on pages 153, 163, 175, 177, 187).

[159] Mercuri, Simone. "Introduction to Loop Quantum Gravity". In: *arXiv preprint arXiv:1001.1330* (2010). Web-link (cited on page 153).

[160] Gourgoulhon, Éric. "3+ 1 Decomposition of Einstein Equation". In: *3+ 1 Formalism in General Relativity*. Springer, 2012, pages 73–99. Web-link (cited on pages 153, 157–159).

[161] Geroch, Robert. "Domain of Dependence". In: *Journal of Mathematical Physics* 11.2 (Feb. 1, 1970), pages 437–449. DOI: 10.1063/1.1665157. Web-link (cited on page 153).

[162] Lawrie, Ian D. *A Unified Grand Tour of Theoretical Physics*. CRC Press, 2012. Web-link (cited on pages 155, 159, 162).

[163] Baez, John and Muniain, Javier P. *Gauge Fields, Knots and Gravity*. Singapore ; River Edge, N.J: World Scientific Publishing Company, Oct. 24, 1994 (cited on page 158).

[164] Yepez, Jeffrey. "Einstein's Vierbein Field Theory of Curved Space". In: *arXiv preprint arXiv:1106.2037* (2011). Web-link (cited on pages 161, 162).

[165] Gambini, Rodolfo and Pullin, Jorge. *A First Course in Loop Quantum Gravity*. Oxford University Press, 2011. Web-link (cited on pages 161, 185, 271).

[166] Bodendorfer, Norbert. "An Elementary Introduction to Loop Quantum Gravity". In: (July 18, 2016). arXiv: 1607.05129. Web-link (cited on page 163).

[167] Holst, Sören. "Barbero's Hamiltonian Derived from a Generalized Hilbert-Palatini Action". In: *Physical Review D* 53.10 (1996), page 5966. Web-link (cited on page 163).

[168] Ashtekar, Abhay and Lewandowski, Jerzy. "Background Independent Quantum Gravity: A Status Report". In: *Classical and Quantum Gravity* 21.15 (2004), R53. Web-link (cited on page 164).

[169] Han, Muxin, Ma, Yongge, and Huang, Weiming. "Fundamental Structure of Loop Quantum Gravity". In: *International Journal of Modern Physics D* 16 (09 Sept. 1, 2007), pages 1397–1474. DOI: 10.1142/S0218271807010894. Web-link (cited on pages 166, 167, 170–172, 174, 175).

[170] Nicolai, Hermann, Peeters, Kasper, and Zamaklar, Marija. "Loop Quantum Gravity: An Outside View". In: *Classical and Quantum Gravity* 22.19 (Oct. 7, 2005), R193–R247. DOI: 10.1088/0264-9381/22/19/R01. arXiv: hep-th/0501114. Web-link (cited on page 168).

[171] Ashtekar, Abhay, Corichi, Alejandro, and Zapata, Jose A. "Quantum Theory of Geometry III: Non-Commutativity of Riemannian Structures". In: *Classical and Quantum Gravity* 15.10 (Oct. 1, 1998), pages 2955–2972. DOI: 10.1088/0264-9381/15/10/006. arXiv: gr-qc/9806041. Web-link (cited on pages 169, 172).

[172] Flori, Cecilia. "Approaches To Quantum Gravity". In: (Nov. 11, 2009). arXiv: 0911.2135. Web-link (cited on pages 170–172, 174, 175).

[173] Achour, Jibril Ben. "Towards Self Dual Loop Quantum Gravity". In: (Nov. 23, 2015). arXiv: 1511.07332. Web-link (cited on pages 172, 175).

[174] Doná, Pietro and Speziale, Simone. "Introductory Lectures to Loop Quantum Gravity". In: (July 2, 2010). arXiv: 1007.0402. Web-link (cited on pages 175–178, 181, 182).

[175] Thiemann, T. "Quantum Spin Dynamics (QSD)". In: *Classical and Quantum Gravity* 15.4 (Apr. 1, 1998), pages 839–873. DOI: 10.1088/0264-9381/15/4/011. arXiv: gr-qc/9606089. Web-link (cited on pages 183, 195).

[176] Page, Don N. and Wootters, William K. "Evolution without Evolution: Dynamics Described by Stationary Observables". In: *Physical Review D* 27.12 (1983), page 2885. Web-link (cited on page 185).

[177] Gambini, Rodolfo et al. “Conditional Probabilities with Dirac Observables and the Problem of Time in Quantum Gravity”. In: *Physical Review D* 79.4 (2009), page 041501. Web-link (cited on pages 185, 186).

[178] Gambini, Rodolfo and Pullin, Jorge. “The Montevideo Interpretation of Quantum Mechanics: A Short Review”. In: (Feb. 11, 2015). arXiv: 1502.03410. Web-link (cited on page 185).

[179] Carrozza, Sylvain. “The Problem of Time in Quantum Gravity”. Web-link (cited on page 185).

[180] Banerjee, Kinjal. “Introduction to Loop Quantum Cosmology”. In: *Symmetry, Integrability and Geometry: Methods and Applications* (Mar. 25, 2012). DOI: 10.3842/SIGMA.2012.016. Web-link (cited on page 187).

[181] Ashtekar, Abhay, Pawlowski, Tomasz, and Singh, Parampreet. “Quantum Nature of the Big Bang”. In: *Physical review letters* 96.14 (2006), page 141301. Web-link (cited on page 193).

[182] Bonzom, Valentin and Dittrich, Bianca. “Dirac’s Discrete Hypersurface Deformation Algebras”. In: *Classical and Quantum Gravity* 30.20 (Oct. 21, 2013), page 205013. DOI: 10.1088/0264-9381/30/20/205013. arXiv: 1304.5983. Web-link (cited on page 192).

[183] Blohmann, Christian, Fernandes, Marco Cezar Barbosa, and Weinstein, Alan. “Groupoid Symmetry and Constraints in General Relativity”. In: *Communications in Contemporary Mathematics* 15 (01 Feb. 2013), page 1250061. DOI: 10.1142/S0219199712500617. arXiv: 1003.2857. Web-link (cited on page 192).

[184] Bojowald, Martin et al. “Hypersurface-Deformation Algebroids and Effective Space-Time Models”. In: (Oct. 26, 2016). arXiv: 1610.08355. Web-link (cited on page 192).

[185] Bojowald, Martin and Mielczarek, Jakub. “Some Implications of Signature-Change in Cosmological Models of Loop Quantum Gravity”. In: *Journal of Cosmology and Astroparticle Physics* 2015 (08 2015), page 052. Web-link (cited on pages 192, 273).

[186] Barrau, Aurélien and Grain, Julien. “Cosmology without Time: What to Do with a Possible Signature Change from Quantum Gravitational Origin?” In: (July 26, 2016). arXiv: 1607.07589. Web-link (cited on pages 192, 273).

[187] Bojowald, Martin and Brahma, Suddhasattwa. “Signature Change in 2-Dimensional Black-Hole Models of Loop Quantum Gravity”. In: (Oct. 27, 2016). arXiv: 1610.08850. Web-link (cited on page 192).

[188] Bojowald, Martin and Brahma, Suddhasattwa. “Signature Change in Loop Quantum Gravity: General Midisuperspace Models and Dilaton Gravity”. In: (Oct. 27, 2016). arXiv: 1610.08840. Web-link (cited on page 192).

[189] Bodendorfer, Norbert, Thiemann, Thomas, and Thurn, Andreas. “New Variables for Classical and Quantum Gravity in All Dimensions: I. Hamiltonian Analysis”. In: *Classical and Quantum Gravity* 30.4 (2013), page 045001. Web-link (cited on pages 193–195).

[190] Bodendorfer, Norbert, Thiemann, Thomas, and Thurn, Andreas. “New Variables for Classical and Quantum Gravity in All Dimensions: IV. Matter Coupling”. In: *Classical and Quantum Gravity* 30.4 (2013), page 045004. Web-link (cited on pages 193, 195).

[191] Bodendorfer, Norbert, Thiemann, Thomas, and Thurn, Andreas. “New Variables for Classical and Quantum Gravity in All Dimensions: II. Lagrangian Analysis”. In: *Classical and Quantum Gravity* 30.4 (2013), page 045002. Web-link (cited on pages 193, 195).

[192] Bodendorfer, Norbert, Thiemann, Thomas, and Thurn, Andreas. "New Variables for Classical and Quantum Gravity in All Dimensions: III. Quantum Theory". In: *Classical and Quantum Gravity* 30.4 (2013), page 045003. Web-link (cited on pages 193, 195).

[193] Bodendorfer, Norbert, Thiemann, Thomas, and Thurn, Andreas. "Towards Loop Quantum Supergravity (LQSG): I. Rarita–Schwinger Sector". In: *Classical and Quantum Gravity* 30.4 (2013), page 045006. Web-link (cited on pages 193, 195).

[194] Bodendorfer, Norbert, Thiemann, Thomas, and Thurn, Andreas. "Towards Loop Quantum Supergravity (LQSG): II. P-Form Sector". In: *Classical and Quantum Gravity* 30.4 (2013), page 045007. Web-link (cited on pages 193, 195).

[195] Einstein, Albert. "Näherungsweise Integration Der Feldgleichungen Der Gravitation". In: *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 688-696.* (1916). Web-link (cited on page 194).

[196] Dirac, Paul AM. "Generalized Hamiltonian Dynamics". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Volume 246. The Royal Society, 1958, pages 326–332. Web-link (cited on page 194).

[197] Dirac, Paul AM. "The Theory of Gravitation in Hamiltonian Form". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Volume 246. The Royal Society, 1958, pages 333–343. Web-link (cited on page 194).

[198] Bergmann, Peter G. "Conservation Laws in General Relativity as the Generators of Coordinate Transformations". In: *Physical Review* 112.1 (1958), page 287. Web-link (cited on page 194).

[199] Bergmann, Peter G. "Observables in General Relativity". In: *Reviews of Modern Physics* 33.4 (1961), page 510. Web-link (cited on page 194).

[200] DeWitt, Bryce S. "Quantum Theory of Gravity. I. The Canonical Theory". In: *Physical Review* 160.5 (1967), page 1113. Web-link (cited on page 194).

[201] DeWitt, Bryce S. "Quantum Theory of Gravity. II. The Manifestly Covariant Theory". In: *Physical Review* 162.5 (1967), page 1195. Web-link (cited on page 194).

[202] DeWitt, Bryce S. "Quantum Theory of Gravity. III. Applications of the Covariant Theory". In: *Physical Review* 162.5 (1967), page 1239. Web-link (cited on page 194).

[203] Rovelli, Carlo and Smolin, Lee. "Knot Theory and Quantum Gravity". In: *Physical Review Letters* 61.10 (1988), page 1155. Web-link (cited on page 195).

[204] Rovelli, Carlo and Smolin, Lee. "Spin Networks and Quantum Gravity". In: *Physical Review D* 52.10 (1995), page 5743. Web-link (cited on page 195).

[205] Thiemann, T. "Quantum Spin Dynamics (QSD) II". In: *Classical and Quantum Gravity* 15.4 (Apr. 1, 1998), pages 875–905. DOI: 10.1088/0264-9381/15/4/012. arXiv: gr-qc/9606090. Web-link (cited on page 195).

