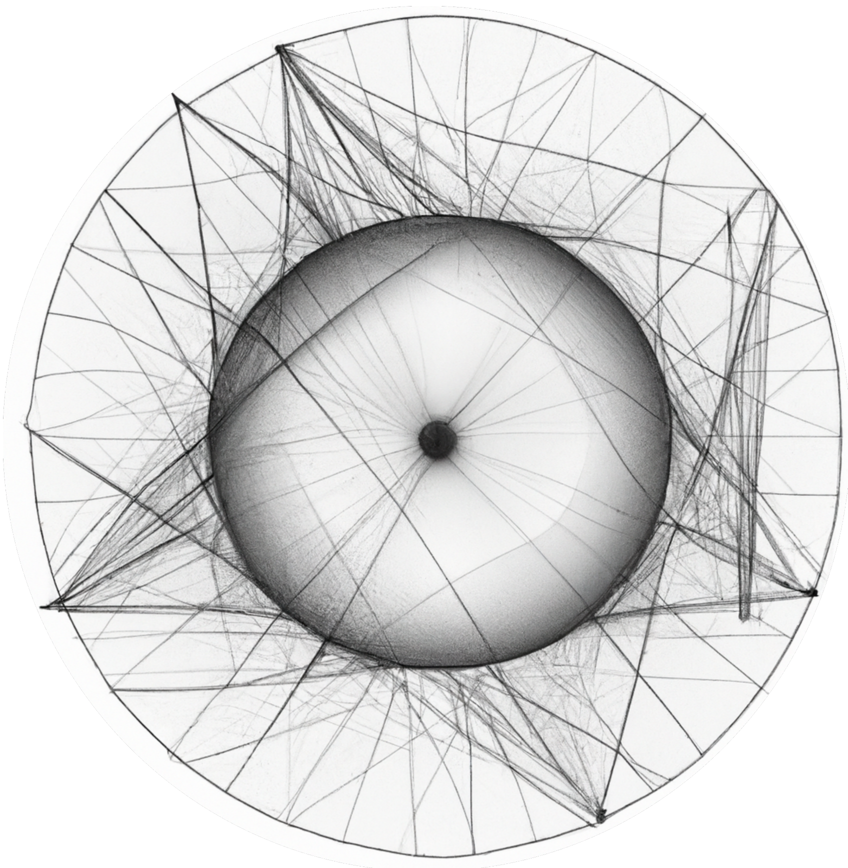
Quantum Field Theory in Curved Space-Time

Pieter Bomans

Mathematical Institute
University of Oxford



These notes are meant to complement the lectures and may be updated over the course of the term. Feedback is very welcome, especially if there are typos or places where the text lacks clarity. Please send any comments, corrections or questions to

pieter.bomans@maths.ox.ac.uk.

About this course

Welcome to the 2023-2024 edition of **Quantum Field Theory in Curved Space-Time** of the Oxford Mathematical and Theoretical Physics master course. This course builds on the courses Quantum Field Theory, Advanced Quantum Field Theory and General Relativity I and II. In these courses, the fundamental concepts of quantum field theory and general relativity were introduced. Here we will further develop these subjects more broadly by applying the tools of quantum field theory in some of the curved space-times encountered in general relativity. Along the way several concepts from differential geometry, representation theory as well as special functions will arise. Familiarity with these topics is not required but will be helpful.

Synopsis

The course will consist of the following topics:

- Quantum field theory in flat space
- Lorentzian geometry and causality
- Quantum field theory in curved space-time
- The Unruh effect
- Hawking Radiation
- Quantum fields in de Sitter space
- Quantum cosmology
- Quantum fields in Anti-de Sitter space
- Holography

References

There are a variety of excellent textbooks and lecture notes available in the literature. This course borrows from a number of them, the most relevant ones are listed below.

Standard textbooks on quantum field theory in curved spaces are:

- N.D. Birrell and P. Davies, *Quantum Fields in Curved Space*, (Cambridge University Press 1982)
- R. Wald, *QFT in Curved Space-time and Black Hole Thermodynamics*, (Chicago University Press 1994)
- V. Mukhanov and S. Winitzki, *Introduction to Quantum Effects in Gravity*, (Cambridge University Press 2007)
- L. Parker and D. Toms, *Quantum Field Theory in Curved Spacetime*, (Cambridge University Press 2009)

The following lecture notes are also recommended for reference:

- L. Mason, *Quantum Field Theory in Curved Space-Time*, Lecture notes.
- M. Mariño, *QFT in curved space*, Lecture notes.
- L.H. Ford, *Quantum Field theory in curved spacetime*, arXiv:gr-gc/9707062.
- P.K. Townsend, *Black Holes*, arXiv:gr-gc/9707012.
- T. Jacobson, *Introduction to Quantum Fields in Curved Spacetime and the Hawking Effect*, arXiv:gr-qc/0308048.
- E. Pajer, *Field Theory in Cosmology*, Lecture notes .
- N. Arkani-Hamed and Y. Kats, *Lecture Notes on Quantum Mechanics and Spacetime*. (PDF version available upon request)
- C. Bär and K. Fredenhagen, *Quantum Field Theory on Curved Spacetimes*, (Springer-Verlag Heidelberg 2009), DOI.
- D. Anninos, De Sitter Musings, Int. J. Mod. Phys. A 27 (2012), arXiv:hep-th/1205.3855.
- J. Penedones, TASI lectures on AdS/CFT, TASI 2015, 75-136, arXiv:hep-th/1608.04948.

Contents

About this course **i**

Synopsis i

References ii

I FUNDAMENTAL ASPECTS **1**

1 Why you should take this course **2**

1.1 What is in this course? 4

1.2 What is not? 5

2 Quantum fields in flat space **7**

2.1 Canonical quantisation in Minkowski space 7

2.2 Particle interpretation 10

2.3 Vacuum energy 12

2.4 Two-point functions 13

2.5 Charged scalars, gauge fields and spinors 17

3 Lorentzian geometry **20**

3.1 Lorentzian manifolds 20

3.2 Causal structure 21

3.3 Conformal infinity and Penrose diagrams 26

3.4 Asymptotics and peeling 30

4 Quantum fields in curved space **33**

4.1 Classical fields in curved space-time 33

4.2 Canonical quantisation in curved spacetime 36

4.3 Quantisation in generic curved spacetimes 40

4.4 Cosmological particle creation 41

5 Observables quadratic in fields 46

5.1 Two-point functions 46

5.2 Adiabatic expansion of the Green's function 47

5.3 Stress tensor renormalisation 50

II APPLICATIONS 53

6 The Unruh effect 54

6.1 Thermal states 54

6.2 Particle detectors 56

6.3 The Unruh effect 58

7 Hawking radiation 66

7.1 Quantum fields in a black hole background 66

7.2 The Hawking thermal state and friends 73

7.3 Black hole thermodynamics 75

7.4 The information paradox 77

8 Quantum fields in de Sitter space 80

8.1 de Sitter basics 80

8.2 Quantum fields in de Sitter space 82

8.3 De Sitter temperature 88

8.4 De Sitter entropy 91

9 Cosmological perturbations 94

9.1 Cosmological models 94

9.2 QFT on a cosmological background 97

9.3 Cosmological correlators 99

9.4 Fluctuating gravitons 100

10 Quantum fields in AdS **102**

10.1 Geometry of AdS spaces 102

10.2 Quantum fields in AdS 105

10.3 Towards a conformal theory on the boundary 108

10.4 The AdS/CFT correspondence 113

10.5 Hawking radiation in AdS 117

III APPENDICES **119**

A Conventions **120**

A.1 Signs, signatures and curvature 120

B Differential forms **123**

B.1 Connections and curvature 125

C Hypersurfaces **127**

D Variational calculus **129**

E Ingredients from general relativity **132**

E.1 Maximally symmetric spaces 132

E.2 Warped product manifolds and FLRW spaces 140

E.3 Black holes 142

F Hypergeometric functions **151**

G A mini introduction to CFT **152**

G.1 Conformal Transformations 152

G.2 Local Operators 153

G.3 Ward identities 154

G.4 State-Operator Map 156

G.5 Operator Product Expansion 157

G.6 Conformal Bootstrap 157

G.7 Embedding Space Formalism 158

G.8 Large N Factorization 160

Bibliography **162**

Part I

FUNDAMENTAL ASPECTS

Chapter 1

Why you should take this course

The incorporation of gravity into quantum physics remains one of the most important outstanding problems in theoretical physics. With the ever growing experimental support for both Einstein's theory of general relativity, which relies on the differential geometry of Lorentzian manifolds, as well as the standard model of elementary particle physics, based on quantum field theory, the issue of their mutual compatibility becomes increasingly pressing. In particular in situations where both theories must be simultaneously applied, such as in early universe cosmology or in the vicinity of a black hole.

Early attempts to incorporate gravity into quantum field theory by treating gravity as one of the quantum fields encountered conceptual and practical difficulties due to the intrinsic non-renormalisability.[1] As a result, new approaches such as string theory (and more controversial approaches such as loop quantum gravity or causal dynamical triangulation theory) arose. Despite their mathematical elegance, explicit computations in these frameworks are often prohibitively hard and a fully satisfactory understanding of quantum gravity remains elusive. Therefore, developing tools to access this regime where gravity and quantum effects coexist is extremely valuable.

To illustrate in what sort of regime we are interested, consider a body of mass M and size L. The "quantumness" of the interactions is governed by the dimensionless constant $\frac{\hbar}{MLc}$, where c is the speed of light and $\hbar$ is Planck's constant.[2] Quantum effects are important whenever

$$\frac{\hbar}{MLc} \gtrsim 1\,, \tag{1.1}$$

and are suppressed when $\frac{\hbar}{MLc} \ll 1$. On the other hand, the curve $\frac{G_N M}{c^2 L}$, where G_N is the effective gravitational Newton's constant, divides the regions in parameter space where the dynamical effects of gravity become important or can be ignored. When we have

$$\frac{G_N M}{c^2 L} \gtrsim 1\,, \tag{1.2}$$

the dynamical effects of gravity are important. In this situation we need the laws of general relativity to describe gravity. When $\frac{G_N M}{c^2 L} \ll 1$ gravitational effects are suppressed and we can use classical (Newtonian) gravity to describe gravitational interactions. In Figure 1.1 we sketch the various domains in this parameter space and indicate the validity of the relevant approximation in each region.

In this course we are probing the bottom right region of this diagram where both (1.1) and (1.2)

[1] A somewhat oversimplified but instructive way to see this is through counting the dimension of the coupling constant. Recall that in $d+1$ dimensions Newton's gravitational constant has mass dimension $[G_N] = 2-d$. Hence, the Einstein-Hilbert action has dimension $d-2$ which suggests that gravity is non-renormalisable in $d > 2$. There are various subtleties that need to be taken into account to make this naive argument rigorous but in the context of gravity it turns out to be correct.

[2] In the remainder of this course we almost exclusively use natural units where $\hbar = c = G_N = k_B = 1$.

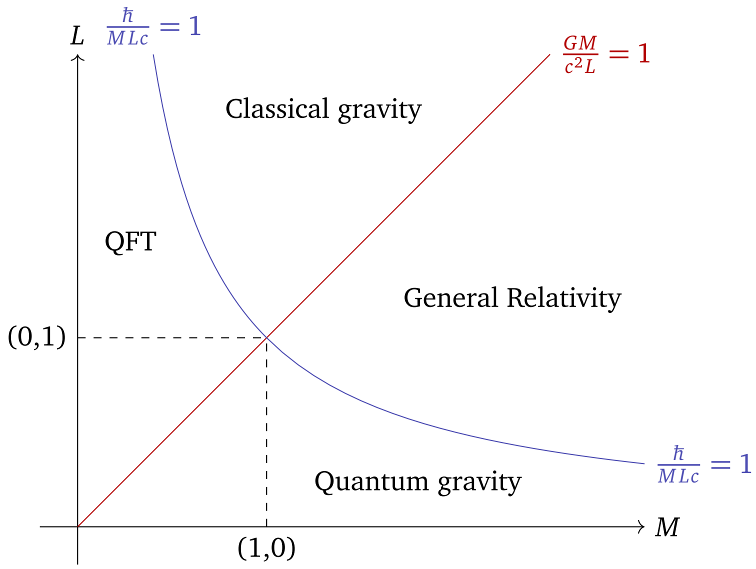


Figure 1.1: This diagram describes the range of validity of various approximations. Right of the blue hyperbola quantum effects are suppressed and we can approximate our system using classical physics, resp. Newtonian mechanics or general relativity depending on the position with respect to the red line. On the other hand, on the left of the hyperbola quantum effects are important and have to be included in an accurate description. In this case we either need QFT or a theory of quantum gravity.

are satisfied. To properly describe this situation we need an honest theory of quantum gravity. However, due to the relative weakness of the gravitational force it is often possible to neglect the backreaction of the quantum fields on the space-time as a first approximation. Consequently the problem then boils down to understanding quantum fields on generic Lorentzian space-time manifolds. In other words, when the space-time curvature is very small on the Planck scale[3] we can describe gravity classically, through general relativity, and consider fluctuations in the matter fields in the gravitational background described by a solution to the equations of motion of general relativity. In this approximation we therefore do not capture the dynamical nature of space-time itself but nonetheless this modest extension of quantum field theory is surprisingly rich in consequences. Among other things, it gives rise to the process of particle creation in cosmological and black hole space-times.

Moreover, we can go beyond this first application and perturbatively[4] include the effects of backreaction of the quantum fields on the gravitational background. We still treat the background as classical but now take the backreaction into account. To do so we have to solve the Einstein equation,

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G_N \left\langle T_{\mu\nu} \right\rangle_{\psi} \tag{1.3}$$

where the right hand side is given by the vacuum expectation value of the stress-tensor in the state $|\psi\rangle$. This approach takes into account the one-loop gravitational interactions. In principle one should be able to step-by-step add higher loop contributions but this procedure is not well understood at

[3]See Appendix A for a definition of the relevant length scales.

[4]The perturbative parameter here is $\frac{G_N}{L^2} \simeq \frac{l_p^2}{L^2} \ll \alpha$ where α parameterises the interaction strength of the matter, such as e.g. the fine structure constant $\alpha \sim \frac{1}{137}$.

present and would take us beyond the scope of these lectures.

From another point of view, studying quantum field theory in curved space-times can help to clarify various conceptual issues and teach us general lessons about the structure of quantum field theories. Traditionally, quantum field theory in Minkowski space relies on concepts such as a unique vacuum, particles, Fock spaces and S-matrices. In curved space-time many of these concepts break down and need to be reconsidered. For example, we will learn that particles are an observer-dependent concept. Rethinking these concepts leads us to a deeper understanding of quantum field theory and eventually quantum gravity. Even more, in some cases computations can be more tractable in curved space then in flat space. Many quantum field theories, such as Yang-Mills theory, suffer from various IR problems on flat space, making semi-classical methods such as instantons useless. When working on a compact space, such as the sphere, the 'radius' of the space provides a natural IR regulator and one can use semi-classical intuition once again. A modern application of this phenomenon is supersymmetric localisation, which allows one to exactly compute various observables non-perturbatively on a variety of manifolds.

For a physicist, this should be more than enough motivation to jump into the study of quantum field theory in curved spaces. It is an active area of research with many questions left unexplored. This course is but the starting point and aims to optimally equip you to attack such problems.

If you are more mathematically inclined, and do not care so much about the rich physical applications of quantum field theory in curved space, do not leave just yet. On top of all these physical applications, quantum field theory in curved spaces is mathematically very interesting in itself. Through its study many new directions in mathematics have opened up such as Donaldson-Witten theory and Chern-Simons theory, providing various new invariants characterizing manifolds. A hallmark of mathematical physics is the connection between geometry and gauge theories. Donaldson-Witten theory aims to characterize four-manifolds through geometric observables. This study was initiated by Donaldson but a breakthrough came when Witten showed that such invariants can be computed efficiently using (supersymmetric) gauge theories on said manifolds. Similarly, Chern-Simons theory in three-dimensions can be used to compute a variety of knot invariants characterising knots in three-dimensional manifolds. In this course we will not have time to dive into these topics but this course, together with the course on supersymmetry and supergravity, provides you with all the tools you need to start exploring the literature.

1.1 What is in this course?

Before we start let us take a moment to outline the scope of this course in some more detail. As this course is about quantum fields, we start in the first chapter with a brief recapitulation of the theory of quantum fields in flat space, with emphasis on those aspects which are important for our exploration in curved space. The wave equations we encounter, such as the Klein-Gordon equation or Dirac equation are linear differential equations which in flat space are straightforward to solve. Upon placing quantum field theories in curved space, a variety of subtleties arise concerning the well-definedness of said linear equations as well as issues regarding causality which now have to be addressed properly.

For this reason we continue with a careful introduction of Lorentzian geometries and their local and global causal structure in Chapter 3. Following this, we come back to the main goal of this course and proceed in Chapter 4 with the general procedure of canonical quantisation of free fields in curved space-time. This chapter will introduce all the important concepts such as the non-uniqueness of the vacuum and the concept of particle creation in a gravitational background. Following this Chapter 5 introduces how to compute Green's functions and other observables quadratic in fields, such as the stress tensor expectation value.

Having introduced the general framework, the remainder of the course will focus on applying it to a plethora of physically interesting backgrounds. The first example is the Unruh effect, discussed in Chapter 6, which describes the physics of an accelerated observer in flat space. This example perfectly sets us up for Chapter 7, where we discuss maybe the most well-known consequence of quantum field theory in curved space-time, the Hawking radiation of black holes.

Having deepened our understanding of quantum mechanical black holes we move on to cosmological consequences of quantum field theory in curved space. In the distant future, as well as during the inflationary era, our universe can be described as de Sitter universe. In Chapter 8 we start with a discussion of quantum fields in de Sitter space and explain why the second law of thermodynamics forces the universe towards a cold empty space. In Chapter 9 we take a look at more generic cosmological space-times and explain how to deal with cosmological perturbations.

Finally, the last topic of this course is quantum field theory in Anti-de Sitter space. Up to a sign, Anti-de Sitter space is identical to de Sitter space, but this sign turns out to have deep implications. In particular Anti-de Sitter space is the space in which holography is best understood. In Chapter 10 we discuss various subtleties in describing quantum fields in Anti-de Sitter space with an eye towards defining the holographic principle, which teaches us that the theory in the bulk of Anti-de Sitter can equivalently be described by a conformal theory on the conformal boundary!

In order not to obscure the message in the main text some of the background information as well as some technical details are relegated to the appendices. In Appendix A we introduce the main conventions used in these notes as well as a comparison to other conventions used in the literature. In Appendix B we review the theory of differential forms on a generic manifold $\mathcal{M}$. Appendix C contains various useful ingredients from general relativity which will be used in this course. The background on special functions and various useful identities for hypergeometric functions can be found in F. Finally, the last two appendices **??** and G contain a short review on how to deal with spinors in curved space as well as a lightning introduction to conformal field theory.

1.2 What is not?

Unfortunately, time is limited so we necessarily have to skip a variety of interesting topics. As such we will not have time to discuss the rich subject of one-loop effective actions, which describe the one-loop quantum corrections to the partition function or on-shell action of a variety of gravitational backgrounds. Similarly, we are forced to skip the rich subject of supersymmetric theories on curved backgrounds. In general putting a supersymmetric theory on a curved background breaks all the supersymmetry. However, some supersymmetry can be preserved by placing the theory in backgrounds which are solutions not to Einsteins equations of motion but those of a supergravity theory [FS11]. In particular, these two topics together provide one of the main testing grounds for the holographic correspondence and provide a active research direction furthering our understanding of the holography. Furthermore, the theory of supersymmetric theories in curved spaces gives rise to rich mathematics such as the theory of Donaldson-Witten or Chern-Simons invariants. We will not be able to cover this topic in this course but the interested reader can have a look at the lecture notes [Mar01, Moo12, Lab00, Mar05, LM05].

Another glaring omission from this course is the theory of interacting quantum fields in curved space-times. For weakly coupled theories one can use similar tools as in flat space, such as perturbation theory as will be hinted at in various instances in the main text. However, the presence of curvature and/or non-trivial topology add additional subtleties which need to be treated carefully. When the theory is strongly coupled many familiar tools break down, and even in flat space the problem of solving these theories is extremely hard and in general not well-understood. However, by putting the theories in curved space-times we can make significant progress, especially for supersymmetric theories. In particular, supersymmetric localisation provides us with a rare tool allowing us to compute a set of protected supersymmetric observables exactly.

Finally, in these lecture notes we took the approach of canonical quantisation. There are various other approaches such as path integral quantisation or algebraic quantum field theory. Path integral quantisation yields a simple, manifestly covariant framework which generalises the stationary action principle of classical mechanics by integrating over all classical paths, weighted by their action. Algebraic quantum field theory on the other hand starts from the algebra of observables acting on some Hilbert space and formalises the quantum field theory as a C^*-algebra with a set of axioms known as the Haag-Kastler axioms. Each of these approaches can be generalised to curved space-times and highlights different aspects. Each of them has their advantages and disadvantages such as the manifest (or not so manifest) presence of unitarity, covariance and locality.

Chapter 2

Quantum fields in flat space

Quantum field theory in curved space-time is a generalisation of quantum field theory in flat space. It is not surprising that in many respects the behaviour of quantum fields in curved space-time can be directly inferred from the flat space theory. Local entities, such as commutation relations or field equations are determined by the principle of general covariance and the equivalence principle and will therefore remain unchanged. On the other hand, there are various global entities which will behave radically different in curved space. For example, in Minkowski space the vacuum is unambiguously determined by Poincaré invariance. However, as we will see in Chapter 4, the concept of a vacuum becomes ambiguous in curved space.

For this reason we will take a moment to review certain aspects of quantum field theory in flat space, fix our conventions and notation and highlight certain aspects which carry over to curved space as well as those which lose their meaning. To avoid having to deal carefully with gauge invariance etc. we mostly focus on scalar fields as they suffice to illustrate the properties we are interested in. For more details on spinors, vector or higher spin fields in Minkowski space we refer the reader to their favourite textbook on quantum field theory, for example [Wei95, PS95, Sre07].

2.1 Canonical quantisation in Minkowski space

In this chapter we exclusively deal with $(d+1)$-dimensional Minkowski space $\mathbb{R}^{1,d}$. The signature is $(1,d)$ and we use mostly minus conventions such that the metric on Minkowski space is given by

$$\eta_{\mu\nu} = \operatorname{diag}(1,-1,\cdots,-1)\,. \tag{2.1}$$

We denote the coordinates collectively by $x = (x_0, x_1, \cdots, x_d)$ and often we use $x_0 = t$ for the time direction. Sometimes it is useful to separate the time direction and denote it by $x_0 = t$ while we collect the space coordinates in the vector $\mathbf{x} = (x_1, \cdots, x_d)$.

Consider a classical scalar field $\phi(x)$ in $(d+1)$-dimensional Minkowski space $\mathbb{R}^{1,d}$, satisfying the Klein-Gordon field equation,

$$\left(\Box + m^2\right)\phi = 0\,, \tag{2.2}$$

where $\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu$ is the d'Alembertian, and m is the mass of the scalar field. This field equation can be derived from the action,

$$S = \int_{\mathbb{R}^{1,d}} \mathcal{L}\left(\phi, \partial_\mu\phi\right) \mathrm{d}^{d+1}x\,, \qquad \text{with} \qquad \mathcal{L}\left(\phi, \partial_\mu\phi\right) = \frac{1}{2}\left(\eta^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi - m^2\phi^2\right), \tag{2.3}$$

by demanding that the variation δS with respect to ϕ vanishes. The conjugate momentum $\pi(x)$ is defined through the following definition,

$$\pi(x) = \frac{\delta \mathcal{L}(\phi, \partial\phi)}{\delta(\partial_t \phi)}, \tag{2.4}$$

such that for the free scalar we find $\pi(x) = \partial_t \phi$.

We proceed by defining the (classical) phase space V_ϕ as the vector space of fields satisfying the linear field equations, i.e. the Klein-Gordon equation in the case of the free scalar. There are some caveats regarding the asymptotic fall-off of these fields at infinity for the various formulae that we use to make sense. Without going into too much detail we assume that our fields have been appropriately restricted.[1] For the scalar field we denote the phase space as

$$V_\phi = \left\{\phi \in C^\infty(\mathbb{R}^{1,d}) \,\middle|\, (\Box + m^2)\phi = 0\right\} = \left\{(\phi, \partial_t\phi) \in C^\infty(\mathbb{R}^d) \times C^\infty(\mathbb{R}^d)\right\}, \tag{2.5}$$

where in the second equality we identified the phase space with the initial data along a time slice $t = t_0$. A similar phase space can be defined for fermions and gauge fields but we leave the precise definition as an exercise to the reader.[2] The phase space comes with a natural symplectic structure, or skew symmetric form,[3]

$$\Omega(\phi_1, \phi_2) = \int_{\mathbb{R}^d} \phi_1 \star \mathrm{d}\phi_2 - \phi_2 \star \mathrm{d}\phi_1 = \int_{\mathbb{R}^d} (\phi_1 \partial_t \phi_2 - \phi_2 \partial_t \phi_1)\, \mathrm{d}^d x \tag{2.6}$$

More concretely, consider a basis $\{u_\mathbf{k}(x)\}$ of the phase space, and denote their conjugate momenta by $\{\pi_\mathbf{k}(x)\}$. In this basis the symplectic form takes the canonical form $\Omega = \sum_\mathbf{k} \mathrm{d}u_\mathbf{k}(x) \wedge \mathrm{d}\pi_\mathbf{k}$ and the expression above can be obtained straightforwardly by expressing the fields ϕ_i in this basis. This skew symmetric form is dual to the Poisson bracket on V_ϕ which can be expressed as,

$$\left\{\phi(x), \pi(x')\right\} = \delta^{(d)}(x - x'), \qquad \left\{\phi(x), \phi(x')\right\} = \left\{\pi(x), \pi(x')\right\} = 0. \tag{2.7}$$

Canonical quantisation then proceeds by promoting the fields and canonical momenta to operators on an appropriate Hilbert space and imposing the canonical (equal time) commutation relations, [4]

$$\begin{aligned} \left[\phi(t,\mathbf{x}), \phi(t,\mathbf{x}')\right] &= 0, \\ \left[\pi(t,\mathbf{x}), \pi(t,\mathbf{x}')\right] &= 0, \\ \left[\phi(t,\mathbf{x}), \pi(t,\mathbf{x}')\right] &= \mathrm{i}\,\delta(\mathbf{x} - \mathbf{x}'). \end{aligned} \tag{2.8}$$

To develop the quantum theory we relate the phase space to a one-particle Hilbert space $\mathcal{H}$ and define

[1] For the interested reader: The relevant space is often taken to be the Schwartz space, which consists of infinitely differentiable functions that at infinity fall off faster than any reciprocal power of x. Crucially, this space has the property that it allows for the Fourier transform to be applied.

[2] Note that the Dirac equation is first order so that the initial value problem only needs the value of the field as boundary condition. For vector fields one has to deal with gauge invariance and quotient out gauge equivalent field configurations.

[3] For fermionic fields the phase space comes with a symmetric form.

[4] Note that here and throughout the text we put $\hbar = 1$ explaining the absence of the characteristic factor of $\hbar$ on the right hand side of the equations in (2.8).

the Fock space to be

$$\mathcal{F} = \mathbb{C} \oplus \mathcal{H} \oplus (\mathcal{H} \otimes \mathcal{H}) \oplus \cdots = \bigoplus_{n=0}^{\infty} \otimes^n \mathcal{H} \,. \tag{2.9}$$

We can make this more precise using the Fourier transform, by decomposing complex fields into plane waves. An appropriately normalised set of solutions to the wave equation (2.2) is given by the following plane waves,

$$u_{\mathbf{k}}(t, \mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d 2\omega}} \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{x} - \omega t)} \,, \tag{2.10}$$

with $\omega = \sqrt{\mathbf{k}^2 + m^2}$. The plane waves $u_{\mathbf{k}}$ are called positive frequency or positive energy solutions with respect to t, while the the complex conjugate solutions, $u^*_{\mathbf{k}}$, are the negative frequency/energy solutions. They are eigenfunctions of the time translation operators with eigenvalues $\mp \mathrm{i}\omega$,

$$\partial_t u_{\mathbf{k}} = -\mathrm{i}\, \omega \, u_{\mathbf{k}} \,, \qquad \partial_t u^*_{\mathbf{k}} = \mathrm{i}\, \omega \, u^*_{\mathbf{k}} \,. \tag{2.11}$$

Such plane waves, together with their complex conjugates, form a complete set of solutions and therefore any solution to the Klein-Gordon equation can be Fourier expanded as

$$\phi(x) = \int \mathrm{d}^d \mathbf{k} \left[a_{\mathbf{k}} u_{\mathbf{k}}(x) + a^\dagger_{\mathbf{k}} u^*_{\mathbf{k}}(x) \right] . \tag{2.12}$$

Such that the phase space for a complex scalar decomposes as $\mathbb{C} \otimes V_\phi = V^+_\phi \oplus V^-_\phi$ and the choice of ω above is precisely made so that the fields satisfy the on-shell condition, $k^2 = m^2$.

The skew symmetric form on the phase space translates to a (positive definite) inner product on the Hilbert space defined as,

$$\langle \phi_1, \phi_2 \rangle = \mathrm{i} \int \mathrm{d}^d \mathbf{x} \, (\phi_1(x) \partial_t \phi_2(x) - \partial_t \phi_1(x) \phi_2(x)) \,. \tag{2.13}$$

The main property of this inner product is that for two solutions to the Klein-Gordon equation the product is conserved under time translation. The plane waves defined in (2.10) are orthonormal with respect to this product,

$$\langle u_{\mathbf{k}}, u_{\mathbf{k}'} \rangle = \delta(\mathbf{k} - \mathbf{k}') \,, \qquad \langle u_{\mathbf{k}}, u^*_{\mathbf{k}'} \rangle = 0 \,. \tag{2.14}$$

Exercise 2.1. *Prove that the plane waves $u_{\mathbf{k}}$ together with their complex conjugates form an orthonormal basis of $L^2(\mathbb{R}^{1,d})$, i.e. square integrable functions on $\mathbf{R}^{1,d}$.*[5]

We can therefore express the Fourier coefficients as follows,

$$a_{\mathbf{k}} = \langle u_{\mathbf{k}}, \phi \rangle \,, \qquad a^\dagger_{\mathbf{k}} = \langle u^*_{\mathbf{k}}, \phi \rangle \,. \tag{2.15}$$

From the equal time commutation relations (2.8), it follows that

$$\left[a_{\mathbf{k}}, a^\dagger_{\mathbf{k}'} \right] = \delta(\mathbf{k} - \mathbf{k}') \,, \qquad \left[a_{\mathbf{k}}, a_{\mathbf{k}'} \right] = \left[a^\dagger_{\mathbf{k}}, a^\dagger_{\mathbf{k}'} \right] = 0 \,, \tag{2.16}$$

[5]Note that strictly speaking plane waves do not belong to $L^2(\mathbf{R}^{1,d})$ but nonetheless they form a complete basis for it. Similar comments apply to the dual basis given by the delta functions, $\delta(x - x_*)$.

Exercise 2.2. *Prove the commutation relations,* (2.16) *, for the creation and annihilation operators starting from* (2.8).

Remark. In the above, $\mathbf{k}$ is a continuous parameter. Sometimes it is useful to introduce periodic boundary conditions on a torus T^{d+1} of volume L^{d+1}, a.k.a. putting the theory in a box. In particular, this avoids all kinds of volume divergences. Doing so, the wave-vectors become quantised $\mathbf{k} = \frac{2\pi}{L}\mathbf{n}$ and all integrals in the above get replaced by infinite sums.

It is often easier to compute physical quantities using such a (IR) regulator and taking the limit $L \to \infty$ at the end of the computation to obtain the infinite volume result.

When the quantum theory enjoys some global symmetries, there will be associated conserved currents. The theories we consider in this course are all covariant under space-time diffeomorphisms. In particular, they are invariant under space and time translations. The associated conserved current is given by the stress tensor,

$$\partial^\mu T_{\mu\nu}, \qquad \text{where} \qquad T_{\mu\nu} = \frac{\delta \mathcal{L}}{\delta \partial^\mu \phi} \partial_\nu \phi - \delta_{\mu\nu} \mathcal{L} . \tag{2.17}$$

Note that in order for this to serve as a source for the Einstein equations we need to improve it to be symmetric under interchanging the indices. We will present a manifestly symmetric stress tensor formulation of the stress tensor later when dealing with curved space-times. Similarly, if the theory has additional global symmetries we can write the conserved current as

$$\partial^\mu J_\mu = 0, \qquad \text{where} \qquad J^\mu = \frac{\delta \mathcal{L}}{\delta \partial^\mu \phi} \delta\phi - T_{\mu\nu} \delta x^\nu , \tag{2.18}$$

where $\delta\phi = \phi' - \phi$ denotes the change of the field under an infinitesimal symmetry transformation and similarly $\delta x^\nu = x^{\nu\prime} - x^\nu$. For an example, see the complex scalar field below.

2.2 Particle interpretation

We can interpret the operators $a_{\mathbf{k}}$ and $a_{\mathbf{k}'}$ as annihilation and creation operators for an infinite amount of harmonic oscillators labelled by their momentum. In the Heisenberg picture, the states span a Hilbert space. A convenient basis for this Hilbert space is given by the Fock representation introduced above.

The state space consists of the vacuum $|0\rangle$, which is annihilated by all the annihilation operators,

$$a_{\mathbf{k}} |0\rangle = 0 . \tag{2.19}$$

All the excited states in the Hilbert space can be constructed by acting on the vacuum with the creation operators $a^\dagger_{\mathbf{k}}$,

$$a^\dagger_{\mathbf{k}} |0\rangle = |1_{\mathbf{k}}\rangle . \tag{2.20}$$

Successively acting with the creation operators we can then construct the most general states,

$$\prod_{i=1}^{p} \frac{1}{\sqrt{n_i!}} \left(a^\dagger_{\mathbf{k}_i}\right)^{n_i} |0\rangle = \left| n_1^{(\mathbf{k}_1)} n_2^{(\mathbf{k}_2)} n_3^{(\mathbf{k}_3)} \cdots n_p^{(\mathbf{k}_p)} \right\rangle . \tag{2.21}$$

These basis vectors are normalised such that

$$\left\langle n_1^{(\mathbf{k}_1)} n_2^{(\mathbf{k}_2)} n_3^{(\mathbf{k}_3)} \cdots n_p^{(\mathbf{k}_p)} \middle| m_1^{(\mathbf{k}_1)} m_2^{(\mathbf{k}_2)} m_3^{(\mathbf{k}_3)} \cdots m_q^{(\mathbf{k}_q)} \right\rangle = \delta_{pq} \sum_{\sigma} \delta_{n_1 m_{\sigma(1)}} \delta_{n_2 m_{\sigma(2)}} \cdots \delta_{n_p m_{\sigma(p)}} , \tag{2.22}$$

where the sum runs over all permutations σ of the integers $1, ..., p$.

In analogy with the harmonic oscillator we can then introduce the number operator,

$$N = \sum N_{\mathbf{k}} , \qquad N_{\mathbf{k}} = a^\dagger_{\mathbf{k}} a_{\mathbf{k}} , \tag{2.23}$$

whose expectation value in a generic state is given by

$$\langle \psi | N | \psi \rangle = \sum_i n_i , \tag{2.24}$$

where $|\psi\rangle = \left| n_1^{(\mathbf{k}_1)} n_2^{(\mathbf{k}_2)} n_3^{(\mathbf{k}_3)} \cdots n_p^{(\mathbf{k}_p)} \right\rangle$. Hence, we can interpret $N_{\mathbf{k}}$ and N as counting the number of quanta with momentum $\mathbf{k}$ and the total number of quanta respectively.

Note that the vacuum, as defined in (2.19) is unique. Naively it may seem to depend on a choice of inertial frame but an easy argument shows otherwise. Indeed, consider a second inertial frame $\tilde{x}^\mu = \Lambda^\mu{}_\nu x^\nu$ with Λ a Lorentz transformation. Analogous to the above we can define the positive frequency functions

$$\tilde{u}_{\tilde{\mathbf{k}}}(\tilde{t}, \tilde{\mathbf{x}}) = \frac{1}{\sqrt{(2\pi)^d 2\tilde{\omega}}} \mathrm{e}^{\mathrm{i}(\tilde{\mathbf{k}}\cdot\tilde{\mathbf{x}} - \tilde{\omega}\tilde{t})} . \tag{2.25}$$

and expand the field ϕ as

$$\phi(\tilde{x}) = \mathrm{d}^d \tilde{\mathbf{k}} \left[\tilde{a}_{\tilde{\mathbf{k}}} \tilde{u}_{\tilde{\mathbf{k}}}(\tilde{x}) + \tilde{a}^\dagger_{\tilde{\mathbf{k}}} \tilde{u}^*_{\tilde{\mathbf{k}}(x)} \right] . \tag{2.26}$$

The "new" vacuum is then defined by the condition that

$$\tilde{a}_{\mathbf{k}} \left|\tilde{0}\right\rangle = 0 . \tag{2.27}$$

To show that this is nothing but the old vacuum consider the mode function and notice that

$$\tilde{u}_{\mathbf{k}} = \frac{1}{\sqrt{(2\pi)^d 2\omega}} \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\tilde{\mathbf{x}} - \omega\tilde{t})} = \left(\frac{\tilde{\omega}}{\omega}\right)^{\frac{1}{2}} \frac{1}{\sqrt{(2\pi)^d 2\tilde{\omega}}} \mathrm{e}^{\mathrm{i}(\tilde{\mathbf{k}}\cdot\mathbf{x} - \tilde{\omega}t)} = \left(\frac{\tilde{\omega}}{\omega}\right)^{\frac{1}{2}} u_{\tilde{\mathbf{k}}} . \tag{2.28}$$

Since we restrict to the orthochronous subgroup of the Lorentz group, to preserve time orientation, we have that $\tilde{\omega} > 0$ implies $\omega > 0$ and thus we find that

$$a_{\tilde{\mathbf{k}}} \left|\tilde{0}\right\rangle = 0 \quad \forall \tilde{\mathbf{k}} \qquad \Rightarrow \qquad a_{\mathbf{k}} |0\rangle = 0 \quad \forall \mathbf{k} , \tag{2.29}$$

and the converse follows by symmetry. Hence the vacuum is indeed unique and independent of the choice of frame.