[206] Thiemann, Thomas. "QSD III : Quantum Constraint Algebra and Physical Scalar Product in Quantum General Relativity". In: *Classical and Quantum Gravity* 15.5 (May 1, 1998), pages 1207–1247. DOI: 10.1088/0264-9381/15/5/010. arXiv: gr-qc/9705017. Web-link (cited on page 195).

[207] Sahlmann, Hanno and Thiemann, Thomas. "Irreducibility of the Ashtekar-Isham-Lewandowski Representation". In: *Classical and Quantum Gravity* 23.13 (July 1, 2006), pages 4453–4471. DOI: 10.1088/0264-9381/23/13/010. arXiv: gr-qc/0303074. Web-link (cited on page 195).

References for Chapter 9: String theory

[11] Carr, Bernard. *Universe Or Multiverse?* Cambridge University Press, June 21, 2007 (cited on pages 9, 240, 241).

[35] Nielsen, Michael A. and Chuang, Isaac L. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Dec. 9, 2010 (cited on pages 26, 254, 256).

[43] Mandl, Franz and Shaw, Graham. *Quantum Field Theory*. John Wiley & Sons, June 6, 2013 (cited on pages 29, 31, 43, 246).

[44] Weinberg, Steven. *The Quantum Theory of Fields*. Volume 1. Cambridge University Press, Aug. 13, 1996 (cited on pages 30, 31, 38, 40–42, 44, 204, 242, 371).

[45] Wald, Robert M. *General Relativity*. University of Chicago Press, May 15, 2010 (cited on pages 30, 33, 36–38, 44, 53, 68, 134, 244, 331).

[141] Freedman, Daniel Z. and Proeyen, Antoine Van. *Supergravity*. Cambridge University Press, Apr. 5, 2012 (cited on pages 134, 205).

[142] Green, Michael B., Schwarz, John H., and Witten, Edward. *Superstring Theory: Volume 1, Introduction*. Cambridge University Press, July 29, 1988 (cited on pages 134, 200, 206, 211, 222, 226, 229, 242, 243, 259).

[143] Polchinski, Joseph. *String Theory: Volume 1, An Introduction to the Bosonic String*. Cambridge University Press, Oct. 13, 1998 (cited on pages 134, 223, 224).

[144] Bodendorfer, Norbert. "A Note on Quantum Supergravity and AdS/CFT". In: *arXiv preprint arXiv:1509.02036* (2015). Web-link (cited on pages 134, 194, 252, 274, 281, 290).

[208] Becker, Katrin, Becker, Melanie, and Schwarz, John H. *String Theory and M-Theory: A Modern Introduction*. Cambridge University Press, Dec. 7, 2006 (cited on pages 200, 218, 222, 227, 229, 249).

[209] Green, Michael B., Schwarz, John H., and Witten, Edward. *Superstring Theory: Volume 2, Loop Amplitudes, Anomalies and Phenomenology*. Cambridge University Press, 1987 (cited on pages 200, 226, 234, 236).

[210] Zwiebach, Barton. *A First Course in String Theory*. Cambridge University Press, Jan. 22, 2009 (cited on pages 200, 230, 233, 238, 239).

[211] Blumenhagen, Ralph, Lüst, Dieter, and Theisen, Stefan. *Basic Concepts of String Theory*. Springer Science & Business Media, Oct. 3, 2012 (cited on pages 200, 207–211, 214–218, 221, 222, 225, 226, 233, 234, 236–238, 242).

[212] Schwarz, John H. "The Early History of String Theory and Supersymmetry". In: (Jan. 4, 2012). arXiv: 1201.0981. Web-link (cited on pages 202, 204).

[213] Overduin, J. M. and Wesson, P. S. "Kaluza-Klein Gravity". In: *Physics Reports* 283 (5-6 Apr. 1997), pages 303–378. DOI: 10.1016/S0370-1573(96)00046-4. arXiv: gr-qc/9805018. Web-link (cited on pages 202, 205).

[214] Arnowitt, R., Chamseddine, A. H., and Nath, Pran. "The Development of Supergravity Grand Unification: Circa 1982-85". In: *International Journal of Modern Physics A* 27.27 (Oct. 30, 2012), page 1230028. DOI: 10.1142/S0217751X12300281. arXiv: 1206.3175. Web-link (cited on page 202).

[215] Duff, M. J., Nilsson, B. E. W., and Pope, C. N. "Kaluza-Klein Supergravity". In: *Physics Reports* 130.1 (Jan. 1, 1986), pages 1–142. DOI: 10.1016/0370-1573(86)90163-8. Web-link (cited on page 202).

[216] Kaluza, Theodor. “Zum Unitätsproblem Der Physik”. In: *Sitzungsber. Preuss. Akad. Wiss. Berlin.(Math. Phys.)* 1921.966972 (1921), page 45. Web-link (cited on page 202).

[217] Klein, Oskar. “Quantentheorie Und Fünfdimensionale Relativitätstheorie”. In: *Zeitschrift für Physik* 37.12 (1926), pages 895–906. Web-link (cited on page 203).

[218] DeWitt, Bryce Seligman. *Dynamical Theory of Groups and Fields*. Gordon and Breach, 1965 (cited on page 203).

[219] Hoddeson, Lillian. *The Rise of the Standard Model: A History of Particle Physics from 1964 to 1979*. Cambridge University Press, Nov. 13, 1997 (cited on page 203).

[220] Witten, Edward. “Search for a Realistic Kaluza-Klein Theory”. In: *Nuclear Physics B* 186.3 (Aug. 10, 1981), pages 412–428. DOI: 10.1016/0550-3213(81)90021-3. Web-link (cited on page 204).

[221] Fecko, Marián. *Differential Geometry and Lie Groups for Physicists*. Cambridge University Press, Oct. 12, 2006 (cited on page 204).

[222] Garling, D. J. H. *Clifford Algebras: An Introduction*. Cambridge University Press, June 23, 2011 (cited on page 204).

[223] Witten, Edward. “Fermion Quantum Numbers in Kaluza-Klein Theory”. In: *et al: Modern Kaluza-Klein Theories* (1983), pages 438–511. Web-link (cited on page 204).

[224] Ramond, P. “Dual Theory for Free Fermions”. In: *Physical Review D* 3.10 (May 15, 1971), pages 2415–2418. DOI: 10.1103/PhysRevD.3.2415. Web-link (cited on pages 204, 259).

[225] Gervais, J. -L. and Sakita, B. “Field Theory Interpretation of Supergauges in Dual Models”. In: *Nuclear Physics B* 34.2 (Nov. 15, 1971), pages 632–639. DOI: 10.1016/0550-3213(71)90351-8. Web-link (cited on page 204).

[226] Freedman, Daniel Z., Van Nieuwenhuizen, P., and Ferrara, Sergio. “Progress toward a Theory of Supergravity”. In: *Physical Review D* 13.12 (1976), page 3214. Web-link (cited on pages 204, 260).

[227] Nahm, W. “Supersymmetries and Their Representations”. In: *Nuclear Physics B* 135.1 (Mar. 27, 1978), pages 149–166. DOI: 10.1016/0550-3213(78)90218-3. Web-link (cited on page 204).

[228] Collins, P. D. B., Martin, Alan D., and Squires, E. J. *Particle Physics and Cosmology*. Wiley, June 6, 1989 (cited on pages 205, 259).

[229] Tanii, Yoshiaki. *Introduction to Supergravity*. Volume 1. SpringerBriefs in Mathematical Physics. Tokyo: Springer Japan, 2014. Web-link (cited on page 205).

[230] Deser, Stanley, Kay, J. H., and Stelle, K. S. “Renormalizability Properties of Supergravity”. In: *Physical Review Letters* 38.10 (1977), page 527. Web-link (cited on pages 205, 260).

[231] Van Nieuwenhuizen, P. and Vermaseren, J. A. M. “One-Loop Divergences in the Quantum Theory of Supergravity”. In: *Physics Letters B* 65.3 (Nov. 22, 1976), pages 263–266. DOI: 10.1016/0370-2693(76)90178-7. Web-link (cited on page 205).

[232] Stelle, K.s. “Is N = 8 Supergravity a Finite Field Theory?” In: *Fortschritte der Physik* 57 (5-7 June 4, 2009), pages 446–450. DOI: 10.1002/prop.200900063. Web-link (cited on page 205).

[233] Green, Michael B. and Schwarz, John H. “Anomaly Cancellations in Supersymmetric D = 10 Gauge Theory and Superstring Theory”. In: *Physics Letters B* 149.1 (Dec. 13, 1984), pages 117–122. DOI: 10.1016/0370-2693(84)91565-X. Web-link (cited on pages 205, 260).

[234] Gross, David J. et al. “Heterotic String Theory (I). The Free Heterotic String”. In: *Nuclear Physics B* 256 (1985), pages 253–284. Web-link (cited on pages 205, 218, 260).

[235] Gross, David J. et al. “Heterotic String”. In: *Physical Review Letters* 54.6 (1985), page 502. Web-link (cited on pages 205, 218, 260).

[236] Gross, David J. et al. “Heterotic String theory:(II). The Interacting Heterotic String”. In: *Nuclear Physics B* 267.1 (1986), pages 75–124. Web-link (cited on pages 205, 218, 260).

[237] Candelas, P. et al. “Vacuum Configurations for Superstrings”. In: *Nuclear Physics B* 258 (Jan. 1, 1985), pages 46–74. DOI: 10.1016/0550-3213(85)90602-9. Web-link (cited on pages 205, 206, 233, 236, 260, 279).

[238] Polchinski, Joseph. “Dirichlet-Branes and Ramond-Ramond Charges”. In: *Physical Review Letters* 75.26 (Dec. 25, 1995), pages 4724–4727. DOI: 10.1103/PhysRevLett.75.4724. arXiv: hep-th/9510017. Web-link (cited on pages 206, 245, 248, 260).

[239] Witten, Edward. “String Theory Dynamics in Various Dimensions”. In: *Nuclear Physics B* 443.1 (June 5, 1995), pages 85–126. DOI: 10.1016/0550-3213(95)00158-O. Web-link (cited on pages 206, 229, 230, 260).

[240] Duff, M. J. *The World in Eleven Dimensions: Supergravity, Supermembranes and M-Theory*. 1 edition. Bristol; Philadelphia: CRC Press, Jan. 1, 1999 (cited on page 206).

[241] Kaku, Michio. *Introduction to Superstrings*. Corr., 2. print. Graduate texts in contemporary physics. New York Berlin: Springer, 1990 (cited on pages 206, 240).

[242] Schottenloher, Martin. *A Mathematical Introduction to Conformal Field Theory*. Springer Science & Business Media, Sept. 26, 2008 (cited on pages 211, 225–227, 275).

[243] West, Peter. *Introduction to Strings and Branes*. Cambridge University Press, July 5, 2012 (cited on pages 214, 261).

[244] Ortín, Tomás. *Gravity and Strings*. Cambridge University Press, Jan. 29, 2007 (cited on page 214).

[245] Belopolsky, Alexander. “New Geometrical Approach to Superstrings”. In: (Mar. 25, 1997). arXiv: hep-th/9703183. Web-link (cited on page 217).

[246] Cornwell, J. F. *Group Theory in Physics - Volume 3*. Academic Press, 1989 (cited on pages 218, 261, 344).

[247] Carmeli, Claudio, Caston, Lauren, and Fioresi, Rita. *Mathematical Foundations of Supersymmetry*. European Mathematical Society, 2011 (cited on page 218).