2.3 Vacuum energy

To further explore the meaning of the Fock states we can compute their energy and momentum, which can be obtained from the expectation value of the energy-momentum tensor which for a scalar field is given by

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - \frac{1}{2}\eta_{\mu\nu}\left(\partial^\rho \phi \partial_\rho \phi - m^2 \phi^2\right) . \tag{2.30}$$

We can define the conserved momentum operator as

$$P_\mu = \int_{t=t_0} T_{\mu 0} \mathrm{d}^d \mathbf{x} , \tag{2.31}$$

with the time-like component $P_0 = H$, the Hamiltonian. The expection value in a state $|\psi\rangle$ can then be computed as $\langle\psi| H |\psi\rangle$.

Exercise 2.3. *Give an expression for the Hamiltonian and conserved momentum in terms of the creation and annihilation operators and show that they commute with the number operator N.*

Naively compute the expectation value of the Hamiltonian density H and momentum density P_μ and show that they are divergent.

Similarly, consider the expectation value of the energy-momentum density $T_{\mu\nu}$ in the vacuum as well as in a generic state and find an expression in terms of the mode functions.

Computing the energy we encounter our first infinite result. Such troubling results are well-known to plague the subject of quantum field theory but can be cured through renormalisation. In this setup it will suffice to define the normal ordering operation, $: \bullet :$, which is understood to act on products of annihilation and creation operators such that it puts all the annihilation operators on the right of the creation operators.

Exercise 2.4 (Vacuum energy divergence)**.** *Compute the momentum and energy of the vacuum and show that it naively diverges. Use the normal ordering prescription to regularise this result and find the resulting vacuum energy.*

In the above we regularised the vacuum energy by passing through a normal ordered prescription which simply throws away the infinite contribution. However, as already mentioned before, in curved space, especially when gravity is included, the energy of the vacuum is physical since it gravitates. For this reason it is instructive to take a closer look at the vacuum expectation value of the energy in a situation where it becomes important. Consider a massless neutral scalar in $(3+1)$-dimensional Minkowski space in the presence of two parallel plates at $x_3 = 0$ and $x_3 = a$ with $a \ll 1$ and impose boundary conditions $\phi(0) = \phi(a)$ on the plates. In addition we impose periodic boundary conditions in the two other spatial directions with $x_{1,2} \sim x_{1,2} + L$, with $L \gg a$. This setup represents a modified version of the usual Casimir effect for two neutral conducting plates in a vacuum electric field vanishing on the plates.

Exercise 2.5. *Quantise the scalar field in the presence of the two plates and show that the average energy density is given by*

$$\rho(a) = a^{-1}L^{-2}\langle 0|\, T_{00}\,|0\rangle_a = \frac{1}{2aL^2}\sum_{\mathbf{k}}\omega_k\,, \tag{2.32}$$

and give an expression for ω_k.

We can regularise this sum by writing

$$\rho(a) = -\frac{1}{2aL^2}\lim_{\epsilon\to 0}\frac{\mathrm{d}}{\mathrm{d}\epsilon}\sum_{\mathbf{k}}\mathrm{e}^{-\epsilon\omega_k} \tag{2.33}$$

Remark. The parameter ϵ here is dimensionful so it is good practice to introduce an explicit length scale $\epsilon \to \epsilon/\Lambda$ so that ϵ becomes dimensionless. In the final result all the Λ dependence should drop out as can be easily verified.

Exercise 2.6. *Compute the regularised vacuum energy and show that the sum $S(\epsilon,a) = L^{-2}\sum_{\mathbf{k}}\exp(-\epsilon\omega_k)$, takes the form*

$$\pi S(\epsilon,a) = aG(\alpha) + \frac{\pi^3}{45a^3}\epsilon + \mathcal{O}(\epsilon^2)\,. \tag{2.34}$$

Before taking the limit in (2.33), subtract the infinite part to arrive at the renormalised vacuum energy. Doing so compute $\rho(a)$.

What would have changed if we instead imposed vanishing boundary conditions at the plates?

Remark. In the last exercise we proceeded in a rather cavalier way. In order to obtain the same result in a mathematically more satisfying manner one can employ ζ-function regularisation. We invite the interested reader to explore this method and repeat this exercise in a more rigorous manner.

2.4 Two-point functions

On top of the vacuum energy and Hilbert space, much information about a quantum field theory is encoded in its n-point functions. Since we are so far only working with free scalars, all the non-trivial information is encoded in the two-point functions, i.e. the Green functions of the wave equations. There are various types of Green functions depending on the choice of integration contour in the complex plane.

A useful set of Green functions are given by the following expectation values,

$$\begin{aligned}
\mathrm{i}G(x_1,x_2) &= \langle 0|\,[\phi(x_1),\phi(x_2)]\,|0\rangle\,,\\
G^{(1)}(x_1,x_2) &= \langle 0|\,\{\phi(x_1),\phi(x_2)\}\,|0\rangle\,,\\
G^{+}(x_1,x_2) &= \langle 0|\,\phi(x_1)\phi(x_2)\,|0\rangle\,,\\
G^{-}(x_1,x_2) &= \langle 0|\,\phi(x_2)\phi(x_1)\,|0\rangle\,,
\end{aligned} \tag{2.35}$$

G is known as the Pauli-Jordan, or Schwinger function, while $G^{(1)}$ is called the Hadamard elementary function. These Green functions can be split into their positive and negative part,

$$\mathrm{i}G = G^+ - G^-\,, \qquad \text{and} \qquad G^{(1)} = G^+ + G^-\,, \tag{2.36}$$

where the positive and negative part $G^\pm$ are known as the Wightman functions. Finally, the Feynman propagator G_F and retarded/advanced Green functions $G_{R/A}$ are defined as

$$G_F(x_1,x_2) = \langle 0|\, T\,[\phi(x_2)\phi(x_1)]\,|0\rangle = \theta(t_1-t_2)G^+(x_1,x_2) + \theta(t_2-t_1)G^-(x_1,x_2)\,, \tag{2.37}$$

where T denotes time ordering, and

$$G_R(x_1,x_2) = \theta(t_1-t_2)G(x_1,x_2)\,, \qquad G_A(x_1,x_2) = -\theta(t_2-t_1)G(x_1,x_2)\,. \tag{2.38}$$

All these two-point functions have the same form in momentum space

$$\widetilde{G}(k) = \frac{-\mathrm{i}}{k^2+m^2}\,, \tag{2.39}$$

but they have different iϵ prescriptions and are used in different contexts. Time ordered products, such as the Feynman or retarded/advanced Green's functions are relevant for S-matrix calculations. The Green's functions involving (anti-)commutators on the other hand are useful to describe how the field responds to a source. Finally, the Wightman functions are useful since they describe the effect of the field on a moving detector, as will be described in Chapter 6.

Exercise 2.7. *Show that the average of the retarded and advanced Green functions $\overline{G} = \frac{1}{2}(G_R + G_A)$ is given by*

$$G_F(x_1,x_2) = \mathrm{i}\overline{G}(x_1,x_2) + \frac{1}{2}G^{(1)}(x_1,x_2)\,. \tag{2.40}$$

Using the field equations (2.2) one can show that the Green functions $\mathcal{G} \in \{G\,, G^{(1)}\,, G^\pm\}$ all satisfy the homogeneous equation

$$(\Box_{x_1} + m^2)\mathcal{G}(x_1,x_2) = 0\,, \tag{2.41}$$

while the Feynman and retarded/advanced Green functions satisfy

$$\begin{aligned}(\Box_{x_1} + m^2)G_F(x_1,x_2) &= -\,\delta^{(d+1)}(x_1-x_2)\,,\\ (\Box_{x_1} + m^2)G_{A/R}(x_1,x_2) &= \delta^{(d+1)}(x_1-x_2)\,.\end{aligned} \tag{2.42}$$

We can use translation invariance to translate one of the points, say x_2 to the origin in which case we can denote the Green functions by $\mathcal{G}(x) = \mathcal{G}(x,0)$. Having done so, an integral representations for the Green functions can be obtained by substituting the mode decomposition (2.12) in the definitions above. All the Green functions can then be represented as

$$\mathcal{G}(x) = \frac{1}{(2\pi)^{d+1}}\int \mathrm{d}^{d+1}k\,\frac{\mathrm{e}^{-\mathrm{i}k\cdot x}}{k^2-m^2}\,. \tag{2.43}$$

This integral has poles at $\omega = \pm\sqrt{|\mathbf{k}^2 + m^2|}$ and the various Green functions correspond to various contour prescriptions for the integration. In Figure 2.1, the contours are shown for the various Green functions.

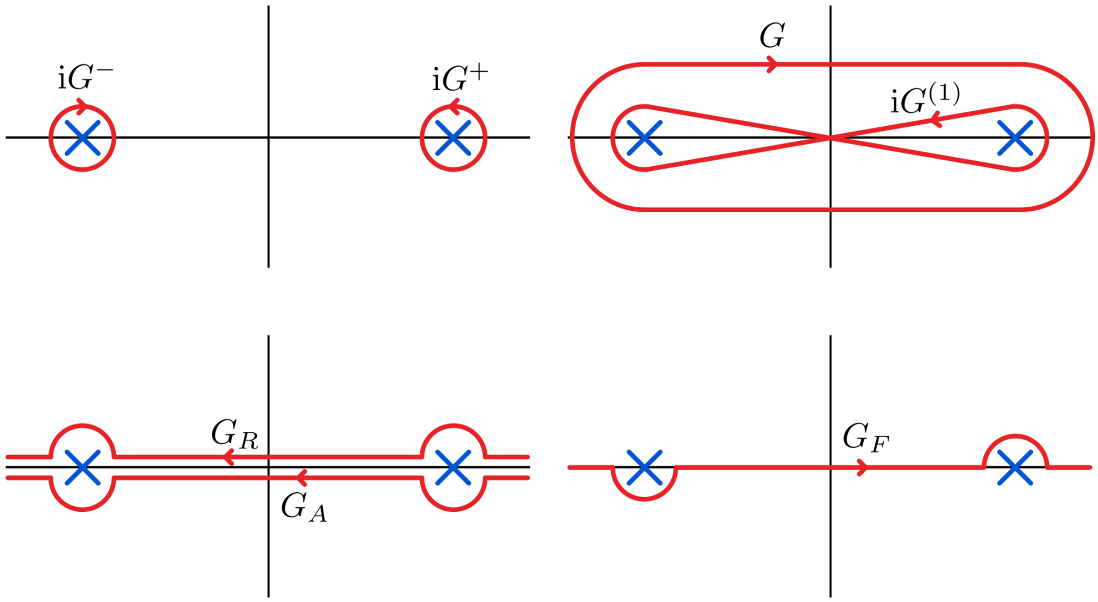


Figure 2.1: THe various Green functions are associated with the above contours for the integral (2.43). The open contours should be interpreted as closed by an infinitely large semicirle in the upper/lower plane.

From (2.36)–(2.38) we can see that we can obtain all the Green functions from the Wightman functions $G^{\pm}$ so we mostly focus on those.

Exercise 2.8. *Use the mode expansion of the scalar field to show that the Green functions take the manifestly Lorentz invariant form* (2.43).

Example 2.1 (Massless scalar). *To illustrate the formalism let us consider a massless scalar in four-dimensional Minkowski space. To compute the Wightman functions perform the relevant contour integration as indicated in Figure 2.1. We focus on G^+ but the computation for G^- is entirely analogous. Near the pole the integrand reduces to*

$$I_+ = \frac{1}{2\omega_{\mathbf{k}}} \mathrm{e}^{-ik\cdot x}\,, \tag{2.44}$$

where $\omega_{\mathbf{k}} = |\mathbf{k}|$. Using Cauchy's residue theorem we obtain

$$G^+(x) = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3 2|\mathbf{k}|} \mathrm{e}^{-\mathrm{i}(|\mathbf{k}|t - \mathbf{k}\cdot\mathbf{x})}\,. \tag{2.45}$$

We can use spherical coordinates to rewrite this as

$$G^+(x) = \int \frac{|\mathbf{k}|\mathrm{d}|\mathbf{k}|\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi}{2(2\pi)^3} \mathrm{e}^{\mathrm{i}|\mathbf{k}|(r\cos\theta - t)} = -\frac{\mathrm{i}}{2(2\pi)^2 r}\int_0^\infty \mathrm{d}|\mathbf{k}|\left(\mathrm{e}^{-\mathrm{i}|\mathbf{k}|(t-r)} - \mathrm{e}^{-\mathrm{i}|\mathbf{k}|(t+r)}\right). \tag{2.46}$$

This integral is divergent and needs to be regularised. We do so by shifting $t \to t - \mathrm{i}\epsilon$, with $\epsilon > 0$ so that

$$\int_0^\infty \mathrm{d}|\mathbf{k}|\mathrm{e}^{-\mathrm{i}|\mathbf{k}|(t - \mathrm{i}\epsilon \pm r)} = -\frac{\mathrm{i}}{t \pm r - \mathrm{i}\epsilon}\,. \tag{2.47}$$

This has to be understood as a distribution, since we have (Sokhotski-Plemelj theorem)

$$\lim_{\epsilon\to 0}\frac{1}{x\pm \mathrm{i}\epsilon}=\mathcal{P}\frac{1}{x}\mp \mathrm{i}\pi\delta(x)\,, \tag{2.48}$$

where $\mathcal{P}$ denotes the Cauchy principal value. Using this we find the Wightman function

$$G^+(x)=-\frac{1}{4\pi^2}\mathcal{P}\frac{1}{t^2-r^2}+\frac{\mathrm{i}}{8\pi r}\left(\delta(t+r)-\delta(t-r)\right). \tag{2.49}$$

Exercise 2.9. *Use this result to compute the Feynman propagator and Hadamard's elementary function,*

$$G_F(x)=\frac{\mathrm{i}}{4\pi^2}\mathcal{P}\frac{1}{x^2}-\frac{1}{4\pi}\delta(x^2)\,,\qquad G^{(1)}=-\frac{1}{2\pi^2}\mathcal{P}\frac{1}{x^2}\,. \tag{2.50}$$

Exercise 2.10 (Massive scalar)**.** *Use spherical coordinates in the spatial directions to simplify the integral and find an explicit expression for the Wightman functions for a massive scalar.*

Show that for space-like separated points, the Wightman function takes the form

$$G^+(x_1,x_2)=\frac{m}{4\pi^2 s}K_1(ms)\,,\qquad s=\sqrt{-(x_1-x_2)^2}\,, \tag{2.51}$$

where s is the proper distance and K_1 is a Bessel function of the second kind. Show that for timelike separated points it becomes

$$G^+(x_1,x_2)=\frac{\mathrm{i}m}{8\pi\tau}H_1^{(2)}(m\tau)\,,\qquad \tau=\sqrt{(x-y)^2}\,, \tag{2.52}$$

where τ is the proper time and $H_1^{(2)}$ is a Hankel function of the second kind.

What is the behaviour for large space-like separation $r\gg 1$? Give a physical explanation of this behaviour.

What happens for light-like separated points? You can study this by taking the limit $s\to 0$ in (2.51) or $\tau\to 0$ in (2.52). Show that there is a branch point singularity in the Wightman distribution. This essential singularity is known as the lightcone singularity of Wightman functions at null separation.

In order to perform the calculations of the (Feynman) Green's function it is often useful to rotate the contour by 90 degrees to obtain the Euclidean Green's function G_E. The integration variables are changed to $\kappa=-\mathrm{i}\omega$ and similarly we replace $\tau=-\mathrm{i}t$, such that we have the relation

$$G_F(t,\mathbf{x})=-\mathrm{i}G_E(\mathrm{i}\tau,\mathbf{x})\,. \tag{2.53}$$

where

$$G_E(\tau,\mathbf{x})=\frac{1}{(2\pi)^{d+1}}\int\frac{\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{x}+\omega\tau)}}{\omega^2+|\mathbf{k}|^2+m^2}\mathrm{d}\kappa\,\mathrm{d}^d\mathbf{k}\,,\qquad (\Box-m^2)G_E=-\delta^{(d+1)}(x)\,, \tag{2.54}$$

where $\Box$ now denotes the d'Alembertian on $(d+1)$-dimensional Euclidean space. The advantage of the Euclidean theory is that the operator $\Box-m^2$ has a unique well-defined inverse because the poles now lie on the imaginary rather than the real axis. Hence it is often much easier to work in Euclidean space and Wick rotate the result to obtain the Lorentzian Feynman Green's function. Note that this only works for the Feynman propagator, as all of the other contours in Figure 2.1 can not be rotated without crossing poles. A useful way to compute Euclidean Green's functions is using heat kernel regularisation, as demonstrated in the following exercise.

Exercise 2.11. *An alternative and often useful way to compute Green's functions is by analytically continuing the expressions to Euclidean signature and using heat kernel regularisation.*

Let us consider the Laplacian operator $A = -\Box$ in Euclidean space $\mathbb{R}^d$. The heat kernel of an operator A is defined as

$$K(x,x';\tau) = \langle x|\,\mathrm{e}^{-\tau A}\,\big|x'\big\rangle\,. \tag{2.55}$$

Show that, in the case of the Laplacian, the heat kernel solves the heat equation

$$\Box_x K(x,x';\tau) = \frac{\partial K(x,x';\tau)}{\partial \tau}\,, \tag{2.56}$$

with boundary condition

$$K(x,x';0) = \delta(x-x')\,. \tag{2.57}$$

Show that the Euclidean Green's function of A is then given by

$$G(x,x') = \int_0^\infty K(x,x';\tau)\mathrm{d}\tau\,, \tag{2.58}$$

and use the explicit expression for the heat kernel to derive the Euclidean Green's function for a massless scalar in $\mathbb{R}^d$. Compare the result with a direct calculation of $G^+(x,x')$ in four-dimensional Minkowski space-time.

The last Green's function we want to introduce is the thermal Green's function about which we will have much more to say in Chapter 6. Instead of looking at the Green's functions in the vacuum state, the thermal Green's functions $G_\beta(x)$ are obtained by considering a thermal state at temperature $T = \frac{1}{\beta}$. These Green's functions have the important property that they are periodic in imaginary time,

$$G_\beta(t,\mathbf{x}) = G_\beta(t+\mathrm{i}\beta,\mathbf{x})\,. \tag{2.59}$$

2.5 Charged scalars, gauge fields and spinors

To finish this chapter, we briefly consider charged scalars, gauge fields and Dirac spinors. Many details are omitted as this section mainly serves to set our notation and conventions.

So far we considered real, neutral scalar fields. A charged scalar on the other hand can be described by a pair of Hermitian scalar fields ϕ_1 and ϕ_2 which we can collect in a single complex field

$$\phi(x) = \phi_1(x) + \mathrm{i}\phi_2(x)\,. \tag{2.60}$$

The Lagrangian density is given by the slightest generalisation of (2.3) as

$$\mathcal{L} = \eta^{\mu\nu}\partial_\mu\phi^\dagger\partial_\nu\phi - m^2\phi^\dagger\phi\,, \tag{2.61}$$

which is invariant under the global symmetry transformation

$$\phi(x) \to \phi'(x) = e^{i\alpha}\phi(x), \qquad \alpha \in \mathbb{R}. \tag{2.62}$$

Exercise 2.12. *What is the conserved current for this symmetry J_μ? And what is the symmetry generator $G(\bullet) = \int d^d J_0 \bullet$?*

Show that the generator can be written as

$$G = \alpha \sum_{\mathbf{k}} \left(N^- - N^+\right), \tag{2.63}$$

where $N^\pm$ are respectively the number operators for positively and negatively charged particles.

We can couple the charged scalar field to an external electromagnetic field with field strength $F = dA$.[6] The minimal coupling prescription is defined by replacing the derivatives ∂ in $\mathcal{L}$ with covariant derivatives $D = \partial + igA$, where g is the coupling constant. The Lagrangian is then invariant under the local symmetry transformation

$$\begin{aligned} A_\mu(x) \to& A'_\mu(x) = A_\mu(x) - \frac{1}{g}\partial_\mu\theta(x), \\ \phi(x) \to& \phi'(x) = e^{i\theta(x)}\phi(x). \end{aligned} \tag{2.64}$$

If the gauge field A_μ is dynamic we must add to the Lagrangian a Maxwell term, gauging the original global symmetry,

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + D^\mu\phi^\dagger D_\mu\phi - m^2\phi^\dagger\phi. \tag{2.65}$$

Remark. In the above we considered a U(1) gauge field. If ϕ is instead charged under a G symmetry we can proceed analogously. The Lagrangian remains unchanged but the gauge field A now transforms in the adjoint representation of the gauge group G,

$$A_\mu(x) = A^a_\mu(x)T^a, \tag{2.66}$$

where T^a are the generators of the gauge group which satisfy

$$\left[T^a, T^b\right] = i f^{ab}{}_c T^c, \tag{2.67}$$

with the structure constants $f^{ab}{}_c$. The (covariant) field strength in turn is given by

$$F = -ig^{-1}[D, D] = dA + ig[A, A]. \tag{2.68}$$

In this case the kinetic term for the gauge fields, also known as the Yang-Mills action contains a quartic interaction term, making this case significantly harder to analyse.

---

[6] If this notation is unfamiliar, see Appendix B for a review on differential forms.

Finally, the Dirac spinor ψ is governed by the action

$$S = \overline{\psi}\,(\mathrm{i}\partial\!\!\!/ - m)\,\psi\,, \qquad\qquad \partial\!\!\!/ = \gamma^\mu \partial_\mu\,. \tag{2.69}$$

where the gamma matrices γ^μ satisfy the following commutation relations

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}\,. \tag{2.70}$$

The Dirac adjoint is defined as $\overline{\psi} = \psi^\dagger \gamma^0$ and transforms such that the Lagrangian density is a scalar.

For more information on the quantisation of Maxwell theory, Yang-Mills theory or the (charged) Dirac spinor we refer the reader to any textbook on quantum field theory cited above. For additional information on representations of gamma matrices and minimal spinors in various dimensions see [VP99, FVP12].

Chapter 3

Lorentzian geometry

In the previous chapter, we reviewed some salient features of quantum field theory in flat space. In particular we focused on the canonical quantisation of free theories and the concepts of particles and vacuum energy. In a classical field theory, one obtains the physical field configurations through variational principles, i.e. Euler-Lagrange equations etc. Once a solution has been found its stability can be studied through a local analysis in field space. Quantisation on the other hand is a global procedure where we need to take into account the full phase space. Indeed, already in quantum mechanics there is the possibility for a particle to jump over any potential barrier allowing it to probe the full phase space.

As we saw in the previous chapter, in order to quantise the theory (using canonical quantisation) we need a complete set of solutions to certain (linear) wave equations. Before moving on to quantum fields in curved space, we will therefore need some basic global notions from Lorentzian geometry. In particular, we focus on carefully defining causality and various related concepts, such as Cauchy hypersurfaces and global hyperbolic manifolds. On such manifolds, the above mentioned wave functions behave particularly nicely. More details can be found in for example [Wal84, O'N83].

3.1 Lorentzian manifolds

Before stating the definition of a Lorentzian manifold let us start to define what we mean by a Lorentzian scalar product on a vector space V.

Definition 3.1. Let V be a $(d+1)$-dimensional real vector space. A Lorentzian scalar product on V is a non-degenerate symmetric bilinear form, $\langle\cdot,\cdot\rangle$, of signature $(1,d)$.

This means that we can find a basis $\{e_\mu\}$, $\mu = 1,\ldots,d+1$, of V such that

$$\langle e_\mu, e_\nu\rangle = \eta_{\mu\nu}\,. \tag{3.1}$$

where $\eta_{\mu\nu}$ is the usual Minkowski metric, $\eta_{\mu\nu} = \mathrm{diag}(1,-1,\ldots,-1)$. With this definition at hand we can now give a precise definition of a Lorentzian manifold.

Definition 3.2. A Lorentzian manifold is a pair $(\mathcal{M},g)$, where $\mathcal{M}$ is a smooth $(d+1)$-dimensional manifold, and g is a Lorentzian metric, i.e. g associates with each point $p\in\mathcal{M}$ a Lorentzian scalar product g_p on the tangent space $T_p\mathcal{M}$.

As usual in differential geometry we require that g_p depends smoothly on p. For a choice of local coordinates, $(x_1,\ldots,x_{d+1}) : U \to V$, where $U \subset \mathcal{M}$ and $V \subset \mathbb{R}^{1,d}$ are open subsets, and for any

$\mu, \nu = 1, \ldots, d+1$, the functions $g_{\mu\nu} : V \to \mathbb{R}$, defined by $g(\partial_\mu, \partial_\nu)$, are smooth. Here $\partial_\mu = \frac{\partial}{\partial x_\mu}$ denote the usual coordinate vector fields. With respect to these coordinates we write the line element $\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu \otimes \mathrm{d}x^\nu$.

Before we proceed let us give a few examples.

Example 3.1 (Minkowski space)**.** *Minkowski space* $(\mathbb{R}^{1,d}, \eta)$ *is obviously a Lorentzian manifold.*

Example 3.2 (Warped product spaces)**.** *Let* $(\mathcal{N}, h)$ *be a connected Riemannian manifold, and* $I \subset \mathbb{R}$ *an open interval. For any smooth positive function* $f : I \to (0, \infty)$*, we can define a metric* $g_{\mu\nu}$ *with line element* $\mathrm{d}s^2 = \mathrm{d}t^2 - f(t)^2 h$ *on* $\mathcal{M} = I \times \mathcal{N}$*. For any two vectors* $X_i = (a_i \partial_t \oplus Y_i) \in T_{(t,p)}(\mathcal{M})$*, with* $Y_i \in T_p\mathcal{N}$ *we have* $g(X_1, X_2) = a_1 a_2 - f(t)^2 h(Y_1, Y_2)$*.*[1] *This type of Lorentzian metrics is called a warped product metric.*

Many familiar Lorentzian manifolds are of the form of this second example. Friedman-Lemaître-Robertson-Walker space-times [Fri22, Fri24, Lem31, Lem33, Rob35, Rob36a, Rob36b, Wal37] are obtained by requiring $(\mathcal{N}, h)$ to be a maximally symmetric Riemannian manifold with a constant curvature metric. This type of metric is of particular relevance when studying cosmological models describing the big bang or the expansion of the universe. A special case is the de Sitter (dS) space-time in global coordinates, where $I = \mathbb{R}$, $\mathcal{N} = S^{n-1}$, with h the canonical metric on the $(n-1)$-sphere with unit radius, and $f(t) = \cosh(t)$.

As a final example, consider the Schwarzschild black hole.

Example 3.3 (Schwarzschild black hole)**.** *For a fixed mass* $M > 0$*, consider the function*

$$h : \mathbb{R}_+ \to \mathbb{R} : r \mapsto 1 - \frac{2M}{r}. \tag{3.2}$$

This function has a pole at $r = 0$ *and a root at* $r = 2M$*. On both patches* $P_I = \{(r,t) \in \mathbb{R}^2 | r > 2M\}$ *and* $P_{II} = \{(r,t) \in \mathbb{R}^2 | 0 < r < 2M\}$ *we define the Lorentzian metric as*

$$g = h(r)\mathrm{d}t^2 - \frac{1}{h(r)}\mathrm{d}r^2. \tag{3.3}$$

The singularity of the metric g at $r = 2M$ might seem problematic, but one can easily show (by going to Kruskal coordinates for example) that this is simply a coordinate singularity. For more details on the Schwarzschild black hole and its rotating and electromagnetically charged cousins we refer the reader to the course GR II.

3.2 Causal structure

Given a Lorentzian manifold with associated metric g, we can associate to each point $p \in \mathcal{M}$ the quadratic form

$$\gamma_p : T_p\mathcal{M} \to \mathbb{R} : \gamma_p(X) = g_p(X, X). \tag{3.4}$$

[1] For any $t \in I$ and $p \in \mathcal{N}$ we identify the tangent space at (t,p) as follows, $T_{(t,p)}\mathcal{M} = T_t I \oplus T_p\mathcal{N}$.

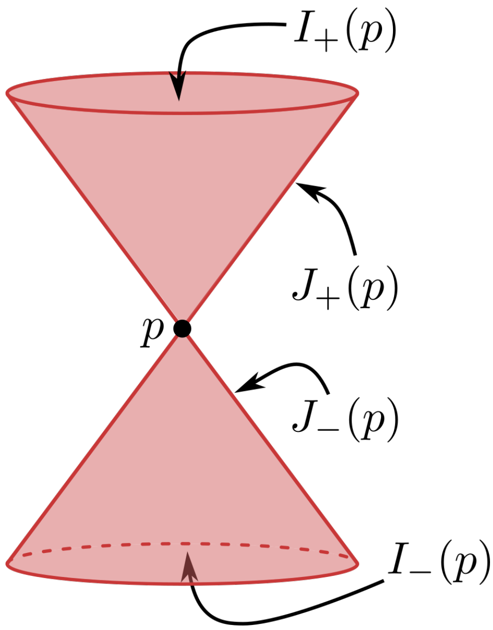


Figure 3.1: The lightcone associated to the point p.

A vector $X \in T_p\mathcal{M}$ is called time-like, light-like or space-like respectively if

$$\begin{cases} \gamma_p(X) > 0\,, & \text{time-like}\,, \\ \gamma_p(X) = 0\,, & \text{light-like}\,, \\ \gamma_p(X) < 0\,, & \text{space-like}\,. \end{cases} \tag{3.5}$$

A vector is called causal if it is time-like or light-like. For $d \geq 1$, the set of time-like vectors consists of two connected components, assigning a time-orientation consists of choosing one of these components which we denote by $I_+(p)$ and call future-directed. The closure $J_+(p) = \overline{I_+(p)}$ consists of the set of future-directed causal vectors. Analogously we call $I_-(p)$ resp. $J_-(p)$ the past directed light-like and causal vectors. The double cone formed as such is called the lightcone at the point p, see Figure 3.1. Similarly, we call a (piecewise) differentiable curve $s \in \mathcal{C}^1(\mathcal{M})$ in $\mathcal{M}$ time-like, light-like or space-like, if all of its tangent vectors are respectively time-like, light-like or space-like.

Having defined these concepts locally at each point p we now want to extend them to global properties of the manifold. A first step in this direction is the following definition.

Definition 3.3 (time-oriented manifold)**.** A Lorentzian manifold $(\mathcal{M}, g)$ is time-orientable if there exists a continuous time-like vector field τ on $\mathcal{M}$. A Lorentzian manifold together with such a vector is called time-oriented.

We will call time-oriented Lorentzian manifolds space-times. It should be noted that the concept of orientability depends only on the topology of $\mathcal{M}$, while the notion of time-orientability depends on the choice of Lorentzian metric. The question of whether a manifold can be equipped with some time-orientable metric is 'topological' indeed, we have the following equivalent statements (see [O'N83]):

- $\mathcal{M}$ admits a smooth non-vanishing vector field.

- $\mathcal{M}$ can be equipped with a smooth Lorentzian metric.
- $\mathcal{M}$ can be equipped with a time-orientable Lorentzian metric.

Example 3.4. *Some (non-)time-orientable manifolds:*

- *The standard Minkowski space is orientable and time-orientable.*
- *The two-sphere is orientable but cannot be equipped with a time-orientable metric. Indeed, famously S^2 does not admit a smooth non-vanishing vector field.*
- *The Mobius strip is not orientable but it can be equipped with Lorentzian metrics that are either time-orientable or not.*
- *The cylinder $\mathbb{R} \times S^1$ is orientable and can be equipped with a metric that is time orientable,*

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}x^2\,. \tag{3.6}$$

It can also be equipped with a metric that is not time orientable,

$$\mathrm{d}s^2 = -\cos\theta \mathrm{d}\theta^2 + 2\sin\theta \mathrm{d}x\mathrm{d}\theta + \cos\theta \mathrm{d}x^2\,. \tag{3.7}$$

For future reference, we define the following causality relations. Let $p, q \in \mathcal{M}$, we have

$$\begin{cases} p \ll q & \leftrightarrow \quad \exists \text{ a future directed timelike curve in } \mathcal{M} \text{ connecting } p \text{ and } q\,, \\ p < q & \leftrightarrow \quad \exists \text{ a future directed causal curve in } \mathcal{M} \text{ connecting } p \text{ and } q\,, \\ p \leq q & \leftrightarrow \quad p < q \text{ or } p = q\,. \end{cases} \tag{3.8}$$

Unless otherwise mentioned, we always consider both space and time orientable manifolds. However, time-orientability is not quite strong enough to rule out all problematic cases. For linear operators to have well-defined unique, causal solutions we need something more.

In general relativity, worldlines of particles are modelled by causal curves. If the space-time is compact, something strange happens. Namely, in every compact space-time $\mathcal{M}$ there exists a closed time-like curve (CTC). When such curves exist, one easily runs into paradoxes as travel into the past is now a clear possibility. Therefore, when dealing with Lorentzian space-times we want to exclude such examples.

Definition 3.4 (Causal manifold)**.** A space-time $(\mathcal{M}, g)$ is causal if it does not contain any closed causal curve.

Definition 3.5 (Strongly causal manifold)**.** A space-time $(\mathcal{M}, g)$ is strongly causal if for any point $p \in \mathcal{M}$ and any neighbourhood U of p, there exists a causally convex neighbourhood V of p, contained in U. A neighbourhood is said to be causally convex if any causal curve with endpoints in V is entirely contained in V.

This property implies that there cannot be time-like curves that pass through V more then once. In other words, it is not possible to return to the same point in space-time by following a time-like curve, i.e. particles travelling slower then light cannot return to the same point in space-time. Strong causality obviously implies causality. For technical reasons we will always assume strong causality.

The wave equations that we consider, such as the wave equation for a scalar field $(\Box + m^2)\phi = 0$, are all hyperbolic (linear) partial differential equations. This means that an equation of order n has a well-posed initial value problem for the first $n-1$ derivatives.[2] More precisely, the Cauchy problem can be locally solved for arbitrary initial data along a non-characteristic hypersurface Σ. In order to have well-defined global solutions on Lorentzian space-times we need some further technical definitions.

Definition 3.6 (Achronal hypersurface)**.** A hypersurface Σ is achronal if no pair of points $p, q \in \Sigma$ can be connected by a time-like curve.

Since solutions propagate along causal curves the data on the hypersurface Σ can only influence a restricted region. We can define the regions

$$\begin{aligned} J_+(\Sigma) &= \left\{ p \in \mathcal{M} \,\middle|\, \exists \text{ a future directed causal curve starting on } \Sigma \text{ to } p \right\} , \\ J_-(\Sigma) &= \left\{ p \in \mathcal{M} \,\middle|\, \exists \text{ a future directed causal curve from } p \text{ ending on } \Sigma \right\} . \end{aligned} \tag{3.9}$$

These sets are sometimes denoted as the future/past domain of influence. Analogously we can define $I_\pm(\Sigma)$ by restricting to the interior of $J_\pm(\Sigma)$.

Definition 3.7 (Domain of dependence)**.** The domain of dependence of a subset Σ is defined as the set of points, $D(\Sigma)$, in $\mathcal{M}$ through which every inextendable causal curve in $\mathcal{M}$ meets Σ, i.e.

$$D(\Sigma) = \{ p \in \mathcal{M} | \text{ every inextendable causal curve passing through } p \text{ intersects } \Sigma \} . \tag{3.10}$$

Analogously we define the future/past domain of dependence $D_\pm(\Sigma)$ as the intersection $D(\Sigma) \cap J_\pm(\Sigma)$.

The domain of dependence is the region on which the initial value problem for wave equations can be proved to be well-posed by various PDE techniques. If p is a point lying on a causal curve that cannot be extended through Σ, then one can imagine waves coming in along that curve that are not determined by the data on Σ and therefore they would violate the uniqueness assumption. The interior of the domain of dependence is sometimes also called the Cauchy development of a set.

Definition 3.8 (Cauchy hypersurface)**.** A subset Σ of a connected space-time $\mathcal{M}$ is a Cauchy hypersurface if each inextendable causal curve in $\mathcal{M}$ meets Σ at exactly one point.

In other words a Cauchy hypersurface is a hypersurface for which $D(\Sigma) = \mathcal{M}$. Any two Cauchy hypersurfaces in $\mathcal{M}$ are homeomorphic. They are topological.

In analogy with the nomenclature for PDEs, we call a space-time globally hyperbolic when the future state of the system is entirely specified by initial conditions. There are several equivalent definitions of global hyperbolicity of which we note two.

Definition 3.9 (globally hyperbolic manifold I)**.** A space-time $\mathcal{M}$ is a globally hyperbolic manifold if it is strongly causal and if for all $p, q \in \mathcal{M}$ the intersection $J_+(p) \cap J_-(q)$ is compact.

[2] We say that the initial value problem on some region U with given data on Σ is well-posed if there exists a unique solution on U with given initial data on Σ.

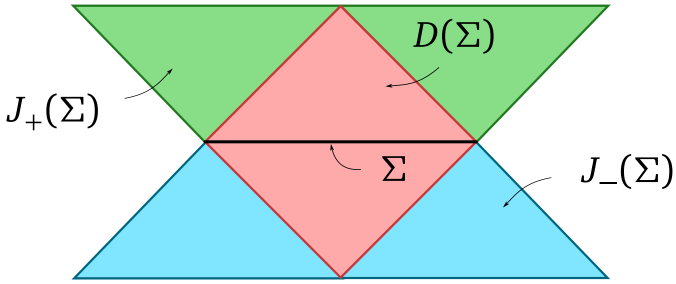


Figure 3.2: The domains of dependence and influence of a set Σ. The green area is the future domain of influence, the blue area the past domain of influence and the red area denotes the full future + past domain of dependence.

Definition 3.10 (globally hyperbolic manifold II)**.** A space-time $\mathcal{M}$ is a globally hyperbolic manifold if it admits a Cauchy surface.

A very useful theorem about globally hyperbolic manifolds goes as follows.

Theorem 3.1. *Let $\mathcal{M}$ be a connected time-oriented Lorentzian manifold. Then the following are equivalent:*

1. *$\mathcal{M}$ is globally hyperbolic.*
2. *$\mathcal{M}$ has a Cauchy hypersurface.*
3. *$\mathcal{M}$ is isometric to $\mathbb{R} \times \Sigma$ with metric $\mathrm{d}s^2 = -\beta \mathrm{d}t^2 + \mathrm{d}s^2_\Sigma(t)$, where $\mathrm{d}s^2_\Sigma(t)$ depends smoothly on t and each $\{t\} \times \Sigma$ is a Cauchy hypersurface in $\mathcal{M}$.*

The proof is rather technical so we do not state it here but the crucial step is proving that *1.* follows from *3.*. The proof can be found in [BS05] using a theorem by Geroch [Ger70]. Once this step has been proven the other implications follow straightforwardly.