[248] Gliozzi, F., Scherk, J., and Olive, D. “Supersymmetry, Supergravity Theories and the Dual Spinor Model”. In: *Nuclear Physics B* 122.2 (Apr. 25, 1977), pages 253–290. DOI: 10.1016/0550-3213(77)90206-1. Web-link (cited on page 221).

[249] Tanedo, Philip. “Lectures on String Theory”. 2010. Web-link (cited on pages 221, 222).

[250] Tong, David. “Lectures on String Theory”. In: (Aug. 3, 2009). arXiv: 0908.0333. Web-link (cited on pages 222, 223, 225, 372).

[251] Staessens, Wieland and Vercnocke, Bert. “Lectures on Scattering Amplitudes in String Theory”. In: (Nov. 1, 2010). arXiv: 1011.0456. Web-link (cited on pages 223, 225).

[252] Blumenhagen, Ralph and Plauschinn, Erik. *Introduction to Conformal Field Theory: With Applications to String Theory*. Springer Science & Business Media, July 31, 2009 (cited on page 225).

[253] Francesco, Philippe Di, Mathieu, Pierre, and Sénéchal, David. *Conformal Field Theory*. Springer Science & Business Media, 1997 (cited on page 225).

[254] Witten, Edward. “What Every Physicist Should Know about”. In: *Physics Today* 11 (2015), page 38. Web-link (cited on pages 226, 227).

[255] *String Theory FAQ in nLab*. Web-link (cited on page 226).

[256] Bern, Zvi and Kosower, David A. “Efficient Calculation of One-Loop QCD Amplitudes”. In: *Physical Review Letters* 66.13 (Apr. 1, 1991), pages 1669–1672. DOI: 10.1103/PhysRevLett.66.1669. Web-link (cited on page 226).

[257] Strassler, Matthew J. “Field Theory Without Feynman Diagrams: One-Loop Effective Actions”. In: *Nuclear Physics B* 385 (1-2 Oct. 1992), pages 145–184. DOI: 10.1016/0550-3213(92)90098-V. arXiv: hep-ph/9205205. Web-link (cited on page 226).

[258] Strassler, Matthew James. “The Bern-Kosower Rules and Their Relation to Quantum Field Theory”. In: (1993). Web-link (cited on page 226).

[259] Roggenkamp, Daniel and Wendland, Katrin. “Limits and Degenerations of Unitary Conformal Field Theories”. In: *Communications in mathematical physics* 251.3 (2004), pages 589–643. Web-link (cited on pages 227, 268, 275, 281).

[260] Roggenkamp, Daniel and Wendland, Katrin. “Decoding the Geometry of Conformal Field Theories”. In: *arXiv preprint arXiv:0803.0657* (2008). Web-link (cited on pages 227, 268, 275, 281).

[261] Polchinski, Joseph. “TASI Lectures on D-Branes”. In: (Nov. 7, 1996). arXiv: hep-th/9611050. Web-link (cited on pages 229, 248, 260).

[262] Townsend, P. K. “The Eleven-Dimensional Supermembrane Revisited”. In: *Physics Letters B* 350.2 (May 1995), pages 184–188. DOI: 10.1016/0370-2693(95)00397-4. arXiv: hep-th/9501068. Web-link (cited on page 229).

[263] Ginsparg, Paul. “On Toroidal Compactification of Heterotic Superstrings”. In: *Physical Review D* 35.2 (Jan. 15, 1987), pages 648–654. DOI: 10.1103/PhysRevD.35.648. Web-link (cited on page 230).

[264] Dai, Jin, Leigh, R.g., and Polchinski, Joseph. “New Connections between String Theories”. In: *Modern Physics Letters A* 04.21 (Oct. 20, 1989), pages 2073–2083. DOI: 10.1142/S0217732389002331. Web-link (cited on pages 230, 260).

[265] Hull, C. M. and Townsend, P. K. “Unity of Superstring Dualities”. In: *Nuclear Physics B* 438 (1-2 Mar. 1995), pages 109–137. DOI: 10.1016/0550-3213(94)00559-W. arXiv: hep-th/9410167. Web-link (cited on page 230).

[266] Horava, Petr and Witten, Edward. “Eleven-Dimensional Supergravity on a Manifold with Boundary”. In: *Nuclear Physics B* 475 (1-2 Sept. 1996), pages 94–114. DOI: 10.1016/0550-3213(96)00308-2. arXiv: hep-th/9603142. Web-link (cited on page 230).

[267] Horava, Petr and Witten, Edward. “Heterotic and Type I String Dynamics from Eleven Dimensions”. In: *Nuclear Physics B* 460.3 (Feb. 1996), pages 506–524. DOI: 10.1016/0550-3213(95)00621-4. arXiv: hep-th/9510209. Web-link (cited on page 230).

[268] Louis, Jan. “Introduction to String Phenomenology”. 2015. Web-link (cited on page 231).

[269] Baaquie, Belal E. and Willeboordse, Frederick H. *Exploring the Invisible Universe: From Black Holes to Superstrings*. World Scientific, Mar. 25, 2015 (cited on pages 231, 232).

[270] Groot-Nibbelink, Stefan. *Some Recent Developments in Non-Supersymmetric String Model Building - - Groot-Nibbelink.pdf*. 2015. Web-link (cited on page 232).

[271] Ibáñez, Luis E. and Uranga, Angel M. *String Theory and Particle Physics: An Introduction to String Phenomenology*. Cambridge University Press, Feb. 9, 2012 (cited on pages 233, 234, 236–240).

[272] Moore, Douglas G. *The Landscape of Free Fermionic Gauge Models*. Springer, Feb. 3, 2016 (cited on page 233).

[273] Scherk, J. and Schwarz, J. H. "Dual Field Theory of Quarks and Gluons". In: *Physics Letters B* 57.5 (Aug. 4, 1975), pages 463–466. DOI: 10.1016/0370-2693(75)90269-5. Web-link (cited on page 233).

[274] Cremmer, E. and Scherk, J. "Dual Models in Four Dimensions with Internal Symmetries". In: *Nuclear Physics B* 103.3 (Feb. 9, 1976), pages 399–425. DOI: 10.1016/0550-3213(76)90507-1. Web-link (cited on page 233).

[275] Randall, Lisa and Sundrum, Raman. "An Alternative to Compactification". In: *Physical Review Letters* 83.23 (Dec. 6, 1999), pages 4690–4693. DOI: 10.1103/PhysRevLett.83.4690. arXiv: hep-th/9906064. Web-link (cited on page 233).

[276] Dixon, L. et al. "Strings on Orbifolds". In: *Nuclear Physics B* 261 (Jan. 1, 1985), pages 678–686. DOI: 10.1016/0550-3213(85)90593-0. Web-link (cited on pages 233, 236).

[277] Dixon, L. et al. "Strings on Orbifolds (II)". In: *Nuclear Physics B* 274.2 (Sept. 15, 1986), pages 285–314. DOI: 10.1016/0550-3213(86)90287-7. Web-link (cited on pages 233, 236).

[278] Sagnotti, Augusto. "Open Strings and Their Symmetry Groups". In: *Nonperturbative Quantum Field Theory*. Springer, 1988, pages 521–528. Web-link (cited on page 234).

[279] Ellis, John. "The Superstring: Theory of Everything, or of Nothing?" In: *Nature* 323.6089 (Oct. 16, 1986), pages 595–598. DOI: 10.1038/323595a0. Web-link (cited on page 235).

[280] Polchinski, Joseph. *String Theory: Volume 2, Superstring Theory and Beyond*. Cambridge University Press, Oct. 13, 1998 (cited on page 237).

[281] Baer, Howard and Tata, Xerxes. *Weak Scale Supersymmetry: From Superfields to Scattering Events*. Cambridge University Press, 2006. Web-link (cited on page 237).

[282] Binetruy, Pierre. *Supersymmetry: Theory, Experiment, and Cosmology*. Oxford University Press, 2012. Web-link (cited on page 237).

[283] Douglas, Michael R. and Kachru, Shamit. "Flux Compactification". In: *Reviews of Modern Physics* 79.2 (May 25, 2007), pages 733–796. DOI: 10.1103/RevModPhys.79.733. arXiv: hep-th/0610102. Web-link (cited on pages 239, 240).

[284] Anderson, Lara B. et al. "Vacuum Varieties, Holomorphic Bundles and Complex Structure Stabilization in Heterotic Theories". In: (Apr. 9, 2013). arXiv: 1304.2704. Web-link (cited on page 240).

[285] Schellekens, A. N. "Life at the Interface of Particle Physics and String Theory". In: *Reviews of Modern Physics* 85.4 (2013), page 1491. Web-link (cited on page 240).

[286] Kaku, Michio. *Strings, Conformal Fields, and Topology: An Introduction*. Softcover reprint of the original 1st ed. 1991 edition. New York, NY: Springer, Jan. 1, 1991 (cited on page 240).

[287] Susskind, Leonard. "The Anthropic Landscape of String Theory". In: (Feb. 27, 2003). arXiv: hep-th/0302219. Web-link (cited on pages 240, 261).

[288] Schellekens, A. N. "The Emperor's Last Clothes?" In: *Reports on Progress in Physics* 71.7 (July 1, 2008), page 072201. DOI: 10.1088/0034-4885/71/7/072201. arXiv: 0807.3249. Web-link (cited on page 240).

[289] Linde, Andrei. "A Brief History of the Multiverse". In: (Dec. 3, 2015). arXiv: 1512.01203. Web-link (cited on page 240).

[290] Linde, Andrei. “Particle Physics and Inflationary Cosmology”. In: *arXiv preprint hep-th/0503203* (2005). Web-link (cited on pages 240, 241).

[291] Polchinski, Joseph. “Why Trust a Theory? Some Further Remarks (Part 1)”. In: (Jan. 22, 2016). arXiv: 1601.06145. Web-link (cited on pages 241, 371).

[292] Schwarzschild, K. “On the Gravitational Field of a Mass Point according to Einstein’s Theory”. In: (1916). arXiv: physics/9905030. Web-link (cited on page 244).

[293] Hawking, S. W. and Ellis, G. F. R. *The Large Scale Structure of Space-Time*. Cambridge University Press, 1973 (cited on page 244).

[294] Hawking, S. W. “Black Holes in General Relativity”. In: *Communications in Mathematical Physics* 25.2 (), pages 152–166. DOI: 10.1007/BF01877517. Web-link (cited on page 244).

[295] Bardeen, J. M., Carter, B., and Hawking, S. W. “The Four Laws of Black Hole Mechanics”. In: *Communications in Mathematical Physics* 31.2 (), pages 161–170. DOI: 10.1007/BF01645742. Web-link (cited on page 244).

[296] Bekenstein, Jacob D. “Black Holes and the Second Law”. In: *Lettere Al Nuovo Cimento (1971–1985)* 4.15 (1972), pages 737–740. Web-link (cited on page 244).

[297] Bekenstein, Jacob D. “Black Holes and Entropy”. In: *Physical Review D* 7.8 (1973), page 2333. Web-link (cited on page 244).

[298] Bekenstein, Jacob D. “Generalized Second Law of Thermodynamics in Black-Hole Physics”. In: *Physical Review D* 9.12 (1974), page 3292. Web-link (cited on page 244).

[299] Hawking, Stephen W. “Black Hole Explosions”. In: *Nature* 248.5443 (1974), pages 30–31. Web-link (cited on page 244).

[300] Hawking, Stephen W. “Particle Creation by Black Holes”. In: *Communications in mathematical physics* 43.3 (1975), pages 199–220. Web-link (cited on page 244).

[301] Unruh, William G. “Notes on Black-Hole Evaporation”. In: *Physical Review D* 14.4 (1976), page 870. Web-link (cited on page 244).

[302] Natsuume, Makoto. *AdS/CFT Duality User Guide*. Springer, Apr. 1, 2015 (cited on pages 244, 245, 247).

[303] Hawking, S. W. “Breakdown of Predictability in Gravitational Collapse”. In: *Physical Review D* 14.10 (Nov. 15, 1976), pages 2460–2473. DOI: 10.1103/PhysRevD.14.2460. Web-link (cited on page 244).

[304] Mathur, Samir D. “The Information Paradox: A Pedagogical Introduction”. In: *Classical and Quantum Gravity* 26.22 (Nov. 21, 2009), page 224001. DOI: 10.1088/0264-9381/26/22/224001. arXiv: 0909.1038. Web-link (cited on pages 244, 251).

[305] Maldacena, Juan M. “The Large N Limit of Superconformal Field Theories and Supergravity”. In: *International Journal of Theoretical Physics* 38.4 (1999), pages 1113–1133. DOI: 10.1023/A:1026654312961. arXiv: hep-th/9711200. Web-link (cited on pages 245, 250, 260).

[306] Nastase, Horatiu. *Introduction to AdS/CFT Correspondence*. Cambridge University Press, Sept. 10, 2015 (cited on pages 245, 250, 251).

[307] Ammon, Martin and Erdmenger, Johanna. *Gauge/Gravity Duality*. Cambridge University Press, Apr. 9, 2015 (cited on page 245).

[308] Zaanen, Jan et al. *Holographic Duality in Condensed Matter Physics*. Cambridge University Press, Nov. 5, 2015 (cited on page 245).

[309] 't Hooft, G. "A Planar Diagram Theory for Strong Interactions". In: *Nuclear Physics B* 72.3 (Apr. 18, 1974), pages 461–473. DOI: 10.1016/0550-3213(74)90154-0. Web-link (cited on pages 245, 246).

[310] 't Hooft, G. "Dimensional Reduction in Quantum Gravity". In: (Oct. 19, 1993). arXiv: gr-qc/9310026. Web-link (cited on page 245).

[311] Susskind, L. "The World as a Hologram". In: *Journal of Mathematical Physics* 36.11 (1995), page 6377. DOI: 10.1063/1.531249. arXiv: hep-th/9409089. Web-link (cited on page 245).

[312] Brown, J. D. and Henneaux, Marc. "Central Charges in the Canonical Realization of Asymptotic Symmetries: An Example from Three Dimensional Gravity". In: *Communications in Mathematical Physics* 104.2 (), pages 207–226. DOI: 10.1007/BF01211590. Web-link (cited on page 245).

[313] Strominger, A. and Vafa, C. "Microscopic Origin of the Bekenstein-Hawking Entropy". In: *Physics Letters B* 379 (1-4 June 1996), pages 99–104. DOI: 10.1016/0370-2693(96)00345-0. arXiv: hep-th/9601029. Web-link (cited on pages 245, 249, 260).

[314] Hayden, Patrick. "Random Codes and Holographic Duality". Web-link (cited on page 246).

[315] Thomson, Mark. *Modern Particle Physics*. Cambridge University Press, Sept. 5, 2013 (cited on page 246).

[316] Rudra, Arnab. "Large N". Web-link (cited on page 247).

[317] Bena, Iosif, El-Showk, Sheer, and Vercnocke, Bert. "Black Holes in String Theory". In: *Black Objects in Supergravity*. Edited by Bellucci, Stefano. Springer Proceedings in Physics 144. Springer International Publishing, 2013, pages 59–178. DOI: 10.1007/978-3-319-00215-6_2. Web-link (cited on pages 248, 249).

[318] Szabo, Richard J. *An Introduction to String Theory and D-Brane Dynamics*. Imperial College Press, 2004 (cited on page 249).

[319] Witten, Edward. "Anti De Sitter Space And Holography". In: (Feb. 20, 1998). arXiv: hep-th/9802150. Web-link (cited on pages 251, 260).

[320] Gubser, S. S., Klebanov, I. R., and Polyakov, A. M. "Gauge Theory Correlators from Non-Critical String Theory". In: *Physics Letters B* 428 (1-2 May 1998), pages 105–114. DOI: 10.1016/S0370-2693(98)00377-3. arXiv: hep-th/9802109. Web-link (cited on pages 251, 260).

[321] Ryu, Shinsei and Takayanagi, Tadashi. "Aspects of Holographic Entanglement Entropy". In: *Journal of High Energy Physics* 2006 (08 2006), page 045. Web-link (cited on pages 251, 255, 256, 261).

[322] Ryu, Shinsei and Takayanagi, Tadashi. "Holographic Derivation of Entanglement Entropy from the Anti–de Sitter Space/Conformal Field Theory Correspondence". In: *Physical review letters* 96.18 (2006), page 181602. Web-link (cited on pages 251, 256, 261).

[323] McGreevy, John. "Holographic Duality with a View toward Many-Body Physics". In: *Advances in High Energy Physics* 2010 (2010), pages 1–54. DOI: 10.1155/2010/723105. arXiv: 0909.0518. Web-link (cited on page 251).

[324] Witten, Edward. "Anti-de Sitter Space, Thermal Phase Transition, And Confinement In Gauge Theories". In: (Mar. 16, 1998). arXiv: hep-th/9803131. Web-link (cited on page 251).

[325] Auletta, Gennaro, Fortunato, Mauro, and Parisi, Giorgio. *Quantum Mechanics*. Cambridge University Press, Apr. 16, 2009 (cited on page 254).

[326] Srednicki, Mark. “Entropy and Area”. In: *Physical Review Letters* 71.5 (Aug. 2, 1993), pages 666–669. DOI: 10.1103/PhysRevLett.71.666. arXiv: hep-th/9303048. Web-link (cited on pages 254, 255).

[327] Bombelli, Luca et al. “Quantum Source of Entropy for Black Holes”. In: *Physical Review D* 34.2 (July 15, 1986), pages 373–383. DOI: 10.1103/PhysRevD.34.373. Web-link (cited on pages 254, 255).

[328] Rangamani, Mukund and Takayanagi, Tadashi. “Holographic Entanglement Entropy”. In: (Sept. 5, 2016). arXiv: 1609.01287. Web-link (cited on page 255).

[329] Nishioka, Tatsuma, Ryu, Shinsei, and Takayanagi, Tadashi. “Holographic Entanglement Entropy: An Overview”. In: *Journal of Physics A: Mathematical and Theoretical* 42.50 (2009), page 504008. Web-link (cited on pages 255, 256).

[330] Holzhey, C., Larsen, F., and Wilczek, F. “Geometric and Renormalized Entropy in Conformal Field Theory”. In: *Nuclear Physics B* 424.3 (Aug. 1994), pages 443–467. DOI: 10.1016/0550-3213(94)90402-2. arXiv: hep-th/9403108. Web-link (cited on page 255).

[331] Calabrese, Pasquale and Cardy, John. “Entanglement Entropy and Quantum Field Theory”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2004 (06 June 12, 2004), P06002. DOI: 10.1088/1742-5468/2004/06/P06002. arXiv: hep-th/0405152. Web-link (cited on page 255).

[332] Lewkowycz, Aitor and Maldacena, Juan. “Generalized Gravitational Entropy”. In: *Journal of High Energy Physics* 2013.8 (Aug. 2013). DOI: 10.1007/JHEP08(2013)090. arXiv: 1304.4926. Web-link (cited on page 256).

[333] Hubeny, Veronika E., Rangamani, Mukund, and Takayanagi, Tadashi. “A Covariant Holographic Entanglement Entropy Proposal”. In: *Journal of High Energy Physics* 2007 (07 July 23, 2007), pages 062–062. DOI: 10.1088/1126-6708/2007/07/062. arXiv: 0705.0016. Web-link (cited on page 256).

[334] Nishioka, Tastsuma. “Introduction to Entanglement Entropy”. 2014. Web-link (cited on page 256).

[335] Maldacena, Juan M. “Eternal Black Holes in AdS”. In: *Journal of High Energy Physics* 2003 (04 Apr. 12, 2003), pages 021–021. DOI: 10.1088/1126-6708/2003/04/021. arXiv: hep-th/0106112. Web-link (cited on pages 257, 261).

[336] Van Raamsdonk, Mark. “Building up Spacetime with Quantum Entanglement”. In: *International Journal of Modern Physics D* 19.14 (Dec. 2010), pages 2429–2435. DOI: 10.1142/S0218271810018529. arXiv: 1005.3035. Web-link (cited on pages 257, 259, 261).

[337] Van Raamsdonk, Mark. “Lectures on Gravity and Entanglement”. In: (Aug. 31, 2016). arXiv: 1609.00026. Web-link (cited on page 258).

[338] Blanco, David D. et al. “Relative Entropy and Holography”. In: *Journal of High Energy Physics* 2013.8 (Aug. 2013). DOI: 10.1007/JHEP08(2013)060. arXiv: 1305.3182. Web-link (cited on page 257).

[339] Lashkari, Nima, McDermott, Michael B., and Van Raamsdonk, Mark. “Gravitational Dynamics From Entanglement "Thermodynamics"”. In: *Journal of High Energy Physics* 2014.4 (Apr. 2014). DOI: 10.1007/JHEP04(2014)195. arXiv: 1308.3716. Web-link (cited on page 258).

[340] Swingle, Brian and Van Raamsdonk, Mark. “Universality of Gravity from Entanglement”. In: (May 12, 2014). arXiv: 1405.2933. Web-link (cited on page 258).

[341] Freedman, Michael and Headrick, Matthew. “Bit Threads and Holographic Entanglement”. In: (Apr. 1, 2016). arXiv: 1604.00354. Web-link (cited on page 258).

[342] Donnelly, William. “Entanglement Entropy in Loop Quantum Gravity”. In: *Physical Review D* 77.10 (May 6, 2008). DOI: 10.1103/PhysRevD.77.104006. arXiv: 0802.0880. Web-link (cited on page 258).

[343] Bodendorfer, Norbert. “A Note on Entanglement Entropy and Quantum Geometry”. In: *Classical and Quantum Gravity* 31.21 (Nov. 7, 2014), page 214004. DOI: 10.1088/0264-9381/31/21/214004. arXiv: 1402.1038. Web-link (cited on pages 258, 274).

[344] Krasnov, Kirill V. “Counting Surface States in the Loop Quantum Gravity”. In: *Physical Review D* 55.6 (Mar. 15, 1997), pages 3505–3513. DOI: 10.1103/PhysRevD.55.3505. arXiv: gr-qc/9603025. Web-link (cited on pages 258, 274).

[345] Bianchi, Eugenio and Myers, Robert C. “On the Architecture of Spacetime Geometry”. In: (Dec. 20, 2012). arXiv: 1212.5183. Web-link (cited on pages 258, 274, 281, 290).