Exercise 3.1. *Show that on a globally hyperbolic manifold $\mathcal{M}$ there always exists a smooth function $h : \mathcal{M} \to \mathbb{R}$ whose gradient is past-directed time-like at every point and all of whose level sets are space-like Cauchy hypersurfaces. Such a function is called a Cauchy time function.*

Globally hyperbolic manifolds are very useful since this property ensures that for a large class of operators the Cauchy problems are well-posed, with initial data on a Cauchy hypersurface in appropriate function spaces. In particular, most of the operators we will encounter are generalized d'Alembertians P, of the form

$$P = \sum_{i,j=0}^{d} g^{ij}(x)\partial_{x_i}\partial_{x_j} + \sum_{i=0}^{d} a^i(x)\partial_{x_i} + b(x), \tag{3.11}$$

where the inverse metric g^{ij}, and the functions a^i and b are smooth functions of x. On globally hyperbolic manifolds, one can prove a whole range of global existence and uniqueness theorems for the solutions of the homogeneous equation $P\phi = 0$ as well as for the Green's functions. For a more detailed treatment of linear wave equations on Lorentzian manifolds we refer the reader to the lecture notes [BBB+09].

For these reasons, we will mostly consider globally hyperbolic manifolds in the subsequent chapters. Minkowski space, de Sitter space, the Schwarzschild black hole as well as the FLRW solutions are all globally hyperbolic space-times. However, there are various notable exceptions which will be very interesting to study such as for example Anti-de Sitter (AdS) space. We can overcome the issues accompanying the absence of global hyperbolicity by providing appropriate boundary conditions at conformal infinity. In Chapter 10 we will come back to this example in more detail.

3.3 Conformal infinity and Penrose diagrams

To get a good grip on the global structure, a very useful tool is to study the asymptotics. A neat way to do so is via conformal compactifications. This consist of introducing a conformal boundary which corresponds to infinity in the physical space-time. Various aspects of this procedure were discussed in GRII and we refer the reader to that course for more details and references. Here we restrict ourselves to those aspects relevant in the remainder of this course. In particular we will not go into any detail on the conformal compactifications of black hole space-times.

Definition 3.11 (Conformal transformation)**.** A conformal transformation is a map from a space-time $(\mathcal{M}, g)$ to another space-time $(\widetilde{\mathcal{M}}, \tilde{g})$, such that

$$\tilde{g}_{\mu\nu}(x) = \Omega(x)^2 g_{\mu\nu}(x)\,, \tag{3.12}$$

where Ω is a non-vanishing smooth function of the coordinates, $\Omega(x) \neq 0$ for all $x \in \mathcal{M}$.

One reason why conformal transformations are useful is because they preserve the causal structure of space-time. Indeed, two space-times whose metrics are related by a conformal transformation have the same null geodesics.[3] We can use this fact to our advantage to study the causal structure of space-time by using suitably chosen conformal transformations to bring infinity to a finite coordinate distance allowing us to represent the causal structure on a finite sized diagram called a Penrose diagram.[4] The process of mapping space-time to a compact domain is called conformal compactification.

Definition 3.12. A conformal compactification of a space-time $(\mathcal{M}, g)$ is a manifold $\widetilde{\mathcal{M}}$ with boundary $\mathscr{I} = \widetilde{\mathcal{M}}$ and metric $\tilde{g}$ such that

1. $\tilde{g}$ is smooth on $\widetilde{\mathcal{M}}$.
2. $\mathcal{M}$ is diffeomorphic to the interior of $\widetilde{\mathcal{M}}$.
3. On $\mathcal{M}$ we have that $\tilde{g} = \Omega^2 g$ with Ω a smooth function on $\widetilde{\mathcal{M}}$.
4. In the interior, $\widetilde{\mathcal{M}} \backslash \partial \widetilde{\mathcal{M}} \simeq \mathcal{M}$ we have $\Omega \neq 0$.
5. On the boundary $\partial \widetilde{\mathcal{M}} = \mathscr{I}$, we have $\Omega = 0$, and $d\Omega \neq 0$.

[3] Space-like and time-like geodesics on the other hand are not necessarily preserved under conformal transformations.
[4] Or Carter-Penrose diagram if you ask someone from Cambridge.

The boundary $\mathscr{I}$ [5] is called conformal infinity. In addition to the hypersurface $\mathscr{I}$ the conformally extended space-time might contain loci of higher co-dimension, these have to be considered separately and are denoted by i_0 or $i_\pm$ depending on their (time-like/space-like/null) causal properties.

To illustrate these concepts, let us consider in detail the example of Minkowski space.

Example 3.5 (conformal compactification of Minkowski space-time). *Consider Minkowski space-time in $d+1$ dimensions with line element*

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \sum_{i=1}^{d} \mathrm{d}x_i^2 , \qquad t, x_i \in (-\infty, \infty) . \tag{3.13}$$

In spherical coordinates the metric becomes

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}r^2 - r^2 \mathrm{d}s^2_{S^{d-1}} . \tag{3.14}$$

Defining the light-cone coordinates $u = t - r$ and $v = t + r$ and performing a further coordinate transformation[6]

$$u = \tan \tilde{u} , \qquad v = \tan \tilde{v} , \qquad \tilde{u}, \tilde{v} \in \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) , \tag{3.15}$$

with $\tilde{u} \leq \tilde{v}$, the metric becomes

$$\mathrm{d}s^2 = -\frac{1}{4\cos^2\tilde{u}\cos^2\tilde{v}} \left[4\mathrm{d}\tilde{u}\mathrm{d}\tilde{v} - \sin^2(\tilde{v}-\tilde{u})\mathrm{d}s_{S^{d-1}} \right] . \tag{3.16}$$

We can now use a conformal transformation to remove the prefactor. Since the metric is now regular at the points at infinity we can now compactify the space by including the points $\tilde{u}, \tilde{v} = \pm\frac{\pi}{2}$. The Penrose diagram for Minkowski space is illustrated in Figure 3.3.

Alternatively, changing coordinates $\tilde{u} = \frac{1}{2}(T - \rho)$ and $\tilde{v} = \frac{1}{2}(T + \rho)$ we obtain the following metric for the conformal compactification

$$\widetilde{\mathrm{d}s}^2 = \mathrm{d}T^2 - \mathrm{d}s^2_{S^3} , \tag{3.17}$$

known as the Einstein cylinder metric. Consequently, the rescaling procedure described above maps Minkowski space into a (compact) region of the Einstein cylinder, see Figure 3.3.

Having constructed the conformal compactification of Minkowski space let us discuss the structure of conformal infinity.

- Future and past null infinity are defined as the hypersurfaces,

$$\mathscr{I}^\pm = \left\{ p \in \widetilde{\mathcal{M}} \;\middle|\; 0 < \rho(p) < \pi, T(p) = \pm(\pi - \rho(p)) \right\} . \tag{3.18}$$

  By definition, on this hypersurface we have $\Omega = 0$ and $\mathrm{d}\Omega \neq 0$. Moreover, at this locus we have

$$\tilde{g}(\mathrm{d}\Omega, \mathrm{d}\Omega) = 0, \tag{3.19}$$

  so that $\mathscr{I}$ are null hypersurfaces.

[5] Pronounced as scri as a shorthand for script I.

[6] Note that in order for r to be positive we need to restrict both u, v and $\tilde{u}$, $\tilde{v}$ to the domain $u < v$ and $\tilde{u} < \tilde{v}$.

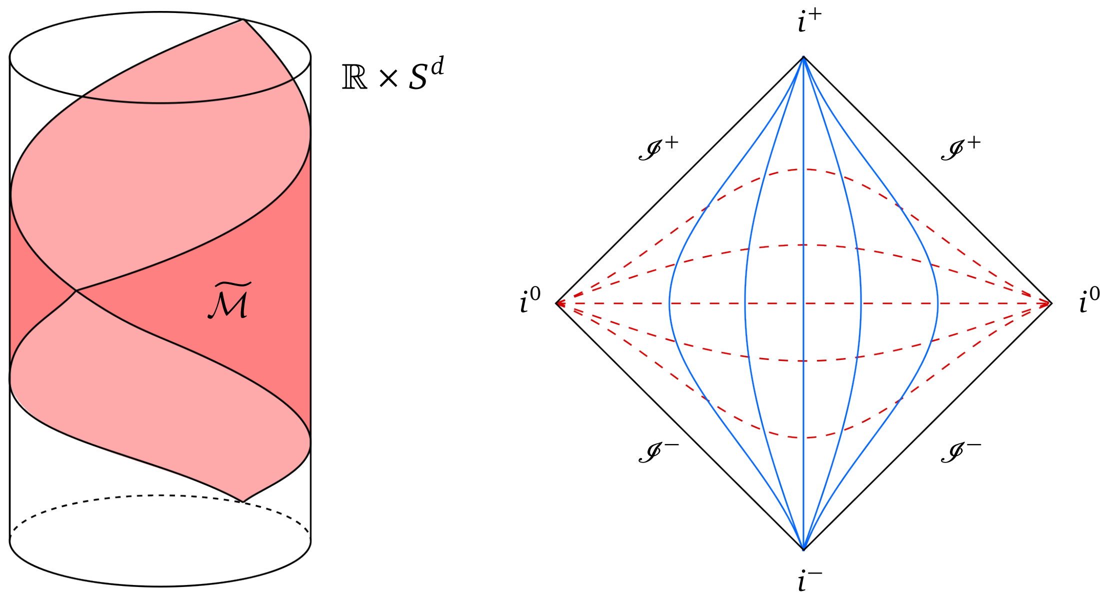


Figure 3.3: The Penrose diagram of Minkowski space. On the left we show the conformal compactification wrapped on the Einstein cylinder. In the Penrose diagram on the right, the time-like geodesics, i.e. lines with constant r, are illustrated as blue lines, while the space-like geodesics, lines with constant t, are illustrated as the dashed red lines.

- Spatial infinity is defined as

$$i^0 = \left\{ p \in \widetilde{\mathcal{M}} \;\middle|\; \rho(p) = \pi \,, T(p) = 0 \right\} . \tag{3.20}$$

  At these points, the radius of the $(d-1)$-sphere vanishes in the usual (d)-sphere degeneration. At this point, both $\Omega = \mathrm{d}\Omega = 0$.

- Future and past time-like infinity are defined as

$$i^\pm = \left\{ p \in \widetilde{\mathcal{M}} \;\middle|\; \rho(p) = 0 \,, T(p) = \pm\pi \right\} . \tag{3.21}$$

  Again, the $(d-1)$-spheres at these points have vanishing radius and $\Omega = \mathrm{d}\Omega = 0$.

The motivations for this nomenclature follows from the analysis of (inextensible) geodesics in the conformally compactified space. Indeed, one can show that space-like geodesics all end and start at i^0, while time-like geodesics start at i^- and end at i_+. Finally, null geodesics start at $\mathscr{I}^-$ and end at $\mathscr{I}^+$.

Next, let us consider the maximally symmetric space-times: Minkowski space, de Sitter space and anti-de Sitter space. These are solutions to the vacuum Einstein equations with resp. vanishing, negative or positive cosmological constant. See Appendix E for more details and a variety of metrics on (A)dS space-times. We leave the explicit construction of the conformal compactification in these cases as an exercise, and restrict ourselves here to a qualitative discussion.

The hypersurface at infinity, $\mathscr{I}$ has a rather different flavour depending on the value of the cosmological constant. In particular it is null for flat space, space-like for de Sitter, and time-like for anti de Sitter.

Moreover, as all of them are conformally flat, they can be mapped to a portion of the Einstein cylinder. de Sitter space occupies a horizontal strip, while Anti de Sitter maps to a vertical strip, in line with the nature of conformal infinity. In Figure 3.4 we show the Penrose diagram for (A)dS space-time.

As a final comment, note that Minkowski space and de Sitter are clearly globally hyperbolic, but that AdS is not. For AdS, we need to present, not just data on an initial $t = const.$ hypersurface, but also data, or at least boundary conditions at time-like infinity. Otherwise, one can imagine incoming radiation from infinity which in turn might cause very problematic instabilities.

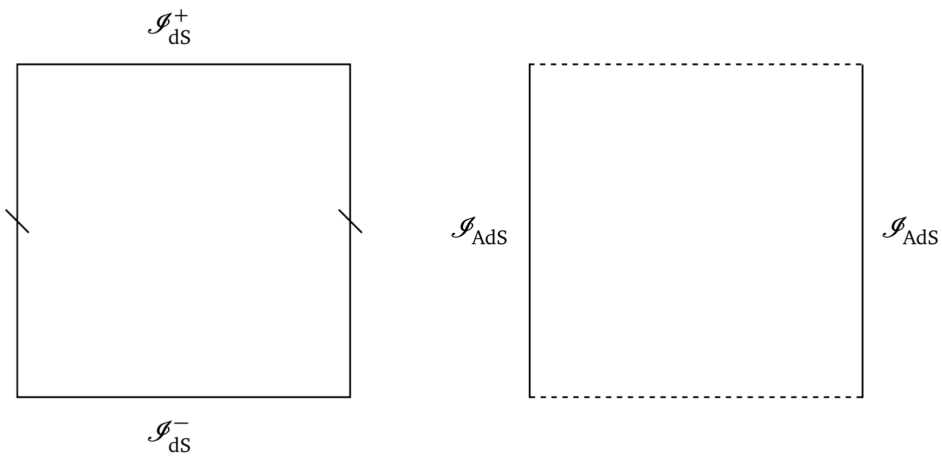


Figure 3.4: The Penrose diagrams for de Sitter (left) and Anti de Sitter (right) space-times. In the de Sitter case, the left and right vertical line can be identified. The topology of $\mathscr{I}^{\pm}_{\text{dS}}$ is $\mathbb{R} \times S^2$, while for AdS conformal infinity is conformal to $\text{Mink}_{1,d-1}$.

For the maximally symmetric spaces there is an elegant alternative construction for their conformal compactification though their embedding in $\mathbb{R}^{2,d+1}$. To see this let us consider the Lorentzian conformal group in $(d+1)$ dimensions, $\text{SO}(2, d+1)$. This group only acts on the compactification, as it interchanges points at finite distance with points at infinity. The conformal group acts on $\mathbb{R}^{d+3}$ by orthogonal transformations preserving the quadratic form

$$X^2 = \eta_{IJ} X^I X^J = s^2 - w^2 + \eta_{\mu\nu} x^\mu x^\nu \,, \tag{3.22}$$

where the coordinates on $\mathbb{R}^{d+3}$ are given by $X^I = (s, w, x^\mu)$ where $I = 0, \ldots, d+2$ and $\mu = 0, \ldots, d$.

To obtain the conformal compactification of the maximally symmetric space-times, we choose a non-zero constant vector $K^I \in \mathbb{R}^{(d+3)}$ and define the metric,

$$\mathrm{d}s^2_{K^2} = \left. \frac{\eta_{IJ} \mathrm{d}X^I \mathrm{d}X^J}{(K \cdot X)^2} \right|_{X^2=0} \tag{3.23}$$

where we defined the product between to vectors $X \cdot Y = X_I Y^I$. Up to $\text{SO}(2, d+1)$ transformations, K^I is distinguished only by its norm K^2 so there are only 3 cases $K^2 = -1, 0, 1$. Since we divided by a quadratic function, the metric is invariant under constant rescalings of the embedding coordinates X^I. Moreover, on $X^2 = 0$ the form $X_I \mathrm{d}X^I = \frac{1}{2} \mathrm{d}X^2$ vanishes so it is easy to see that under $X^I \to f(X) X^I$, the metric $\mathrm{d}s^2_I$ is invariant for any (non-vanishing) function $f(X)$. Thus we can scale X so that $K \cdot X = 1$ on the interior of the conformal compactification. At conformal infinity we find that $K \cdot X = 0$. The isometry group of the metric $\mathrm{d}s^2_{K^2}$ can then be found as the subgroup of $\mathrm{SO}(2, d+1)$ that preserves the vector K^I.

Exercise 3.2. *Show that by taking the vectors:*

$$K = (1, -1, 0, 0, 0, 0)\,, \tag{3.24}$$

$$K = (1, 0, 0, 0, 0, 0)\,, \tag{3.25}$$

$$K = (0, 1, 0, 0, 0, 0)\,, \tag{3.26}$$

the metric (3.23) *reduces resp. to the metric on Minkowksi space, de Sitter space or anti-de Sitter space. Use this to construct the conformal compactifications. Show that the resp. preserved isometry groups are* $\mathrm{SO}(1, d)$*,* $\mathrm{SO}(1, d+1)$ *and* $\mathrm{SO}(2, d)$*.*

Exercise 3.3. *As a final example consider the FLRW backgrounds with spatial sections of constant curvature. The metric is given by*

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2 \mathrm{d}s^2_{d,k}\,, \tag{3.27}$$

where for $k = 0, 1, -1$ *the spatial manifold is respectively flat Euclidean space, the sphere or hyperbolic space.*

Show that these metrics are all conformally flat. Hint: consider the conformal time $\tau = \int_0^t \frac{\mathrm{d}t}{a(t)}$*.*

So far all our examples were conformally flat, allowing us to use a variety of tricks to easily construct their conformal compactification. For non-conformally flat space-times many of our tricks fail making the task of finding the conformal compactification more involved. Conceptually the procedure remains identical, as described in Definition 3.12 but has to be discussed on a case by case basis. However, if M is globally hyperbolic we can see that $\mathscr{I} = \mathscr{I}^+ \cup \mathscr{I}^-$ where future infinity $\mathscr{I}^+$ is to the future of a Cauchy hypersurface and past infinity $\mathscr{I}^-$ to the past.

3.4 Asymptotics and peeling

All the examples discussed above had an important feature in common, namely that they all admit a smooth conformal extension which attaches a conformal boundary to the space-time. A natural question is to what extend this property is shared by more generic manifolds.

Definition 3.13 (Asymptotically simple space-times)**.** A space-time $(\mathcal{M}, g)$ is asymptotically simple if there exists a smooth, oriented, time-oriented causal conformal compactification $(\widetilde{\mathcal{M}}, \tilde{g})$ such that each null geodesic of $(\widetilde{\mathcal{M}}, \tilde{g})$ acquires two distinct endpoints on $\mathscr{I}$.

Note that the completeness requirement in this definition excludes singular space-times such as the Schwarzschild black hole in which there exist null geodesics which do not reach $\mathscr{I}$. Not only those falling into the black hole but also those lying in the photon sphere are incomplete in this sense. Moreover, even without singularities, the fact that a space-time is smooth and geodesically complete does not guarantee that it admits a smooth conformal compactification.

Exercise 3.4. *Consider the Nariai space-time* $\mathcal{M} = \mathbb{R} \times S^1 \times S^2$ *with metric*

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \cosh^2 t\, \mathrm{d}\psi - \mathrm{d}s^2_{S^2}\,. \tag{3.28}$$

The Nariai space-time is both geodesically complete and globally hyperbolic. Show that it does not allow for a smooth conformal extension.

To do so note that under a conformal transformation the squared Weyl tensor transforms as

$$C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma} = \Omega^4 \widetilde{C}_{\mu\nu\rho\sigma}\widetilde{C}^{\mu\nu\rho\sigma}\,. \tag{3.29}$$

Use this fact to show that a smooth conformal extension cannot exist for this space-time.