[346] Hamma, Alioscia et al. “Area Law from Loop Quantum Gravity”. In: (June 4, 2015). arXiv: 1506.01623. Web-link (cited on pages 258, 259, 274, 281, 290).

[347] Jacobson, Ted. “Thermodynamics of Spacetime: The Einstein Equation of State”. In: *Physical Review Letters* 75.7 (Aug. 14, 1995), pages 1260–1263. DOI: 10.1103/PhysRevLett.75.1260. arXiv: gr-qc/9504004. Web-link (cited on page 259).

[348] Jacobson, Ted. “Gravitation and Vacuum Entanglement Entropy”. In: *International Journal of Modern Physics D* 21.11 (Oct. 2012), page 1242006. DOI: 10.1142/S0218271812420060. arXiv: 1204.6349. Web-link (cited on page 259).

[349] Jacobson, Ted. “Entanglement Equilibrium and the Einstein Equation”. In: *Physical Review Letters* 116.20 (May 20, 2016). DOI: 10.1103/PhysRevLett.116.201101. arXiv: 1505.04753. Web-link (cited on page 259).

[350] Swingle, Brian. “Constructing Holographic Spacetimes Using Entanglement Renormalization”. In: (Sept. 14, 2012). arXiv: 1209.3304. Web-link (cited on pages 259, 261).

[351] Cao, ChunJun, Carroll, Sean M., and Michalakis, Spyridon. “Space from Hilbert Space: Recovering Geometry from Bulk Entanglement”. In: (June 27, 2016). arXiv: 1606.08444. Web-link (cited on pages 259, 289).

[352] Noorbala, Mahdiyar. “SpaceTime from Hilbert Space: Decompositions of Hilbert Space as Instances of Time”. In: (Sept. 5, 2016). arXiv: 1609.01295. Web-link (cited on pages 259, 289).

[353] Raasakka, Matti. “Spacetime-Free Approach to Quantum Theory and Effective Spacetime Structure”. In: (May 12, 2016). arXiv: 1605.03942. Web-link (cited on pages 259, 289).

[354] Sabella-Garnier, Philippe-Alexandre. “Geometry from Quantum Mechanics : Entanglement, Energy Conditions and the Emergence of Space”. University of British Columbia, Aug. 24, 2016. Web-link (cited on pages 259, 289).

[355] Veneziano, Gabriele. “Construction of a Crossing-Simmetric, Regge-Behaved Amplitude for Linearly Rising Trajectories”. In: *Il Nuovo Cimento A* 57.1 (1968), pages 190–197. Web-link (cited on page 259).

[356] Susskind, Leonard. “Dual-Symmetric Theory of hadrons.—I”. In: *Il Nuovo Cimento A Series 10* 69.3 (1970), pages 457–496. Web-link (cited on page 259).

[357] Nambu, Y. “Quark Model and the Factorization of the Veneziano Amplitude”. In: Symmetries and Quark Models. 1970, page 269. Web-link (cited on page 259).

[358] Nielsen, H. B. “An Almost Physical Interpretation of the Integrand of the N-Point Veneziano Model”. In: *preprint at Niels Bohr Institute* (1970) (cited on page 259).

[359] Scherk, Joel and Schwarz, John H. “Dual Models for Non-Hadrons”. In: *Nuclear Physics B* 81.1 (1974), pages 118–144. Web-link (cited on page 260).

[360] Scherk, J. and Schwarz, John H. “Dual Model Approach to a Renormalizable Theory of Gravitation”. In: Superstrings: The First 15 Years of Superstring Theory (Reprints + Commentary - in 2 Volumes). Edited by SCHWARZ J H. Published by World Scientific Publishing Co. Pte. Ltd., 1985. ISBN #9789814542456, pp. 218-222. Sept. 1, 1985, pages 218–222. DOI: 10.1142/9789814542456_0022. Web-link (cited on page 260).

[361] Hořava, Petr. “Background Duality of Open-String Models”. In: *Physics Letters B* 231.3 (Nov. 9, 1989), pages 251–257. DOI: 10.1016/0370-2693(89)90209-8. Web-link (cited on page 260).

[362] Berkooz, Micha, Douglas, Michael R., and Leigh, Robert G. “Branes Intersecting at Angles”. In: *Nuclear Physics B* 480 (1-2 Nov. 1996), pages 265–278. DOI: 10.1016/S0550-3213(96)00452-X. arXiv: hep-th/9606139. Web-link (cited on page 260).

[363] Gukov, S., Vafa, C., and Witten, E. “CFT’s From Calabi-Yau Four-Folds”. In: *Nuclear Physics B* 584 (1-2 Sept. 2000), pages 69–108. DOI: 10.1016/S0550-3213(00)00373-4. arXiv: hep-th/9906070. Web-link (cited on page 261).

[364] Swingle, Brian. “Entanglement Renormalization and Holography”. In: *Physical Review D* 86.6 (Sept. 5, 2012). DOI: 10.1103/PhysRevD.86.065007. arXiv: 0905.1317. Web-link (cited on page 261).

[365] Almheiri, Ahmed et al. “Black Holes: Complementarity or Firewalls?” In: *Journal of High Energy Physics* 2013.2 (Feb. 2013). DOI: 10.1007/JHEP02(2013)062. arXiv: 1207.3123. Web-link (cited on page 261).

[366] Maldacena, Juan and Susskind, Leonard. “Cool Horizons for Entangled Black Holes”. In: *Fortschritte der Physik* 61.9 (Sept. 2, 2013), pages 781–811. DOI: 10.1002/prop.201300020. arXiv: 1306.0533. Web-link (cited on page 261).

References for Chapter 10: Evaluation and conclusion

[25] Thiemann, Thomas. *Modern Canonical Quantum General Relativity*. Cambridge University Press, 2007. Web-link (cited on pages 20, 134, 135, 149, 150, 152, 153, 155, 158–160, 162–167, 169–175, 181, 185, 194, 195, 271).

[121] Chamseddine, Ali H., Connes, Alain, and Mukhanov, Viatcheslav. “Geometry and the Quantum: Basics”. In: *Journal of High Energy Physics* 2014.12 (2014), pages 1–25. Web-link (cited on pages 79, 94, 268).

[122] Chamseddine, Ali H., Connes, Alain, and Mukhanov, Viatcheslav. “Quanta of Geometry: Noncommutative Aspects”. In: *Physical review letters* 114.9 (2015), page 091302. Web-link (cited on pages 79, 94, 268).

[123] Chamseddine, Ali H., Connes, Alain, and van Suijlekom, Walter D. “Grand Unification in the Spectral Pati-Salam Model”. In: (July 29, 2015). arXiv: 1507.08161. Web-link (cited on pages 79, 93, 94, 129, 268).

[144] Bodendorfer, Norbert. “A Note on Quantum Supergravity and AdS/CFT”. In: *arXiv preprint arXiv:1509.02036* (2015). Web-link (cited on pages 134, 194, 252, 274, 281, 290).

[145] Date, Ghanashyam and Hossain, Golam Mortuza. “Matter in Loop Quantum Gravity”. In: *Symmetry, Integrability and Geometry: Methods and Applications* 8 (2012), pages 10–26. Web-link (cited on pages 134, 273).

[146] Rovelli, Carlo and Vidotto, Francesca. *Covariant Loop Quantum Gravity: An Elementary Introduction to Quantum Gravity and Spinfoam Theory*. Cambridge University Press, Nov. 13, 2014 (cited on pages 134, 171, 273).

[149] Perez, Alejandro. “The Spin Foam Approach to Quantum Gravity”. In: *Living Rev. Rel* 16.3 (2013), pages 1205–2019. Web-link (cited on pages 145, 272).

[165] Gambini, Rodolfo and Pullin, Jorge. *A First Course in Loop Quantum Gravity*. Oxford University Press, 2011. Web-link (cited on pages 161, 185, 271).

[185] Bojowald, Martin and Mielczarek, Jakub. “Some Implications of Signature-Change in Cosmological Models of Loop Quantum Gravity”. In: *Journal of Cosmology and Astroparticle Physics* 2015 (08 2015), page 052. Web-link (cited on pages 192, 273).

[186] Barrau, Aurélien and Grain, Julien. “Cosmology without Time: What to Do with a Possible Signature Change from Quantum Gravitational Origin?” In: (July 26, 2016). arXiv: 1607.07589. Web-link (cited on pages 192, 273).

[237] Candelas, P. et al. “Vacuum Configurations for Superstrings”. In: *Nuclear Physics B* 258 (Jan. 1, 1985), pages 46–74. DOI: 10.1016/0550-3213(85)90602-9. Web-link (cited on pages 205, 206, 233, 236, 260, 279).

[242] Schottenloher, Martin. *A Mathematical Introduction to Conformal Field Theory*. Springer Science & Business Media, Sept. 26, 2008 (cited on pages 211, 225–227, 275).

[259] Roggenkamp, Daniel and Wendland, Katrin. “Limits and Degenerations of Unitary Conformal Field Theories”. In: *Communications in mathematical physics* 251.3 (2004), pages 589–643. Web-link (cited on pages 227, 268, 275, 281).

[260] Roggenkamp, Daniel and Wendland, Katrin. “Decoding the Geometry of Conformal Field Theories”. In: *arXiv preprint arXiv:0803.0657* (2008). Web-link (cited on pages 227, 268, 275, 281).

[343] Bodendorfer, Norbert. “A Note on Entanglement Entropy and Quantum Geometry”. In: *Classical and Quantum Gravity* 31.21 (Nov. 7, 2014), page 214004. DOI: 10.1088/0264-9381/31/21/214004. arXiv: 1402.1038. Web-link (cited on pages 258, 274).

[344] Krasnov, Kirill V. “Counting Surface States in the Loop Quantum Gravity”. In: *Physical Review D* 55.6 (Mar. 15, 1997), pages 3505–3513. DOI: 10.1103/PhysRevD.55.3505. arXiv: gr-qc/9603025. Web-link (cited on pages 258, 274).

[345] Bianchi, Eugenio and Myers, Robert C. “On the Architecture of Spacetime Geometry”. In: (Dec. 20, 2012). arXiv: 1212.5183. Web-link (cited on pages 258, 274, 281, 290).

[346] Hamma, Alioscia et al. “Area Law from Loop Quantum Gravity”. In: (June 4, 2015). arXiv: 1506.01623. Web-link (cited on pages 258, 259, 274, 281, 290).

[351] Cao, ChunJun, Carroll, Sean M., and Michalakis, Spyridon. “Space from Hilbert Space: Recovering Geometry from Bulk Entanglement”. In: (June 27, 2016). arXiv: 1606.08444. Web-link (cited on pages 259, 289).

[352] Noorbala, Mahdiyar. “SpaceTime from Hilbert Space: Decompositions of Hilbert Space as Instances of Time”. In: (Sept. 5, 2016). arXiv: 1609.01295. Web-link (cited on pages 259, 289).

[353] Raasakka, Matti. “Spacetime-Free Approach to Quantum Theory and Effective Spacetime Structure”. In: (May 12, 2016). arXiv: 1605.03942. Web-link (cited on pages 259, 289).

[354] Sabella-Garnier, Philippe-Alexandre. “Geometry from Quantum Mechanics : Entanglement, Energy Conditions and the Emergence of Space”. University of British Columbia, Aug. 24, 2016. Web-link (cited on pages 259, 289).

[367] Dungen, Koen van den. “Krein Spectral Triples and the Fermionic Action”. In: *Mathematical Physics, Analysis and Geometry* 19.1 (Mar. 2016). DOI: 10.1007/s11040-016-9207-z. arXiv: 1505.01939. Web-link (cited on page 267).

[368] Paschke, Mario. “An Essay on the Spectral Action and Its Relation to Quantum Gravity”. In: *Quantum Gravity*. Edited by Fauser, Bertfried, Tolksdorf, Jürgen, and Zeidler, Eberhard. Birkhäuser Basel, 2006, pages 127–150. DOI: 10.1007/978-3-7643-7978-0_7. Web-link (cited on page 267).

[369] Aastrup, Johannes and Grimstrup, Jesper M. “From Quantum Gravity to Quantum Field Theory via Noncommutative Geometry”. In: (May 1, 2011). arXiv: 1105.0194. Web-link (cited on page 268).

[370] Aastrup, Johannes and Grimstrup, Jesper M. “Intersecting Quantum Gravity with Noncommutative Geometry - a Review”. In: *Symmetry, Integrability and Geometry: Methods and Applications* (Mar. 28, 2012). DOI: 10.3842/SIGMA.2012.018. arXiv: 1203.6164. Web-link (cited on pages 268, 281).

[371] Aastrup, Johannes and Grimstrup, Jesper M. “Quantum Holonomy Theory”. In: (Apr. 27, 2015). arXiv: 1504.07100. Web-link (cited on pages 268, 281).

[372] Oriti, Daniele. “Group Field Theory and Loop Quantum Gravity”. In: (Aug. 29, 2014). arXiv: 1408.7112. Web-link (cited on page 272).

[373] Gambini, Rodolfo, Olmedo, Javier, and Pullin, Jorge. “Quantum Black Holes in Loop Quantum Gravity”. In: *Classical and Quantum Gravity* 31.9 (May 7, 2014), page 095009. DOI: 10.1088/0264-9381/31/9/095009. arXiv: 1310.5996. Web-link (cited on page 273).

[374] Corichi, Alejandro and Montoya, Edison. “Loop Quantum Cosmology of Bianchi IX: Effective Dynamics”. In: (Feb. 8, 2015). arXiv: 1502.02342. Web-link (cited on page 273).

[375] Corichi, Alejandro, Olmedo, Javier, and Rastgoo, Saeed. “Vacuum CGHS in Loop Quantum Gravity and Singularity Resolution”. In: (Aug. 22, 2016). arXiv: 1608.06246. Web-link (cited on page 273).

[376] Gorji, M. A., Nozari, K., and Vakili, B. “Gravity’s Rainbow: A Bridge between LQC and DSR”. In: (June 2, 2016). arXiv: 1606.00910. Web-link (cited on page 273).

[377] Kowalski-Glikman, Jerzy. “Introduction to Doubly Special Relativity”. In: 669 (2005), pages 131–159. DOI: 10.1007/11377306_5. arXiv: hep-th/0405273. Web-link (cited on page 273).

[378] Gambini, Rodolfo and Pullin, Jorge. “Emergence of String-like Physics from Lorentz Invariance in Loop Quantum Gravity”. In: *International Journal of Modern Physics D* 23.12 (Oct. 2014), page 1442023. DOI: 10.1142/S0218271814420231. arXiv: 1406.2610. Web-link (cited on pages 273, 274, 281).

[379] Collins, John et al. “Lorentz Invariance and Quantum Gravity: An Additional Fine-Tuning Problem?” In: *Physical Review Letters* 93.19 (Nov. 5, 2004), page 191301. DOI: 10.1103/PhysRevLett.93.191301. Web-link (cited on page 273).

[380] Kong, Liang. “Conformal Field Theory and a New Geometry”. In: (July 19, 2011). arXiv: 1107.3649. Web-link (cited on page 275).

[381] Gildener, Eldad and Weinberg, Steven. “Symmetry Breaking and Scalar Bosons”. In: *Physical Review D* 13.12 (June 15, 1976), pages 3333–3341. DOI: 10.1103/PhysRevD.13.3333. Web-link (cited on page 279).

[382] Susskind, Leonard. “Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory”. In: *Physical Review D* 20.10 (Nov. 15, 1979), pages 2619–2625. DOI: 10.1103/PhysRevD.20.2619. Web-link (cited on page 279).

[383] Witten, Edward. “Dynamical Breaking of Supersymmetry”. In: *Nuclear Physics B* 188.3 (Oct. 5, 1981), pages 513–554. DOI: 10.1016/0550-3213(81)90006-7. Web-link (cited on page 279).

[384] Langacker, Paul. *The Standard Model and Beyond*. CRC Press, Dec. 1, 2009 (cited on page 279).

[385] Polchinski, Joseph. “Quantum Gravity at the Planck Length”. In: *International Journal of Modern Physics A* 14.17 (July 10, 1999), pages 2633–2658. DOI: 10.1142/S0217751X99001329. arXiv: hep-th/9812104. Web-link (cited on page 279).

[386] Han, Muxin and Hung, Ling-Yan. “Loop Quantum Gravity, Exact Holographic Mapping, and Holographic Entanglement Entropy”. In: (Oct. 6, 2016). arXiv: 1610.02134. Web-link (cited on page 281).

[387] Smolin, Lee. “Holographic Relations in Loop Quantum Gravity”. In: (Aug. 9, 2016). arXiv: 1608.02932. Web-link (cited on page 281).

[388] Chiribella, Giulio, D’Ariano, Giacomo Mauro, and Perinotti, Paolo. “Quantum from Principles”. In: 181 (2016). DOI: 10.1007/978-94-017-7303-4. arXiv: 1506.00398. Web-link (cited on page 288).

[389] Chiribella, G., D’Ariano, G. M., and Perinotti, P. “Informational Derivation of Quantum Theory”. In: *Physical Review A* 84.1 (July 11, 2011). DOI: 10.1103/PhysRevA.84.012311. arXiv: 1011.6451. Web-link (cited on page 288).

[390] D’Ariano, Giacomo Mauro. “Probabilistic Theories: What Is Special about Quantum Mechanics?” In: (July 28, 2008). arXiv: 0807.4383. Web-link (cited on page 288).

[391] Hardy, Lucien. “Reformulating and Reconstructing Quantum Theory”. In: (Apr. 11, 2011). arXiv: 1104.2066. Web-link (cited on page 288).

[392] Hardy, Lucien. “Reconstructing Quantum Theory”. In: (Mar. 6, 2013). arXiv: 1303.1538. Web-link (cited on page 288).

[393] Hardy, Lucien. “Quantum Theory From Five Reasonable Axioms”. In: (Jan. 3, 2001). arXiv: quant-ph/0101012. Web-link (cited on page 288).

[394] Aaronson, Scott. *Quantum Computing Since Democritus*. Cambridge University Press, Mar. 14, 2013 (cited on pages 288, 371).

[395] Hoehn, Philipp A. “Toolbox for Reconstructing Quantum Theory from Rules on Information Acquisition”. In: (Dec. 29, 2014). arXiv: 1412.8323. Web-link (cited on page 288).

[396] Hoehn, Philipp A. and Wever, Christopher. “Quantum Theory from Questions”. In: (Nov. 3, 2015). arXiv: 1511.01130. Web-link (cited on page 288).

[397] Konopka, Tomasz, Markopoulou, Fotini, and Smolin, Lee. “Quantum Graphity”. In: (Nov. 17, 2006). arXiv: hep-th/0611197. Web-link (cited on page 289).

[398] Levin, Michael and Wen, Xiao-Gang. "Quantum Ether: Photons and Electrons from a Rotor Model". In: *Physical Review B* 73.3 (Jan. 23, 2006). DOI: 10.1103/PhysRevB.73.035122. arXiv: hep-th/0507118. Web-link (cited on page 289).

[399] Zeng, Bei et al. "Quantum Information Meets Quantum Matter – From Quantum Entanglement to Topological Phase in Many-Body Systems". In: (Aug. 11, 2015). arXiv: 1508.02595. Web-link (cited on page 289).

[400] Bilson-Thompson, Sundance et al. "Emergent Braided Matter of Quantum Geometry". In: *Symmetry, Integrability and Geometry: Methods and Applications* (Mar. 24, 2012). DOI: 10.3842/SIGMA.2012.014. arXiv: 1109.0080. Web-link (cited on page 289).

[401] Okon, Elias and Sudarsky, Daniel. "Black Holes, Information Loss and the Measurement Problem". In: (July 5, 2016). arXiv: 1607.01255. Web-link (cited on page 291).

[402] Bonder, Yuri, Okon, Elias, and Sudarsky, Daniel. "Can Gravity Account for the Emergence of Classicality?" In: *Physical Review D* 92.12 (Dec. 22, 2015). DOI: 10.1103/PhysRevD.92.124050. arXiv: 1509.04363. Web-link (cited on page 291).

[403] Pikovski, Igor et al. "Universal Decoherence due to Gravitational Time Dilation". In: *Nature Physics* advance online publication (June 15, 2015). DOI: 10.1038/nphys3366. Web-link (cited on page 291).

References for Chapter A: Logic and sets

[10] Takeuti, Gaisi and Zaring, Wilson M. *Introduction to Axiomatic Set Theory*. Springer-Verlag, 1982 (cited on pages 7, 300).

[404] *Home Page - Metamath*. Web-link (cited on page 298).

[405] *Welcome! | The Coq Proof Assistant*. Web-link (cited on page 298).

[406] Mendelson, Elliott. *Introduction to Mathematical Logic*. CRC press, 2009. Web-link (cited on page 300).

[407] Suppes, Patrick. *Axiomatic Set Theory*. Courier Corporation, 1960. Web-link (cited on page 300).

[408] Ciesielski, Krzysztof. *Set Theory for the Working Mathematician*. Volume 39. Cambridge University Press, 1997. Web-link (cited on page 300).

[409] Rubin, Jean E. *Set Theory for the Mathematician*. Holden-Day, 1967 (cited on page 300).

[410] Avigad, Jeremy and Harrison, John. "Formally Verified Mathematics". In: *Communications of the ACM* 57.4 (Apr. 1, 2014), pages 66–75. DOI: 10.1145/2591012. Web-link (cited on page 300).

[411] *Mizar System*. In: *Wikipedia, the Free Encyclopedia*. Sept. 13, 2015. Web-link (cited on page 300).

[412] Kanamori, Akihiro. "The Mathematical Development of Set Theory from Cantor to Cohen". In: *Bulletin of Symbolic Logic* 2.1 (Mar. 1996), pages 1–71. DOI: 10.2307/421046. Web-link (cited on page 300).

References for Chapter B: Category theory

[413] Simmons, Harold. *An Introduction to Category Theory*. Cambridge University Press, 2011. Web-link (cited on page 305).

[414] Lawvere, F. William and Rosebrugh, Robert. *Sets for Mathematics*. Cambridge University Press, 2003. Web-link (cited on page 305).

References for Chapter C: Algebra

[415] Adkins, William and Weintraub, Steven. *Algebra: An Approach via Module Theory*. Volume 136. Springer Science & Business Media, 2012. Web-link (cited on pages 309, 315).

[416] Dummit, David Steven and Foote, Richard M. *Abstract Algebra*. Volume 1999. Prentice Hall Englewood Cliffs, 1991. Web-link (cited on page 315).