The requirement of asymptotic simplicity is very natural from a physical point of view and can be thought of as stating that the matter density has to die off quickly enough at infinity such that the asymptotic geometry is purely described by the cosmological constant. More precisely, we only allow conformally invariant matter in the neighbourhood of $\mathscr{I}$.[7] Furthermore, we can prove the following theorem

Theorem 3.2. *Let $(\mathcal{M}, g)$ have conformal compactification $(\widetilde{\mathcal{M}}, \tilde{g})$, and suppose that the space-time asymptotically satisfies the Einstein equations with conformally invariant matter (so that the trace of the stress-energy tensor vanishes) with cosmological constant λ. Then $\mathscr{I}$ is space-like when $\lambda > 0$, time-like for $\lambda < 0$ and null when $\lambda = 0$.*

Proof: From Einstein's equations in $d + 1$ dimensions we immediately obtain

$$R = \frac{2(d+1)}{d-1}\lambda\,, \tag{3.30}$$

in a neighbourhood of $\mathscr{I}$. Define the Schouten tensor as

$$P_{\mu\nu} = -\frac{1}{d-1}\left(R_{\mu\nu} - \frac{1}{2d}Rg_{\mu\nu}\right)\,. \tag{3.31}$$

Near $\mathscr{I}$ we have that $P^{\mu}_{\mu} = -\frac{d+1}{d(d-1)}\lambda$ and note that near $\mathscr{I}$ it transforms as

$$P_{\mu\nu} = \tilde{P}_{\mu\nu} + \Omega^{-1}\tilde{\nabla}_{\mu}\tilde{\nabla}_{\nu}\Omega - \frac{1}{2}\Omega^{-2}\tilde{g}_{\mu\nu}\tilde{\nabla}_{\rho}\Omega\tilde{\nabla}^{\rho}\Omega\,, \tag{3.32}$$

under a conformal rescaling of the metric. Remember that on $\mathscr{I}$ we have $\Omega = 0$ so that,

$$\widetilde{g}^{\mu\nu}N_{\mu}N^{\nu} = \frac{\lambda}{24}\,, \tag{3.33}$$

on $\mathscr{I}$, where we defined the normal vector $N_{\mu} = -\tilde{\nabla}_{\mu}\Omega$. From this the proposition immediately follows.

Asymptotically simple space-times satisfying the conditions in the theorem above are called asymptotically flat, asymptotically AdS or asymptotically dS, depending on the value of the cosmological constant. Moreover, using the results above one can analyse the topology of conformal infinity in each case, the de Sitter and anti de Sitter cases are rather straightforward. For de Sitter one can show that topologically $\mathscr{I} \approx S^{d-1}$, while for anti de Sitter it is a compact time-like hypersurface. The case

[7] Conformally invariant matter comes with a traceless stress tensor and includes massless matter fields such as for example a massless scalar field or massless gauge fields. Therefore this criterion still allows for radiation to reach infinity.

of vanishing cosmological constant on the other hand is harder to analyse. However, one can prove that in any asymptotically simple space-time for which $\mathscr{I}$ is everywhere null, the topology of each component $\mathscr{I}^{\pm}$ is given topologically by $\mathscr{I}^{\pm} \approx \mathbb{R} \times S^2$, where the $\mathbb{R}$ factor can be thought of as the rays generating null infinity.

To finish this chapter we briefly discuss one of the most important results of the theory of asymptotics of the gravitational field, the so-called peeling theorem. The peeling theorem is based on the observation that in asymptotically simple space-times the Weyl tensor must vanish on $\mathscr{I}$ and quantises the allowed decay. A precise statement of the peeling theorem requires the introduction of a variety of new concepts and goes beyond the scope of these lectures. For this reason we only present a simplified sketch and refer the interested reader to the textbooks [PR84, Kro23] for more information.

As already mentioned before, it is reasonable to expect conformally invariant and massless fields to continue smoothly to $\mathscr{I}$ in the conformal compactification. Consider a (conformally coupled) scalar field ϕ solving the (conformally invariant) wave equation. Then $\tilde{\phi} = \phi/\Omega^{\frac{d-1}{2}}$ should be smooth on $\mathscr{I}$ (at least if it is in the domain of dependence of $\mathcal{M}$). In an asymtotically de Sitter space, where $\lambda > 0$, we can deduce that the massless scalar will evolve past $\mathscr{I}$ as if it wasn't there and so $\tilde{\phi}$ will be smooth and generically non-vanishing near $\mathscr{I}$ in the conformally extended space-time. Translating this back to the physical space-time this gives a sharp asymptotic fall-off of the physical scalar ϕ. It is instructive to compare this fall-off in terms of the affine parameter r along an outward going geodesic terminating on $\mathscr{I}$. When $\mathscr{I}$ is null we find

$$\Omega \sim \frac{1}{r} \qquad \rightarrow \qquad \phi \sim \frac{1}{r^{\frac{d-1}{2}}} . \tag{3.34}$$

On the other hand, in the de Sitter case, it is easily seen that $\Omega \sim \exp(-t)$, where t is the proper time along a time-like geodesic ending on $\mathscr{I}^+$.[8] Hence, for massive fields we find an exponential fall-off for the physical fields. In the asymptotically de Sitter case these conclusions can straightforwardly be extended to higher spin/helicity fields.

These considerations can straightforwardly be generalised to study the fall-off of scalars in asymptotically AdS or flat spaces. When $\lambda = 0$ however, the analysis is more involved for higher spin fields as they can scale differently according to whether they are aligned with $\mathscr{I}$ or transverse to it. Carefully doing so for the Weyl tensor results in a proper statement of the peeling theorem in asymptotically flat spaces, see for example [PR84, Kro23] for a careful statement.

Remark. As a final remark, note that in the peeling theorem one is usually focused on asymptotically flat space-times and the decay of gravitational waves at null infinity. In asymptotically AdS space-times, in particular in the context of holography, a more common way of analysing asymptotic expansions proceed through the use of the Fefferman-Graham expansion [FG85]. See Chapter 10 for more details in the context of holography.

---

[8] This can be easily see by noting that $\mathrm{d}t = \frac{\mathrm{d}T}{T}$, where T is the time coordinate on the Einstein cylinder.

Chapter 4

Quantum fields in curved space

Having introduced all the necessary global properties of Lorentzian spacetimes we are well-equipped to move on to quantum fields in curved spacetime. This chapter introduces all the necessary ingredients and discusses general features of quantisation in curved spacetimes. In the next chapter we continue the general discussion and focus on observables quadratic in the fields, such as the stress tensor vacuum expectation value. The following chapters will then be devoted to applying this general framework to particularly interesting examples. The main new feature in curved space will be that we will no longer have a clear grasp of the concept of particle. Associated is the lack of uniqueness in defining the vacuum state.

4.1 Classical fields in curved space-time

To formulate a classical field theory in curved spacetime we need to know how the various fields couple to the background metric. Let us once more emphasize that the metric will not be a dynamic field and we only wish to consider fixed background metrics. This situation is very similar to the charged scalar we considered in Chapter 2.

Analogous to that case, we can couple the theory to a non-trivial background metric simply by changing all partial derivatives into covariant derivatives, $\partial_\mu \to \nabla_\mu$. However, this is not quite enough. In order to guarantee that the action transforms as a scalar under Lorentz transformations we need to simultaneously change the measure $\mathrm{d}^{d+1}x$ to the Lorentz invariant measure $\sqrt{|g|}\mathrm{d}^{d+1}x$. This procedure is in a sense the minimal consistent way to couple the system to gravity and for that reason goes under the name minimal coupling.[1]

Taking for example the free massive scalar field we find the action

$$S = \frac{1}{2}\int_{\mathcal{M}} \mathrm{d}^{d+1}x\,\sqrt{|g|}\left(g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - m^2\phi^2\right). \tag{4.1}$$

This leads to the equation of motion,

$$\left(\Box + m^2\right)\phi = 0\,, \tag{4.2}$$

[1]Moreover, this coupling is consistent with Einstein's principle of equivalence, according to which local gravitational effects are not present in a neighbourhood of the spacetime origin of a locally inertial frame of reference. A similar comment applies to the theory with conformal coupling.

where now $\Box$ represents the d'Alembertian on the curved space,[2]

$$\Box\phi = g^{\mu\nu}\nabla_\mu\nabla_\nu\phi = |g|^{-1/2}\partial_\mu\left[|g|^{1/2}g^{\mu\nu}\partial_\nu\phi\right]. \tag{4.3}$$

This is not the most general way of coupling the scalar field to a background metric. A slightly more general action one can consider is,

$$S = \frac{1}{2}\int_{\mathcal{M}} \mathrm{d}^{d+1}x\sqrt{|g|}\left(g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - m^2\phi^2 - \xi R\phi^2\right), \tag{4.4}$$

where R is the Ricci scalar of the curved manifold $\mathcal{M}$. The term proportional to R disappears in flat space so in this sense it is a generalization of the usual flat space action. The equation of motion after including the curvature coupling is given by

$$\left(\Box + m^2 + \xi R\right)\phi = 0, \tag{4.5}$$

Remark. In principle one can consider more general higher derivative interactions in the Lagrangian. Such terms could be of the form $R^n\phi^2$ but equally well terms like $(\partial^\mu\phi\partial_\mu\phi)$ could be included. In many context such terms are indeed present, but in this course we will mostly ignore them. However, there are various arguments why to a first approximation this is a reasonable thing to do. First of all, under the renormalisation group flow such terms will always be suppressed with respect to the two-derivative couplings. In many cases these terms will be "irrelevant" and can therefore be ignored at low energies. See the course on renormalisation for more details on this first point. Secondly, when including higher degree terms in the Lagrangian a multitude of issues with causality arise which need a careful treatment which goes beyond the scope of this course.

One reason why the curvature coupling $R\phi$ is useful is that for a specific values of ξ its addition to the Lagrangian of a massless scalar makes the action conformally invariant. For this value $\xi = \xi_{\text{conf.}}$ this coupling to the background metric is called conformal coupling.

Exercise 4.1. *Consider a free massless scalar field in $d+1$ dimensions. Show how the action transforms under a conformal transformation of the background metric,*

$$g_{\mu\nu} \to \Omega(x)^2 g_{\mu\nu}, \qquad \phi \to \Omega(x)^\Delta\phi, \tag{4.6}$$

for any function $\Omega(x) \neq 0$.

For which values of ξ and Δ is the action invariant under conformal transformations? In conformally invariant theories Δ is called the scaling dimension of the field ϕ.

Next, let us briefly consider Maxwell theory, i.e. the Abelian gauge theory with gauge field $A_\mu(x)$ and field strength $F_{\mu\nu} = \nabla_{[\mu}A_{\nu]}$. For gauge fields, the differential form language really starts to show it's

[2] Written as a differential form, the kinetic term in the action takes the form $\int \mathrm{d}\phi \wedge \star\mathrm{d}\phi$, where $\star$ is the Hodge star operator. In this language, the d'Alembertian takes the form $\Box\phi = \star(\mathrm{d}\star\mathrm{d}\phi)$.

elegance. In this language, we write the gauge field as a one form

$$A = A_\mu \mathrm{d}x^\mu \in \Omega^1(\mathcal{M})\,. \tag{4.7}$$

In this language the field strength can be written as the two-form $F = \mathrm{d}A \in \Omega^2(\mathcal{M})$. The action for Maxwell theory coupled to a background metric is given by

$$S = \frac{1}{4}\int_{\mathcal{M}} F \wedge \star F = \frac{1}{2}\int_{\mathcal{M}} \sqrt{|g|}\mathrm{d}^{d+1}x F_{\mu\nu}F^{\mu\nu}\,, \tag{4.8}$$

which is invariant under the gauge transformations, $A_\mu \to A_\mu + \partial_\mu f$, or equivalently $A \to A + \mathrm{d}f$. The Bianchi identity for F can be written as $\mathrm{d}F = \nabla_{[\mu}F_{\nu\rho]} = 0$ while the equations of motion are given by

$$\mathrm{d}\star F = 0\,, \qquad \text{or} \qquad \nabla^\mu F_{\mu\nu} = 0\,. \tag{4.9}$$

The gauge invariance noted above adds a redundancy to Maxwell theory. In order to obtain a deterministic equation, one should first fix a gauge. A common choice of gauge in this situation is provided by the Lorenz gauge defined by $\nabla^a A_a = 0$. Upon this gauge fixing, the equations of motion reduce to the wave equation $\Box A_a = 0$. Note however that there is nevertheless still a residual gauge freedom parameterised by $A \to A + \mathrm{d}f$ provided that $\Box f = 0$. Finally, one can also consider spinors in curved spacetime. Since scalars and occasionally vector fields suffice for most of our purposes, we defer a discussion of spinors in curved spacetimes to Appendix **??**.

Exercise 4.2. *Show how the Maxwell action transforms under conformal transformations. Demonstrate that in $d+1=4$ dimensions the action and equations of motion (and Bianchi identities) are conformally invariant.*

Just as in general relativity (and its generalisations such as e.g. Einstein-Maxwell theory) we can define the energy momentum tensor as

$$T^{\mu\nu} = \frac{2}{|g|^{1/2}}\frac{\delta S}{\delta g_{\mu\nu}}\,, \tag{4.10}$$

which in general relativity this provides a source to the Einstein equation. In this course we consider quantum fields in fixed background metric and do not consider fluctuations of the metric. One can think of this as solving Einsteins equations with a source given by the vacuum expectation value of the stress tensor $\langle T_{\mu\nu}\rangle$. In a second step of our semi-classical treatment we can then consider fluctuations of the dynamical fields around this background metric. Note that up to possible improvement terms this reduces to the definitions in flat space presented in Chapter 2.

Conformal invariance of a theory is manifested in the stress tensor as the property $T^\mu_\mu = 0$. This is a general consequence of the invariance of the action under variations of the form $\delta g_{\mu\nu} = \sigma g_{\mu\nu}$.

Exercise 4.3. *Show that the stress tensor of a conformally invariant theory is necessarily traceless.*

Exercise 4.4. *Compute the stress tensor for a minimal and conformally coupled scalar as well as Maxwell theory in curved space.*

Check that for the latter two the stress tensor is indeed traceless.

4.2 Canonical quantisation in curved spacetime

To avoid subtleties, in this chapter we always assume the spacetime to be globally hyperbolic with a foliation by Cauchy hypersurfaces Σ_t where $t \in \mathbb{R}$ is a time-like coordinate. Having discussed how various fields couple to a background metric we repeat and modify where necessary the same steps as in Chapter 2 to quantise the theory.

Concepts such as the phase space, the associated symplectic form and the commutation relations generalise straightforwardly to curved space. The phase space for a scalar field on a spacetime $\mathcal{M}$ can be defined as

$$V_\phi(\mathcal{M}) = \left\{\phi \in C^\infty(\mathcal{M}) \,\middle|\, (\Box + m^2 + \xi R)\phi = 0\right\} = \left\{(\phi, \dot{\phi}) \in C^\infty(\Sigma_t) \times C^\infty(\Sigma_t)\right\} \tag{4.11}$$

where we allow for an arbitrary coupling to the Ricci scalar. Note that here we defined the initial data on the Cauchy hypersurface at time t and defined $\dot{\phi} = \nabla_n \phi$ with n^μ a unit normal vector to the hypersurface at time t. Similarly, one can define the phase space for fermions or gauge fields entirely analogous to the flat space case. The symplectic form is given by

$$\Omega(\phi_1, \phi_2) = \int_\Sigma \phi_1 \star \mathrm{d}\phi_2 - \phi_2 \star \mathrm{d}\phi_1 = \int_\Sigma (\phi_1 \dot{\phi}_2 - \phi_2 \dot{\phi}_1) \mathrm{d}\Sigma \tag{4.12}$$

where $\mathrm{d}\Sigma$ is the volume form on the hypersurface Σ. For more details on hypersurfaces see Appendix C.

Remark. Note that for the complex scalar field we can write the inner product more invariantly as

$$\langle \phi_1, \phi_2 \rangle = \int_\Sigma \mathrm{d}\Sigma^\mu J_\mu \,, \tag{4.13}$$

where $\mathrm{d}\Sigma^\mu = \mathrm{d}\Sigma n^\mu$ and J_μ is the conserved current for the U(1) global symmetry,

$$J_\mu = \mathrm{i}\left(\phi_1^* \nabla_\mu \phi_2 - \phi_2^* \nabla_\mu \phi_1\right) . \tag{4.14}$$

In order to proceed with the quantisation we need a Hamiltonian description of the theory, and therefore a preferred time coordinate. Assuming there is a splitting (t, x_m) we will quantise the theory using the hypersurfaces Σ_t defined by $t =$ constant. Let us be more explicit and write the metric as

$$\mathrm{d}s^2 = g_{00}\mathrm{d}t^2 + 2g_{0m}\mathrm{d}t\mathrm{d}x^m - h_{mn}\mathrm{d}x^m\mathrm{d}x^n \,, \tag{4.15}$$

where h_{mn} is the induced metric on the Cauchy hypersurface Σ. In terms of these coordinates, we can write the momentum conjugate to ϕ as

$$\pi = \frac{\delta \mathcal{L}}{\delta(\partial_t \phi)} = |g|^{1/2} g^{0\mu} \nabla_\mu \phi = |h|^{1/2} n^\mu \nabla_\mu \phi \,. \tag{4.16}$$

Exercise 4.5. *Prove the second equality in Equation* (4.16)*. The easiest way to do so is by explicit computation.*

Analogous to flat space we can now quantise the theory by promoting all fields to operators on a Hilbert space and imposing the canonical commutation relations

$$\begin{aligned} \left[\phi(t,\mathbf{x}),\phi(t,\mathbf{x}')\right] &= 0\,, \\ \left[\pi(t,\mathbf{x}),\pi(t,\mathbf{x}')\right] &= 0\,, \\ \left[\phi(t,\mathbf{x}),\pi(t,\mathbf{x}')\right] &= \mathrm{i}\,\delta(\mathbf{x}-\mathbf{x}')\,. \end{aligned} \tag{4.17}$$

where now the δ-function is normalised against the volume form on the hypersurface,

$$f(\mathbf{x}') = \int_\Sigma \delta(\mathbf{x}-\mathbf{x}')f(\mathbf{x}')\mathrm{d}\Sigma\,. \tag{4.18}$$

Similarly, this Hilbert space now comes with a generalised inner product,

$$\langle\phi_1,\phi_2\rangle = \mathrm{i}\,\Omega(\phi_1,\phi_2)\,. \tag{4.19}$$

Similar to flat space it follows immediately that this inner product is independent of the choice of Cauchy hypersurface in the given foliation. For this reason it was justified not to specify the precise time t in the above.

To further develop the quantum field theory we again want to relate the above phase space to a one-particle Hilbert space $\mathcal{H}$. However, in curved space there is no canonical analogue of the Fourier transform so this step will need some more care.