[417] Fraleigh, John B. *First Course in Abstract Algebra, A: Pearson New International Edition*. Pearson Higher Ed, 2013 (cited on page 315).

[418] Givant, Steven and Halmos, Paul. *Introduction to Boolean Algebras*. Springer Science & Business Media, Dec. 2, 2008 (cited on page 315).

References for Chapter D: Measure theory and probability

[21] Folland, Gerald B. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, 2013. Web-link (cited on pages 16, 20, 318, 319).

[22] Tao, Terrence. *Algebraic Probability Spaces*. 2014-06-29T06:22:29+00:00. Web-link (cited on pages 17, 319).

References for Chapter E: Topology

[103] Tabak, John. *Beyond Geometry: A New Mathematics of Space and Form*. Facts on File, 2011 (cited on pages 61, 321).

[104] Morris, Sidney A. *Topology without Tears*. University of New England, 1989. Web-link (cited on pages 61, 170, 327).

[419] Munkres, James R. *Topology*. Prentice Hall, Incorporated, 2000 (cited on page 327).

[420] Kelley, John L. *General Topology*. Springer Science & Business Media, 1975. Web-link (cited on page 327).

References for Chapter F: Differential geometry

[33] Rudolph, Gerd and Schmidt, Matthias. *Differential Geometry and Mathematical Physics: Part I. Manifolds, Lie Groups and Hamiltonian Systems*. Springer Science & Business Media, Nov. 9, 2012 (cited on pages 24, 331, 334).

[45] Wald, Robert M. *General Relativity*. University of Chicago Press, May 15, 2010 (cited on pages 30, 33, 36–38, 44, 53, 68, 134, 244, 331).

[90] Martin, Daniel. *Manifold Theory: An Introduction for Mathematical Physicists*. Elsevier, 2002. Web-link (cited on pages 52, 68, 340).

[118] Nakahara, Mikio. *Geometry, Topology and Physics*. CRC Press, 2003. Web-link (cited on pages 78, 340).

[421] Spivak, Michael. *A Comprehensive Introduction to Differential Geometry Vol 1 (Houston, TX: Publish or Perish)*. Inc, 1979 (cited on page 340).

[422] Lee, John M. *Introduction to Smooth Manifolds*. Springer Science & Business Media, 2003 (cited on page 340).

References for Chapter G: Lie groups and Lie algebras

[246] Cornwell, J. F. *Group Theory in Physics - Volume 3*. Academic Press, 1989 (cited on pages 218, 261, 344).

[423] De Kerf and Bäuerle. *Lie Algebras, Part 2: Finite and Infinite Dimensional Lie Algebras and Applications in Physics*. North Holland, Apr. 14, 1999 (cited on page 344).

[424] De Kerf and Bäuerle. *Lie Algebras, Part 1: Finite and Infinite Dimensional Lie Algebras and Applications in Physics*. 1st edition. Amsterdam, The Netherlands : New York, N.Y: North Holland, Dec. 28, 1990 (cited on page 344).

[425] Fuchs, Jürgen and Schweigert, Christoph. *Symmetries, Lie Algebras and Representations: A Graduate Course for Physicists*. Cambridge University Press, 2003. Web-link (cited on page 344).

References for Chapter H: Functional analysis

[13] Strocchi, F. *An Introduction to the Mathematical Structure of Quantum Mechanics: A Short Course for Mathematicians*. World Scientific, 2008 (cited on pages 13, 14, 20–23, 26, 27, 149, 353).

[27] Greiner, Walter and Reinhardt, Joachim. *Field Quantization*. Springer Science & Business Media, June 29, 2013 (cited on pages 20, 135, 358).

[101] Suijlekom, Walter van. *Noncommutative Geometry and Particle Physics*. Springer, July 21, 2014 (cited on pages 61, 65, 73, 74, 78, 81, 84, 89–93, 95, 358).

[105] Ha, Dzung Minh. *Functional Analysis*. Matrix Editions, Jan. 1, 2006 (cited on pages 62, 346, 358).

[107] Blackadar, Bruce. *Operator Algebras: Theory of C*-Algebras and Von Neumann Algebras*. Taylor & Francis, 2006 (cited on pages 65, 66, 358).

[109] Sitarz, A. "3 1/2 Lectures on Noncommutative Geometry". In: *Acta Polytechnica* vol. 48 (2008), pages 34–55 (cited on pages 68–71, 358).

[110] Sontz, Stephen Bruce. *Principal Bundles*. Universitext. Cham: Springer International Publishing, 2015. Web-link (cited on pages 69, 70, 358).

[111] Baehr, H. C., Dimakis, A., and Müller-Hoissen, F. "Differential Calculi on Commutative Algebras". In: *Journal of Physics A: Mathematical and General* 28.11 (June 7, 1995), pages 3197–3222. DOI: 10.1088/0305-4470/28/11/019. arXiv: hep-th/9412069. Web-link (cited on pages 69, 70, 358).

[426] Reed, Michael and Simon, Barry. *Methods of Modern Mathematical Physics: Functional Analysis*. Volume 1. Gulf Professional Publishing, 1980. Web-link (cited on page 358).

[427] Rudin, Walter. *Functional Analysis*. McGraw-Hill, Jan. 1991 (cited on page 358).

[428] Dimakis, A. and M"uller-Hoissen, F. "Noncommutative Differential Calculus: Quantum Groups, Stochastic Processes, and the Antibracket". In: (Jan. 28, 1994). arXiv: hep-th/9401151. Web-link (cited on page 358).

[429] Dubois-Violette, Michel. "Lectures on Graded Differential Algebras and Noncommutative Geometry". In: *Noncommutative Differential Geometry and Its Applications to Physics*. Springer, 2001, pages 245–306. Web-link (cited on page 358).

[430] Carroll, R. W. *Calculus Revisited*. Springer Science & Business Media, Mar. 9, 2013 (cited on page 358).

References for Chapter I: Algebraic topology

[100] Khalkhali, Masoud. *Basic Noncommutative Geometry*. European Mathematical Society, 2009 (cited on pages 61, 63, 65, 68, 365).

[431] Hatcher, Allen. *Algebraic Topology*. 2002. Web-link (cited on page 366).

References for Chapter J: Other math

[44] Weinberg, Steven. *The Quantum Theory of Fields*. Volume 1. Cambridge University Press, Aug. 13, 1996 (cited on pages 30, 31, 38, 40–42, 44, 204, 242, 371).

[50] Mirman, R. *Massless Representations of the Poincaré Group: Electromagnetism, Gravitation, Quantum Mechanics, Geometry*. iUniverse, May 2005 (cited on pages 31, 371).

[51] Schwartz, Matthew D. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2013. Web-link (cited on pages 31, 38, 41, 43, 371).

[64] Weinberg, Steven. "What Is Quantum Field Theory, and What Did We Think It Is?" In: *arXiv preprint hep-th/9702027* (1997). Web-link (cited on pages 38, 40, 42, 44, 371).

[76] Misner, Charles W., Thorne, Kip S., and Wheeler, John Archibald. *Gravitation*. W. H. Freeman, Sept. 15, 1973 (cited on pages 46, 53, 371).

[82] Casola, Eolo Di, Liberati, Stefano, and Sonego, Sebastiano. "Nonequivalence of Equivalence Principles". In: *American Journal of Physics* 83.1 (Jan. 1, 2015), pages 39–46. DOI: 10.1119/1.4895342. Web-link (cited on pages 49, 372).

[84] Aldrovandi, Ruben and Pereira, Jose G. *Teleparallel Gravity: An Introduction*. Volume 173. Springer Science & Business Media, 2012. Web-link (cited on pages 50, 372).

[250] Tong, David. "Lectures on String Theory". In: (Aug. 3, 2009). arXiv: 0908.0333. Web-link (cited on pages 222, 223, 225, 372).

[291] Polchinski, Joseph. "Why Trust a Theory? Some Further Remarks (Part 1)". In: (Jan. 22, 2016). arXiv: 1601.06145. Web-link (cited on pages 241, 371).

[394] Aaronson, Scott. *Quantum Computing Since Democritus*. Cambridge University Press, Mar. 14, 2013 (cited on pages 288, 371).

[432] Levitan, B. M. and Zhikov, V. V. *Almost Periodic Functions and Differential Equations*. CUP Archive, Dec. 2, 1982 (cited on page 369).

[433] Jones, William Thomas. *A History of Western Philosophy*. Harcourt Brace Jovanovich, 1975 (cited on page 371).

[434] Hossenfelder, Sabine. *Backreaction: The LHC "nightmare Scenario" Has Come True.* Web-link (cited on page 371).

[435] Wigner, Eugene P. "The Unreasonable Effectiveness of Mathematics in the Natural Sciences. Richard Courant Lecture in Mathematical Sciences Delivered at New York University, May 11, 1959". In: *Communications on pure and applied mathematics* 13.1 (1960), pages 1–14. Web-link (cited on page 371).

[436] Baez, John. "Higher-Dimensional Algebra and Planck-Scale Physics". In: *Physics Meets Philosophy at the Planck Length, Editors C. Callender and N. Huggett, Cambridge University Press, Cambridge* (2001), pages 177–195. Web-link (cited on pages 371, 372).

[437] Frohlich, J. and Schubnel, B. "Quantum Probability Theory and the Foundations of Quantum Mechanics". In: (Oct. 5, 2013). arXiv: 1310.1484. Web-link (cited on page 371).

[438] Bekaert, Xavier, Boulanger, Nicolas, and Sundell, Per. "How Higher-Spin Gravity Surpasses the Spin Two Barrier: No-Go Theorems versus Yes-Go Examples". In: *Reviews of Modern Physics* 84.3 (July 3, 2012), pages 987–1009. DOI: 10.1103/RevModPhys.84.987. arXiv: 1007.0435. Web-link (cited on page 371).

[439] Sotiriou, Thomas P., Liberati, Stefano, and Faraoni, Valerio. "Theory of Gravitation Theories: A No-Progress Report". In: *International Journal of Modern Physics D* 17 (03n04 2008), pages 399–423. Web-link (cited on page 372).

[440] Arkani-Hamed, Nima. *Robustness of GR*. 2011. Web-link (cited on page 372).

[441] Kastler, Daniel. "Noncommutative Geometry and Basic Physics". In: *Geometry and Quantum Physics*. Edited by Gausterer, H., Pittner, L., and Grosse, Harald. Lecture Notes in Physics 543. Springer Berlin Heidelberg, 2000, pages 131–230. DOI: 10.1007/3-540-46552-9_4. Web-link (cited on page 372).

[442] Oriti, Daniele. *Approaches to Quantum Gravity: Toward a New Understanding of Space, Time and Matter*. Cambridge University Press, Mar. 5, 2009 (cited on page 372).

[443] Baulieu, L. et al. *Strings, Branes and Dualities*. Springer Science & Business Media, Dec. 6, 2012 (cited on page 372).

[444] Linde, A. D. "Eternally Existing Self-Reproducing Chaotic Inflanationary Universe". In: *Physics Letters B* 175.4 (Aug. 14, 1986), pages 395–400. DOI: 10.1016/0370-2693(86)90611-8. Web-link (cited on page 372).

[445] Rovelli, Carlo. "Unfinished Revolution". In: (Apr. 10, 2006). arXiv: gr-qc/0604045. Web-link (cited on page 372).