Let us for the moment consider a complex scalar field ϕ. Given any solution $f \in V_\phi(\mathcal{M})$ to the Klein Gordon equation we can associate to it an annihilation operator,

$$a\,:\,V_\phi(\mathcal{M})\to\mathcal{H}\,:\,f\mapsto a(f)=\langle f,\phi\rangle\,. \tag{4.20}$$

Since ϕ is Hermitian, we have that the associated creation operator can be defined as

$$a^\dagger(f) = -a(f^*)\,. \tag{4.21}$$

More explicitly, we can write the annihilation operator (and similarly for the creation operator) in terms of ϕ and its canonical momentum π as

$$a(f) = \mathrm{i}\int_\Sigma \mathrm{d}^d x\,|h|^{1/2}\,n^\mu\,(f^*\partial_\mu\phi - \partial_\mu f^*\phi) = \mathrm{i}\int_\Sigma \mathrm{d}^d x\left(f^*\pi - |h|^{1/2}n^\mu\partial_\mu f^*\phi\right)\,. \tag{4.22}$$

From the canonical commutation relations introduced in Equation (4.17), we immediately find

$$\begin{aligned}\left[a(f),a^\dagger(g)\right]=&\int \mathrm{d}^d\mathbf{x}\int \mathrm{d}^d\mathbf{y}\left[(f^*\pi-|h|^{1/2}n^\mu\partial_\mu f^*\phi)(t,\mathbf{x})(g\pi-|h|^{1/2}n^\mu\partial_\mu\phi)(t,\mathbf{y})\right]\\ =&\mathrm{i}\int \mathrm{d}^d\mathbf{x}|h|^{1/2}n^\mu\left(f^*(t,\mathbf{x})\partial_\mu g(t,\mathbf{x})-\partial_\mu f^*(t,\mathbf{x})g(t,\mathbf{x})\right)\\ =&\langle f,g\rangle\ .\end{aligned} \tag{4.23}$$

Similarly, one can easily show that

$$[a(f),a(g)]=-\langle f,g^*\rangle\ ,\qquad \left[a^\dagger(f),a^\dagger(g)\right]=-\langle f^*,g\rangle\ . \tag{4.24}$$

With this map a at hand we have provided an appropriate procedure to quantise the scalar field and construct the one-particle Hilbert space. Similar to the flat space case, one can then proceed to find the full Hilbert space in the Fock representation by consecutively applying creation operators to the vacuum. However, this is not quite enough. In order to have a complete understanding of this Hilbert space we need to find a complete orthonormal basis of the phase space, analogous to the plane waves in flat space. I.e. we have to find a complete set of functions u_i, solving the wave equation and satisfying,

$$\left\langle u_i,u_j\right\rangle=\delta_{ij}\,,\qquad \left\langle u_i^*,u_j\right\rangle=0\,,\qquad \left\langle u_i^*,u_j^*\right\rangle=-\delta_{ij}\,. \tag{4.25}$$

Having found such a basis, we can expand the field ϕ as follows,

$$\phi(x)=\sum_i\left(a_iu_i+a_i^\dagger u_i^*\right)\,, \tag{4.26}$$

where the quantum annihilation and creation operators, a_i and $a_i^\dagger$, satisfy the standard commutation relations,

$$\left[a_i,a_j^\dagger\right]=\delta_{ij}\,. \tag{4.27}$$

A very important point is that on curved space, there is no natural way to perform this mode decomposition. This property is key to many odd strange but interesting phenomena in the theory of quantum field theory on curved space and will lie at the origin of many of the observations that will follow.

A special scenario where there exists a "natural" choice of decomposition is when the spacetime admits a globally well-defined, non-vanishing Killing vector $K=\partial_t$, which is irrotational, i.e. when the spacetime is globally static. In this case one can write the metric as

$$\mathrm{d}s^2(t,\mathbf{x})=g_{00}(\mathbf{x})\mathrm{d}t^2-h_{mn}(\mathbf{x})\mathrm{d}x^m\mathrm{d}x^n\,, \tag{4.28}$$

with h and g_{00} are respectively the metric and a positive function on the transverse (space-like) manifold Σ, both independent of t. In this case the quantisation follows very closely in the steps of quantisation in flat space. We can separate variables and analogous to flat space define the positive modes V_ϕ^+ as the set of functions P_n of the form

$$P_n(x) = \frac{\chi_n(\mathbf{x})}{\sqrt{2\omega_n}} \mathrm{e}^{-\mathrm{i}\omega_n t} \,. \tag{4.29}$$

where the functions $\chi_n(\mathbf{x})$ collectively give a basis of functions on the manifold Σ and satisfy the reduced wave equation,

$$\left(\Delta + \omega_n^2\right) \chi_n = 0 \,, \tag{4.30}$$

for some second order operator Δ on Σ. The functions χ_n should satisfy the completeness relation,

$$\sum_n \bar{\chi}_n(x) \chi_n(x') = \delta^{(d)}(x, x') \,, \tag{4.31}$$

where the δ-function is understood to be normalised against the measure $\mathrm{d}^d x \sqrt{g_{00} h}$. Hence, it follows that the inner product reduces to

$$\langle P_m, P_n \rangle = \int \mathrm{d}^d x \sqrt{g_{00} h} \bar{\chi}_m \chi_n = \delta_{mn} \,. \tag{4.32}$$

Similarly, one can introduce negative frequency modes $N_n \in V_\phi^-$ and complete the quantisation exactly like in flat space. However, this limited approach does not generalise to generic curved spacetimes and moreover, even though the Schwarzschild black hole is static we will see that this approach is insufficient to fully understand quantum field theory on a black hole background.

Remark. In the above we silently assumed that Σ is compact such that the labels n take value in a countably infinite set. More generally, on a non-compact spatial slicing such as $\mathbb{R}^3$, n, m must be replaced by continuous variables such as $\mathbf{k}, \mathbf{k}'$ for flat spatial slices. Similarly, the Kronecker delta, δ_{mn}, should then be replaced by $\delta^3(\mathbf{k} - \mathbf{k}')$ and $\sum_n$ by $\int_{\mathbb{R}^3} d^3\mathbf{k}/\omega_{\mathbf{k}}$. For the more formal development we will mostly stick to the discrete modes and introduce continuous alternatives when needed.

Another example of a natural choice of vacuum arises for conformally invariant theories in conformally flat spacetimes. Consider a conformally flat metric, i.e. $g_{\mu\nu}(x) = \Omega^2(x)\eta_{\mu\nu}$. A solution to Klein-Gordon equation for a conformally coupled massless scalar can then be obtained from the Minkowski solution $u_{\mathbf{k}}^M(x)$ simply by using the conformal transformation of the scalar field. Therefore, we have a natural choice of solutions of the wave equation given by

$$u_{\mathbf{k}}(x) = \Omega(x)^{\frac{1-d}{2}} u_{\mathbf{k}}^M(x) \,. \tag{4.33}$$

The vacuum associated to this choice of modes is called the conformal vacuum. In this case the analysis of the two-point functions becomes particularly easy, as one can obtain the Wightman functions simply as,

$$G_\pm(x, x') = \Omega(x)^{\frac{1-d}{2}} G_\pm^M(x, x') \Omega(x')^{\frac{1-d}{2}} \,, \tag{4.34}$$

where $G_\pm^M(x, x')$ is the Wightman function in Minkowski space introduced in Chapter 2. The Wightman functions transform as bi-scalars under conformal transformations.

4.3 Quantisation in generic curved spacetimes

In general there does not exist such a natural choice of orthonormal basis. For example, consider a spacetime that is asymptotically static both in the future and past region, respectively $\mathcal{M}_+$ and $\mathcal{M}_-$. In both asymptotic regions one can write down a metric in the form (4.28) and define the natural basis $P_n^\pm$ and $N_n^\pm$ resp. in the past and future region. This will lead to two notions of positive and negative modes and in general we do not expect these to agree.

We can quantise the theory in the two regions as defined above. In order to construct the transition between them we can impose that the quantum field ϕ is the same in both regions, i.e.

$$\phi(x) = \sum_n a_n^- P_n^- + a_n^{-\dagger} N_n^- = \sum_n a_n^+ P_n^+ + a_n^{+\dagger} N_n^+ , . \tag{4.35}$$

The main question we'll try to answer in this section is how to identify the Fock spaces $\mathcal{F}^+$ and $\mathcal{F}^-$ of the different asymptotic regions. Similarly, in spacetimes with no asymptotically static regions, we have to learn how to identify the Fock spaces at different times.

To do so, consider two different orthonormal bases of the phase space $\{u_n\}_{n\in\mathbb{N}}$ and $\{v_n\}_{n\in\mathbb{N}}$ such that

$$\phi(x) = \sum_n a_n u_n + a_n^\dagger u_n^* = \sum_n b_n v_n + b_n^\dagger v_n^* , , \tag{4.36}$$

where a_n, $a_n^\dagger$ and b_n, $b_n^\dagger$ are the creation and annihilation operators defined with respect to either the u basis or v basis. Since both sets of functions are independent bases of the $\mathbb{C}\otimes V_\phi$ we can relate them through linear maps

$$\begin{pmatrix} v_n \\ v_n^* \end{pmatrix} = \sum_m \begin{pmatrix} \alpha_{nm} & \beta_{nm} \\ \beta_{nm}^* & \alpha_{nm}^* \end{pmatrix} \begin{pmatrix} u_m \\ u_m^\dagger \end{pmatrix} = \sum_m S_{nm} \begin{pmatrix} u_m \\ u_m^\dagger \end{pmatrix} \tag{4.37}$$

where S is the classical S-matrix whose entries are called the Bogoliubov coefficients,

$$\alpha_{mn} = \langle u_n, v_m \rangle , \qquad \beta_{mn} = -\left\langle u_n^*, v_m \right\rangle , \tag{4.38}$$

and analogous for their complex conjugates. Since the u and v basis were both assumed to be orthonormal, the Bogoliubov coefficients have to satisfy the normalisation relation,

$$\alpha\alpha^\dagger - \beta\beta^\dagger = 1 , \qquad \alpha\beta^\dagger - \beta\alpha^\dagger = 0 . \tag{4.39}$$

Going back to (4.36) we find that the (transposed) Bogoliubov coefficients similarly encode the relation between the a and b creation and annihilation operators,

$$\begin{pmatrix} b_n \\ b_n^\dagger \end{pmatrix} = \sum_m \begin{pmatrix} \alpha_{nm}^* & -\beta_{nm}^* \\ -\beta_{nm} & \alpha_{nm} \end{pmatrix} \begin{pmatrix} a_m \\ a_m^\dagger \end{pmatrix} \tag{4.40}$$

Exercise 4.6. *Starting from (4.36) and (4.37), show that the Bogoliubov coefficients encode the relation (4.40) between the a and b operators.*

Exercise 4.7. *Show that in order for the commutation relations of the b operators to be properly* *normalised, we have to impose the condition* (4.39).

Define the vacua with respect to the u or v expansion, $|0\rangle_a$ and $|0\rangle_b$ in the usual way as the state annihilated by all annihilation operators,

$$a_n\,|0\rangle_a = 0\,, \qquad b_n\,|0\rangle_b = 0\,, \forall n \geq 0\,. \tag{4.41}$$

When $\beta = 0$ the two vacua are equivalent and can be identified up to possibly a phase. However, if $\beta \neq 0$, the b vacuum will contain a particles and vice versa! To see this more explicitly, define the number operators,

$$N_n^{(a)} = a_n^\dagger a_n\,, \qquad N^{(a)} = \sum_n N_n^{(a)}\,, \tag{4.42}$$

and compute their expectation values in the b vacuum,

$$\left\langle N_n^{(a)} \right\rangle_b = \sum_m |\beta_{mn}|^2\,. \tag{4.43}$$

Hence, we see that the changing gravitational field creates particles (in pairs). In a sensible physical systems the number of quanta in any vacuum should be finite so we require that the Bogoliubov transformation is such that $\left\langle N_n^{(a)} \right\rangle_b < \infty$ for all n. Indeed, without this condition various problems regarding the convergence of all the expressions in this section arise.

To find an explicit relation between the a and b vacuum we would like to extend the map S from the classical phase space to the Fock space, $\mathcal{S} : \mathcal{F}_a \to \mathcal{F}_b$. As a first step, we can construct the b vacuum using the Bogoliubov coefficients. After some algebra, one can show that (up to phase) we have

$$|0\rangle_b = \frac{1}{|\det \alpha|^2} \exp\left\{ \sum_{m,n} M_{mn} a_m^\dagger a_n^\dagger \right\} |0\rangle_a\,, \qquad M_{mn} = (\alpha^{*-1}\beta^*)_{mn}\,, \tag{4.44}$$

Such states are sometimes called squeezed states, in analogy with the nomenclature for the ground state for an oscillator with a different frequency. The full quantum evolution operator translating between the two Fock spaces then takes the form

$$\mathcal{S} = \frac{1}{|\det \alpha|^2} \exp\left\{ \sum_{m,n} M_{mn} a_m^\dagger a_n^\dagger - M_{mn}^* a_m a_n \right\}. \tag{4.45}$$

This operator is now unitary as the exponent is skew Hermitian so it preserves norms.

Exercise 4.8. *Show that the a and b vacua are related as in Equation* (4.44). *Hint: use the Baker-Campbell-Hausdorff formula to expand expressions of the form as* $\mathrm{e}^{-F} a_n \mathrm{e}^{F}$.

4.4 Cosmological particle creation

To illustrate the above let us consider the following cosmological FLRW model with a spatially flat isotropically changing metric,

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2 \sum_{m=1}^{d} \mathrm{d}x_m^2\,, \tag{4.46}$$

where the cosmological scale factor $a(t)$ is an arbitrary non-vanishing function of time. We take the cosmological scale factor to asymptotically approach the values[3]

$$a(t) = \begin{cases} a_1 & \text{as} \quad t \to -\infty \\ a_2 & \text{as} \quad t \to \infty \end{cases} . \tag{4.47}$$

To analyse this setup it is convenient to change coordinates to conformal time

$$\eta(t) = \int_{t_0}^{t} \frac{\mathrm{d}t'}{a(t')} , \tag{4.48}$$

such that the metric becomes

$$\mathrm{d}s^2 = C(\eta)\left(\mathrm{d}\eta^2 - \sum_{m=1}^{d} \mathrm{d}x_m^2\right) , \tag{4.49}$$

where $C(\eta) = a^2(\eta)$. The d'Alembertian for this metric can be written as

$$\Box\phi = C(\eta)^{-\frac{d+1}{2}} \partial_\eta \left(C(\eta)^{\frac{d-1}{2}} \partial_\eta \phi\right) - C(\eta)^{-1}\nabla^2\phi . \tag{4.50}$$

The spatial translation symmetry allows the spatial dependence to be separated from the time dependence so that we can consider solutions to the wave equation of the form,

$$u_{\mathbf{k}}(\eta, \mathbf{x}) = \frac{\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}}{(2\pi)^{\frac{d}{2}}} C(\eta)^{\frac{1-d}{4}} \chi_{\mathbf{k}}(\eta) . \tag{4.51}$$

Direct computation then results in

$$\begin{aligned} \Box u_{\mathbf{k}} =& \frac{C(\eta)^{-\frac{3+d}{4}}}{(2\pi)^{\frac{d}{2}}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \left[\chi''_{\mathbf{k}}(\eta) + \mathbf{k}^2\chi_{\mathbf{k}}(\eta) + \frac{1-d}{2}\left(\frac{C''(\eta)}{C(\eta)} + \frac{d-5}{4}\left(\frac{C'(\eta)}{C(\eta)}\right)^2\right)\chi_{\mathbf{k}}(\eta)\right] \\ =& \frac{C(\eta)^{-\frac{3+d}{4}}}{(2\pi)^{\frac{d}{2}}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \left[\chi''_{\mathbf{k}}(\eta) + \mathbf{k}^2\chi_{\mathbf{k}}(\eta) - \xi(d)C(\eta)R\chi_{\mathbf{k}}(\eta)\right] . \end{aligned} \tag{4.52}$$

where R is the Ricci scalar of the FLRW metric and $\xi(d) = \frac{1}{4}\frac{1-d}{d}$. The above can also be obtained by noting that the FLRW metric is conformally flat with conformal factor $\Omega^2 = C(\eta)$ and performing a conformal transformation. Putting the above together, we find that $u_{\mathbf{k}}$ solves the Klein-Gordon equation if $\chi_{\mathbf{k}}$ solves

$$\chi''_{\mathbf{k}} + \omega^2(\eta)\chi_k = 0 , \tag{4.53}$$

where

$$\omega^2(\eta) = k^2 + m^2 C(\eta) + C(\eta)R(\eta)\left(\xi - \xi(d)\right) . \tag{4.54}$$

The induced metric on the hypersurface at constant η is given by

$$h_{mn}\mathrm{d}x^m\mathrm{d}x^n = C(\eta)\delta_{mn}\mathrm{d}x^m\mathrm{d}x^n , \tag{4.55}$$

[3]As usual we are not very careful with the appropriate fall-off and smoothness conditions but assume the function $a(t)$ is sufficiently well-behaved.

so that the inner product between two solutions to functions is given by

$$\begin{aligned}\langle u_{\mathbf{k}}, u_{\mathbf{k}'}\rangle =& \mathrm{i}\int \mathrm{d}^d\mathbf{x}\, C(\eta)^{\frac{d-1}{2}}\left(u^*_{\mathbf{k}}\partial_\eta u_{\mathbf{k}'} - \partial_\eta u^*_{\mathbf{k}} u_{\mathbf{k}'}\right)\\ =& \mathrm{i}\delta^{(d)}(\mathbf{k}-\mathbf{k}')\left(\chi^*_{\mathbf{k}}\partial_\eta \chi_{\mathbf{k}'} - \partial_\eta \chi^*_{\mathbf{k}}\chi_{\mathbf{k}'}\right).\end{aligned} \tag{4.56}$$

Hence we find that the modes are properly normalised if

$$\mathrm{i}\left(\chi^*_{\mathbf{k}}\partial_\eta \chi_{\mathbf{k}'} - \partial_\eta \chi^*_{\mathbf{k}}\chi_{\mathbf{k}'}\right) = 1\,. \tag{4.57}$$

This is a normalisation condition on the Wronskian of the solutions. The fact that the Wronskian is independent of η is a manifestation of the fact that the inner product does not depend on the particular hypersurface we use.