Index

A-bimodule, 354
C^k differentiable manifold, 330
C^k-function, 330
ϵ-almost period, 367
$\mathbb{Z}_2$-graded module, 314
σ-algebra, 317
$*$-algebra, 350
$*$-ring, 350
't Hooft coupling, 246
't Hooft double line notation, 246, 247
(A,B)-bimodule, 354
*-automorphism, 351
*-morphism, 350

Adjoint action, 86
Adjoint of an operator, 348
Affine connection, 334
Algebra of functions, 62
Algebra of sets, 317
Algebraic classical system, 20
Algebraic dynamical system, 25
Algebraic probability space, 17, 19
Algebraic quantum system, 22
Algebraic structures, 307
Algebraic vector bundle, 68
Almost commutative space, 75
Almost periodic function, 368
Anti-automorphism, 350
Area operator, 181
Associated JLB algebra, 25
Axiom of choice, 295
Axiomatic set theory, 295

Background free, 52
Background independent, 52
Balanced tensor product, 355
Banach *-algebra, 350
Banach algebra, 347
Banach space, 347
Basis for a topology, 323
Biholomorphic, 339
Bilinear form, 312
Bohr compactification, 368
Borel measure, 319
Borel sets, 319
Bosonic action principle, 88
Boundary, 324
Bounded function, 318
Bounded operators, 348
BPS bound, 248
BPS states, 248
Bundle, 336
Bundle morphism, 337

C*-algebra, 351
Calabi-Yau manifold, 234
Canonical basis, 334
Canonical triple, 73
Casimir element, 342
Category, 302

Cauchy sequence, 346
Cauchy surface, 340
Causal curve, 340
CDP, 40
Centralizer, 352
Chain complex, 359
Chain group, 360
Character of an algebra, 63
Chart, 330
Chebyshev norm, 354
Classical algebraic dynamical system, 25
Classification of Clifford algebras, 314
Clifford algebra, 313
Closed forms, 364
Closed manifold, 331
Closed set, 322, 326
Closure, 324
Cluster decomposition principle, 40
Cochain complex, 360
Cochain group, 360
codimension, 340
Cohomology classes, 361
Cohomology group, 361
Commutant, 352
Commutative Gelfand-Naimark theorem, 65
Commutator, 332
Compact, 325
Compact support, 367
Compatible charts, 330
Complete metric space, 347
Completion, 347
Complex charts, 339
Complex manifold, 339
Complex structure, 369
Complexification, 369
Configuration space, 139
Conjugate momentum, 140
Connected topological space, 325
Connection, 338
Connection on module, 71
Connes differential one-forms, 71
Connes distance function, 73
Constraint surface, 142
Continuous function, 322
Continuum hypothesis, 295
Contravariant functor, 304
Convergent sequence, 346
Coset, 309
Cotangent space, 331
Cotangent space basis, 331
Covariant functor, 304
Covariant theory, 51
Covector, 332
Cover, 325
Curvature, 339
Curve, 164
Cyclic cochain, 365
Cyclic cohains, 365
Cyclic cohomology, 365
Cyclic vector, 353
Cyl, 172
Cylindrical function, 172
Cylindrical functions, 171, 176
Cylindrical inner product, 175
Cylindrical measure, 174

D-brane, 248
Darboux's theorem, 335
de Rham cohomology, 363
Dense set, 326
Density operator, 253
Diffeomorphism constraint, 182
Diffeomorphism invariant theory, 52
Differential algebra, 356
Differential calculus, 357
Differential graded algebra, 315, 356
Dirac observables, 145
Dirac quantization, 150
Directed graph, 312
Directed poset, 297
Directed preorder, 297
Dirichlet boundary condition, 208
Discrete topology, 322
Distance function, 72
Distribution, 358
Distributive, 308
Dolbeault cohomology, 364
Domain and codomain, 303
Dual of triad, 167
Dynamical field, 51

Edge, 165
Elementary functions, 146
Embedding, 340
Energy-momentum tensor, 209
Entanglement entropy, 254
Enveloping algebra, 342
Equivalent categories, 304
Equivalent curves, 165

Essentially bounded function, 318
Evaluation map, 64
Even spectral triple, 74
Exact forms, 364
Exact sequence, 361
Extended configuration space, 169
Exterior, 324
Exterior derivative, 333
Extrinsic curvature, 158

Face, 362
Faithful state, 353
Fermionic action principle, 88
Fiber, 336
Fiber bundle, 336
Finite spectral triple, 75
Finite support, 367
Finitely generated module, 310
First class constraints, 144
First class function, 144
First order differential calculus, 357
First order logic, 294
First-order differential calculus, 69
Flow, 332
Fluctuated Dirac operator, 83
Flux, 167
Foliation, 340
Foliation invariance, 155
Formal systems, 293
Fredholm module, 368
Fredholm operator, 368
Free action, 343
Free vector space, 311

G-module, 343
Gauge fields, 83
Gauss constraint, 178
Gelfand spectrum, 63
Gelfand topology, 63
Gelfand transform, 64
Gelfand triple, 354
Gelfand-Naimark theorem, 66, 353
Generalized connections, 172
Generalized K-derivation, 356
Generating set, 309
Generators, 342
Geroch's theorem, 153
GL(n), 344
Global section, 337
Globally hyperbolic, 340
GNS theorem, 353
Graded algebra, 314
Graded module, 314
Graded ring, 314
Graph, 165
Group action, 342
Group of boundaries, 360
Group of coboundaries, 360
Group of cocycles, 360
Group of cycles, 360
Group of unitary elements, 85
Groupoid, 308
Groupoid of a graph, 169

Hamilton's equations, 141
Hamiltonian, 140
Hamiltonian constraint, 182
Hamiltonian vector field, 335
Hausdorff topological space, 324
Hilbert (A,B)-bimodule, 355
Hilbert bimodule, 355
Hilbert space, 347
Hochschild cohomology, 364
Holomorphic atlas, 339
Holomorphic function, 339
Holonomy, 165, 166
Hom space of a subgrupoid, 170
Homeomorphism, 323
Homology classes, 361
Homology group, 361
Hypersurface, 340

Ideal, 309
Idempotent, 308
Identity functor, 304
Immersion, 340
Index of a Fredholm operator, 368
Infinity norm, 354
Initial topology, 327
Inner automorphism, 85
Inner derivation, 356
Inner fluctuations, 83
Inner product, 346
Inner product space, 346
Inner semidirect product, 311
Integral curve, 332
Interior, 324
Interior point, 324
Introduction, 1
Invariant theory, 52

Involution, 350
Irreducible finite spectral triple, 75
Isomorphic categories, 305

JLB algebra, 25

K-derivation, 355
Killing vector field, 333

Leaf, 340
Left-invariant measure, 343
Lie algebra, 341
Lie algebra of unitaries, 85
Lie algebra representation, 342
Lie derivative, 333
Lie group, 341
Limit point, 326
Linear functional, 352
Linearly independent set, 310
Local section, 337
Locally finite, 325
Locally small category, 302
Logic, 294

Manifold, 329
Maximal C^k-atlas, 330
Measurable function, 318
Measurable space, 317
Measure, 318
Measure space, 318
Measure space morphism, 318
Metric, 345
Metric space, 345
Minimal generating set, 310
Morita equivalence, 81, 355
Musical isomorphism (symplectic), 334

N-simplex, 362
Nambu-Goto action, 207
Neighborhood, 322
Non-dynamical field, 51
Non-planar diagrams, 247
Noncommutative topology, 61
Norm, 345
Normal state, 353
Normal subgroup, 309
Normed *-algebra, 350
Normed algebra, 347
Normed vector space, 346

Open cover, 325
Open neighborhood, 322
Open set, 322
Operations on curves, 164
Operator norm, 348
Orbifold, 234
Orbit of an element, 343
Ordered vector space, 352
Orientable manifold, 331, 333
Oriented atlas, 333
Oriented manifold, 333
Orientifold, 234
Orthogonal projection operator, 349
Orthogonal projection tensor, 156
Outer automorphisms, 85
Outer derivation, 356

p-brane, 248
Paracompact, 325
Parity, 314
Path connected topological space, 326
Peano axioms, 298
Peter-Weyl theorem, 344
Planar diagrams, 247
Poincare duality theorem, 364
Poisson algebra, 335
Poisson manifold, 336
Poisson structure, 335
Polyakov action, 207
Positive linear functional, 352
Positive operator, 348
Primary constraints, 142
Principal bundle, 337
Probability space, 319
Product Hom space, 170
Product triple, 75
Projection, 358
Projection operator, 349
Projective family, 170, 297
Projective limit, 298
Projective representation, 36
Propositional logic system, 294
Pseudo real structure, 369
Pseudo-Riemannian manifold, 338
Pseudo-Riemannian metric, 338

Quadratic form, 312
Quantum algebraic dynamical system, 26
Quotient group, 309
Quotient ring, 309

R-Algebra, 309

R-linear map, 309
Radon measure, 319
Random variable, 319
Real spectral triple, 74
Real structure, 368
Reduced density operator, 254
Refined algebraic quantization, 150
Refinement, 325
Relative topology, 327
Relatively dense, 367
Resolvent, 349
Resolvent set, 349
Riemann surface, 339
Riemannian manifold, 338
Riemannian metric, 338

Second class functions, 144
Secondary constraints, 142
Self-adjoint elements, 351
Self-adjoint operator, 349
Seminorm, 346
Separating set, 326
Separating vector, 75
Sequence, 346
Serre-Swan theorem, 68
Short exact sequence, 361
Simplicial complex, 362
Simply connected topological space, 326
Slice, 340
Small category, 302
Space of global sections, 337
Spatial curvature tensor, 157
Spatial Levi-Civita connection, 157
Spatial projection tensor, 156
Spatial tensor, 157
Spectral radius, 349
Spectral triple, 73
Spectrum of a C*-algebra, 349
Spin manifold, 72
Spin structure, 72
Spinor inner product, 75, 105
Spinors, 314
Split exact sequence, 361
Stabilizer of an element, 343
Standard Clifford algebra, 314
Standard quadratic form, 313
State, 352
SU(n), 344
su(n), 344
Subcover, 325
Support of a function, 367
Supremum norm, 354
Symmetric bilinear form, 312
Symplectic manifold, 334
Symplectic vector space, 334

Tame groupoid, 169
Tangent bundle, 332
Tangent space, 331
Tangent space basis, 331
Test functions, 357
Test physics, 49
The current epistemological challenges, 7
The general epistemology of physics, 2
The K-module of derivations, 356
Topological space, 321
Total Hamiltonian, 143
Totally constrained system, 144
Transitive action, 343
Trivial topology, 322

Undirected graph, 312
Uniform norm, 354
Unimodular gauge group, 88
Unimodular gauge Lie algebra, 88
Unit elements, 351
Unit normal of foliation, 155
Unitary elements, 351
Unitary operator, 349
Universal FODC, 69

Vector, 331
Vector bundle, 337
Vector field, 332
Vector space, 310
Virasoro algebra, 211
Volume form, 333
Volume operator, 181
von Neumann algebra, 352
von Neumann boundary condition, 208
von Neumann entropy, 254

Weak topology, 327
Weaker and stronger topologies, 324
Weakly equivalent, 143
Weakly vanishing, 143
Wigner's theorem, 35
Witt algebra, 211