Going back to the problem at hand, let us consider the special case when the function $C(\eta)$ approaches a constant value in the past and future. In this case the asymptotic past and future are Minkowski spacetimes and the η-dependent frequencies have limiting values

$$\omega(\eta) = \begin{cases} \omega_{\mathrm{in}} & \text{as} \quad \eta \to -\infty \\ \omega_{\mathrm{out}} & \text{as} \quad \eta \to \infty \end{cases}. \tag{4.58}$$

As discussed above, we have two natural vacua, one in the asymptotic past and one in the future. These vacua are defined by considering bases of solution $u^{\mathrm{in/out}}_{\mathbf{k}}$ satisfying the following asymptotic conditions,

$$\begin{aligned}\chi^{\mathrm{in}}_{\mathbf{k}}(\eta) &\to \frac{1}{\sqrt{2\omega_{\mathrm{in}}}}\exp(-\mathrm{i}\omega_{\mathrm{in}}\eta)\,, &&\text{as} \quad \eta\to -\infty\,,\\ \chi^{\mathrm{out}}_{\mathbf{k}}(\eta) &\to \frac{1}{\sqrt{2\omega_{\mathrm{out}}}}\exp(-\mathrm{i}\omega_{\mathrm{out}}\eta)\,, &&\text{as} \quad \eta\to \infty\,.\end{aligned} \tag{4.59}$$

These modes then define annihilation and creation operators $a^{\mathrm{in/out}}_{\mathbf{k}}$ and $\left(a^{\mathrm{in/out}}_{\mathbf{k}}\right)^\dagger$ and the corresponding vacua are annihilated by

$$a^{\mathrm{in/out}}_{\mathbf{k}}\left|0_{\mathrm{in/out}}\right\rangle = 0\,. \tag{4.60}$$

To relate these vacua, we have to compute the Bogoliubov coefficients (4.38). Due to the spatial homogeneity, we immediately see that $\alpha_{\mathbf{k},\mathbf{k}'} \propto \delta(\mathbf{k}-\mathbf{k}')$ and similarly, $\beta_{\mathbf{k},\mathbf{k}'} \propto \delta(\mathbf{k}+\mathbf{k}')$. Hence we have

$$\alpha_{\mathbf{k},\mathbf{k}'} = \alpha_{\mathbf{k}}\delta(\mathbf{k}-\mathbf{k}')\,, \qquad \beta_{\mathbf{k},\mathbf{k}'} = \beta_{\mathbf{k}}\delta(\mathbf{k}+\mathbf{k}')\,. \tag{4.61}$$

Note that this means that the Bogoliubov coefficients only mix modes with wave vectors $\mathbf{k}$ and $-\mathbf{k}$ and in addition they only depend on the magnitude of the wave vector due to rotational symmetry. The Bogoliubov transformation then takes the form

$$u^{\mathrm{in}}_{\mathbf{k}} = \alpha_{\mathbf{k}}u^{\mathrm{out}}_{\mathbf{k}} + \beta_{\mathbf{k}}u^{*\mathrm{out}}_{-\mathbf{k}} \tag{4.62}$$

The first condition in (4.39) is automatically satisfied, while the second gives

$$|\alpha_{\mathbf{k}}|^2 - |\beta_{\mathbf{k}}|^2 = 1 \tag{4.63}$$

The condition that the vacuum in the in region contains only a finite number of modes in the out Hilbert space becomes,

$$\int \mathrm{d}^d \mathbf{k} |\beta_{\mathbf{k}}|^2 < \infty \,. \tag{4.64}$$

Remark. Note that since FLRW metric is conformally flat, in the case of a massless, conformally coupled scalar we can always construct a conformal vacuum. Indeed, in this case we have

$$\omega_k^2(\eta) = k^2 \tag{4.65}$$

and the solution simplifies drastically to

$$\chi_{\mathbf{k}}(\eta) = \frac{1}{\sqrt{2\omega_k}} \mathrm{e}^{-ik\eta} \,. \tag{4.66}$$

In this case the function $u_{\mathbf{k}}$ can simply be obtained as a conformal transformation of the plane waves in standard Minkowski space.

Exercise 4.9 (An exactly solvable model). *To get a better feeling for these types of properties, it is interesting to work out some non-trivial models where the equations for $\chi_{\mathbf{k}}$ can be solved explicitly. One such model is given by the two dimensional spacetime with $\xi = 0$ and $C(\eta) = A + B\tanh(\rho\eta)$ studied in [BD77].*

Verify that the equations of motion for χ_k can be solved by the following two sets of functions,

$$\chi_{\mathbf{k}}^{\mathrm{in}}(\eta) = \frac{1}{\sqrt{2\omega_{\mathrm{in}}}} \exp\left[-\mathrm{i}\omega_+ \eta - \frac{\mathrm{i}\omega_-}{\rho} \log(2\cosh\rho\eta)\right] \tag{4.67}$$

$$\times\, {}_2F_1\left(\frac{\rho + \mathrm{i}\omega_-}{\rho}, \frac{\mathrm{i}\omega_-}{\rho}, \frac{\rho - \mathrm{i}\omega_{\mathrm{in}}}{\rho}, \frac{1}{2}(1 + \tanh\rho\eta)\right), \tag{4.68}$$

$$\chi_{\mathbf{k}}^{\mathrm{out}}(\eta) = \frac{1}{\sqrt{2\omega_{\mathrm{out}}}} \exp\left[-\mathrm{i}\omega_+ \eta - \frac{\mathrm{i}\omega_-}{\rho} \log(2\cosh\rho\eta)\right] \tag{4.69}$$

$$\times\, {}_2F_1\left(1 + \frac{\mathrm{i}\omega_-}{\rho}, \frac{\mathrm{i}\omega_-}{\rho}, 1 - \frac{\mathrm{i}\omega_{\mathrm{out}}}{\rho}, \frac{1}{2}(1 + \tanh\rho\eta)\right), \tag{4.70}$$

where $\omega_\pm$ are defined as $\omega_\pm = \frac{1}{2}(\omega_{\mathrm{out}} \pm \omega_{\mathrm{in}})$.

Show that these functions have the appropriate plane wave limit as $\eta \to \pm\infty$ and show that the squared Bogoliubov coefficients are given by

$$|\alpha_{\mathbf{k}}|^2 = \frac{\sinh^2\left(\frac{\pi\omega_+}{\rho}\right)}{\sinh\left(\frac{\pi\omega_{\mathrm{in}}}{\rho}\right)\sinh\left(\frac{\pi\omega_{\mathrm{out}}}{\rho}\right)}, \quad |\beta_{\mathbf{k}}|^2 = \frac{\sinh^2\left(\frac{\pi\omega_-}{\rho}\right)}{\sinh\left(\frac{\pi\omega_{\mathrm{in}}}{\rho}\right)\sinh\left(\frac{\pi\omega_{\mathrm{out}}}{\rho}\right)} \,. \tag{4.71}$$

Hint: Use the identities for hypergeometric functions introduced in Appendix F.

Remark. In curved space an interesting connection between statistics and dynamics appears that is not present in Minkowski space. The statistics is determined by the algebra of creation and annihilation operators. Commuting operators give rise to bosons, while anti-commuting operators give rise to fermions. This is nothing new, but as we will argue, in curved spacetime this is the only consistent option!

To see this, consider a spin-0 field, i.e. a scalar field in curved space and two possible vacua, the a and b vacuum among which we can interpolate using the Bogoliubov coefficients defined above, satisfying $|\alpha|^2 - |\beta|^2 = 1$. Note that this relation follows purely from the field equations. Assume that the a creation and annihilation operators satisfy the (anti-)commutation relations

$$[a_n, a_m]_\pm = \left[a_m^\dagger, a_n^\dagger\right]_\pm = 0\,, \qquad \left[a_m, a_n^\dagger\right]_\pm = \delta_{mn}\,, \tag{4.72}$$

where the plus sign denotes the anti-commutator and the minus the commutator. Following the above, one can show that in the b vacuum, the (anti-)commutation relations become

$$[b_m, b_n]_\pm = \left[b_m^\dagger, b_n^\dagger\right]_\pm = (\alpha_{mn}\beta_{mn}^* \pm \alpha_{mn}\beta_{mn}^*)\delta_{m,-n}\,, \tag{4.73}$$

and

$$\left[b_m, b_n^\dagger\right]_\pm = (|\alpha|_{mn}^2 \pm |\beta|_{mn}^2)\delta_{m,n}\,, \tag{4.74}$$

Hence, we see that in order for the particles in the b vacuum to satisfy the same statistics as the particles in the a vacuum it is necessary to pick the minus sign such that (4.73) vanishes identically and (4.74) reduces to the standard commutation relations for bosons.

This is a purely curved space derivation, since in flat space $\beta = 0$ and the connection derived above is absent and both statistics seem allowed. Similar conclusions can be reached for fermionic fields as well as higher spin fields.

Chapter 5

Observables quadratic in fields

In the previous chapter we discussed the Hilbert space and non-uniqueness of the vacuum. Another important set of observables is given by the two-point functions. Indeed, a number of physically interesting quantities, such as the action and the energy-momentum tensor are quadratic in the fields and their derivatives, evaluated at a single point. As discussed in Chapter 2, the expectation values of such quantities usually diverges and has to be regularized. In flat space this could be done through normal ordering, but in curved space the implicit gravitational interactions introduce additional divergences. Furthermore, the vacuum energy needs to be treated very carefully since it can give rise to gravitational effects through the gravitational field equations. For these reasons more care is needed in the process of renormalisation.

5.1 Two-point functions

Similar to the situation in flat space, the most elementary Green's functions are the Wightman functions,

$$G^+(x,x') = \langle 0| \phi(x)\phi(x') |0\rangle \,, \qquad G^-(x,x') = \langle 0| \phi(x')\phi(x) |0\rangle \,,. \tag{5.1}$$

Using these expression, we can obtain the various other Green's functions, following the same identities as in flat space, see (2.36), (2.37) and (2.38).

Given the mode expansion

$$\phi(x) = \sum_n \left(a_n u_n(x) + a_n^\dagger u_n^*(x)\right) \,, \tag{5.2}$$

appropriate for a given choice of vacuum one can compute the Wightman function G^+ (and similarly G^-) as

$$G^+(x,x') = \sum_n u_n(x) u_n^*(x') \,. \tag{5.3}$$

Furthermore, since the field ϕ satisfies the wave equation we necessarily have

$$(\Box_x + m^2 + \xi R) G^\pm(x,x') = 0 \,. \tag{5.4}$$

Although conceptually clear, finding the appropriate basis of functions and computing the Wightman functions is often prohibitively hard and only in very special cases will we be able to obtain analytic expressions.

Such cases are usually characterised by some sort of additional symmetry. As in flat space, each symmetry gives rise to some sort of conserved quantity attached to a Cauchy surface Σ,

$$Q = \int_\Sigma J_\mu \mathrm{d}\Sigma^\mu \,, \qquad \nabla^\mu J_\mu = 0\,. \tag{5.5}$$

Due to Gauss law this charge is conserved, as long as the Cauchy surface does not cross any charged objects. This formula is perhaps most familiar for internal global symmetries but is easily extended to include space-time symmetries. Namely, when the space-time has a Killing vector K^μ we immediately find the conserved current $J_\mu = K^\mu T_{\mu\nu}$ with the associated charge being the momenta for (space) translation invariant theories, angular momenta for rotationally invariant theories and the Hamiltonian, or energy, for time translations.

The more symmetries are present in our setup the more constrained are the Green's functions. This will become most clear when working in maximally symmetric spaces, such as (A)dS, where the form of the Green's functions can be explicitly found after imposing all the symmetry constraints. See Chapter 8 and 10 for more details in these cases.

Another case where computations simplify dramatically is when we have a conformal vacuum, i.e. a conformal theory on a conformally flat manifold. In this case the Green's functions can simply be obtained through a conformal transformation from Minkowski space.

5.2 Adiabatic expansion of the Green's function

In the absence of symmetries, we have little hope of explicitly finding the mode functions and constructing the Green's functions. However, not all is lost as in many cases we can make progress by working perturbatively. We will be schematic and highlight the results while skipping many technical steps in the derivation. More details can be found in [BD84, PT09].

When the space-time curvature is small and slowly changing, we do not expect particles with arbitrary large energies to be created. In other words, if the metric changes sufficiently slow, or adiabatic, the particle number in a given mode should not change. We can make this more precise by introducing a formal parameter T in the metric,

$$g_{\mu\nu}(t,\mathbf{x}) \to g^T_{\mu\nu}(t,\mathbf{x}) = g_{\mu\nu}\left(\frac{t}{T},\frac{\mathbf{x}}{T}\right)\,. \tag{5.6}$$

We call the expansion in T^{-1} the adiabatic expansion, which roughly counts the number of derivatives of the metric. Indeed, we find that the scalar curvature R is of 'second adiabatic order'. We will not explicitly write this parameter but it can easily be introduced by counting derivatives. However, in the expansions below we will sometimes write $\mathcal{O}(T^{-3})$ for example which can be understood as up to terms of third adiabatic order.

To illustrate how this expansion can be used, let us consider again a free neutral scalar in $d = 3+1$ space-time dimensions, with equation of motion

$$(\Box + m^2 + \xi R)\phi = 0\,. \tag{5.7}$$

An all-powerful being would now proceed to find the mode functions and subsequently use them to compute the Green's function. For us, mere humans, this is usually not possible. However, one can show that Feynman's Green's function can be obtained directly as an adiabatic expansion. As discussed in flat space, the Feynman Green's function satisfies

$$(\Box + m^2 + \xi R)G_F(x,x') = -\delta(x,x'), \qquad \delta(x,x') = |g(x)|^{-1/2}\delta(x-x'). \tag{5.8}$$

The minus sign is convention, and both $\delta(x,x')$ and $G_F(x,x')$ transform as biscalars. To obtain the adiabatic expansion, rewrite the propagator as

$$G_F(x,x') = -\mathrm{i}\int_0^\infty \mathrm{d}s \mathrm{e}^{-\mathrm{i}m^2 s} K(x,x';s), \tag{5.9}$$

where m^2 is understood to have a small imaginary part $m^2 - \mathrm{i}\epsilon$ so that there is no divergence as $s \to \infty$. The kernel K satisfies the Schrödinger type equation

$$\mathrm{i}\partial_s K(x,x';s) = (\Box_x + \xi R)K(x,x';s), \tag{5.10}$$

with the boundary condition that $K(x,x';s) \sim |g(x)|^{-1/2}\delta(x-x')$ as $s \to 0$.

Exercise 5.1. *Show that the kernel $K(x,x';s)$ satisfies the equation (5.10) with the boundary conditions mentioned above.*

Equation (5.10) implies that we can expand the kernel K in powers of the parameter s. This can be made explicit by writing the kernel K as

$$K(x,x';s) = \mathrm{i}\frac{\Delta^{1/2}(x,x')}{(4\pi^2)(\mathrm{i}s)^2}\mathrm{e}^{\frac{\sigma(x,x')}{2\mathrm{i}s}}F(x,x';\mathrm{i}s), \tag{5.11}$$

where $\Delta(x,x')$ is the Van Vleck-Morette determinant and $\sigma(x,x')$ is related to the proper distance along the geodesic from x to x',[1]

$$\Delta(x,x') = -|g(x)|^{-1/2}\det\left(-\partial_{x^\mu}\partial_{x'^\nu}\sigma(x,x')\right)|g(x')|^{-1/2}, \qquad \sigma(x,x') = \frac{1}{2}\tau(x,x')^2, \tag{5.12}$$

where τ is the proper distance along the geodesic. The adiabatic expansion of G_F can now be rephrased as the following expansion of the function F,

$$F(x,x';s) \sim a_0(x,x') + (\mathrm{i}s)a_1(x,x') + (\mathrm{i}s)^2 a_2(x,x') + \cdots. \tag{5.13}$$

where the first coefficients are given in the coincidence limit $x \to x'$ as

$$a_0(x) = 1, \qquad a_1(x) = \left(\frac{1}{6} - \xi\right)R, \tag{5.14}$$

$$a_2(x) = \frac{1}{180}R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - \frac{1}{180}R_{\mu\nu}R^{\mu\nu} - \frac{1}{6}\left(\frac{1}{5} - \xi\right)\Box R + \frac{1}{2}\left(\frac{1}{6} - \xi\right)^2 R^2. \tag{5.15}$$

In this expansion we did not keep the parameter T explicit but by counting the derivatives on the metric, one can easily see that the term a_n is of adiabatic order $2n$. If the metric is smooth, one can

[1] Here we assume x' is in a normal neighbourhood of x such that only one geodesic goes from x to x'.

continue this expression indefinitely and find a unique expansion. However, typically this expansion is asymptotic so the solutions will in general not be uniquely determined.

Exercise 5.2. *To make this more explicit, consider the FLRW metric with flat spatial slices,*

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2 \mathrm{d}\mathbf{x}^2 \,. \tag{5.16}$$

The wave equation for the scalar field it

$$\left(\Box + m^2 + \xi R\right)\phi = 0 \,. \tag{5.17}$$

(Hint: Note that the Ricci scalar for this metric is $R = 6(\dot{a}^2/a^2 + \ddot{a}/a)$)

Show that the modes for the scalar field can be written as

$$f_{\mathbf{k}}(x) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} h_{\mathbf{k}}(t) \,, \tag{5.18}$$

where $h_{\mathbf{k}}(t)$ satisfies

$$\ddot{h}_{\mathbf{k}} + \left(\omega^2 + \sigma\right) h_{\mathbf{k}} = 0 \,, \tag{5.19}$$

where $\omega = \sqrt{\frac{\mathbf{k}^2}{a^2} + m^2}$. Find an expression for σ in terms of a and its derivatives.

The adiabatic expansion for the field modes is based on the usual WKB ansatz,

$$h_{\mathbf{k}}(t) = \frac{1}{\sqrt{W_{\mathbf{k}}(t)}} \exp\left[-\mathrm{i}\int^t W_{\mathbf{k}}(t')\mathrm{d}t'\right] , \tag{5.20}$$

where $W_{\mathbf{k}}(t)$ can be expanded in an adiabatic expansion,

$$W_{\mathbf{k}}(t) = \omega^{(0)} + \omega^{(0)} + \omega^{(0)} + \mathcal{O}(T^{-2}) \,, \tag{5.21}$$

where each $\omega^{(n)}$ is of nth adiabatic order, i.e. contains n derivatives with respect to t.

Show that $W_{\mathbf{k}}(t)$ satisfies the following equation,

$$W_{\mathbf{k}}^2 = \omega^2 + \sigma + \frac{3}{4}\frac{\dot{W}_{\mathbf{k}}^2}{W_{\mathbf{k}}^2} - \frac{1}{2}\frac{\ddot{W}_{\mathbf{k}}}{W_{\mathbf{k}}} \,, \tag{5.22}$$

and solve this equation perturbatively in a large T expansion. (Hint: explicitly reintroduce T and expand.)

Show that all odd $\omega^{(\mathrm{odd})}$ vanish and that

$$\omega^{(0)} = \omega \,, \tag{5.23}$$

$$\omega^{(2)} = \frac{1}{2\omega^3}\left(\sigma\omega^2 + \frac{3}{4}\dot{\omega}^2 - \frac{1}{2}\omega\ddot{\omega}\right) , \tag{5.24}$$

$$\omega^{(4)} = \frac{1}{2\omega^3}\left(2\sigma\omega\omega^{(2)} - 5\omega^2\left(\omega^{(2)}\right)^2 + \frac{3}{2}\dot{\omega}\dot{\omega}^{(2)} - \frac{1}{2}(\omega\ddot{\omega}^{(2)} + \ddot{\omega}\omega^{(2)}\right) . \tag{5.25}$$

Exercise 5.3. *Using the results from the previous exercise we can compute the Green's function in the* *adiabatic expansion. In the coincidence limit the Green's function takes the form*

$$G(x,x) \simeq \int_0^\infty \mathrm{d}k k^2 W_{\mathbf{k}}^{-1} \tag{5.26}$$

Expand this integral adiabatically and show that only the first to adiabatic orders contain divergences. More precisely, show that the Green's function can be written as,

$$G(x,x) = \frac{R}{288\pi^2} + \frac{1}{4\pi^2 a^3}\int_0^\infty \mathrm{d}k k^2 \left[\frac{1}{\omega} - \left(\xi - \frac{1}{6}\right)\frac{R}{2\omega^3}\right]. \tag{5.27}$$

After removing the divergent terms, we find a finite result for the Green's function (at coincident points).

Note that the adiabatic expansion described above can be continued to arbitrary order and it is even possible to find analytic expressions for all of the higher order terms in terms of first two [dRNS15]. The perturbative expansion is therefore uniquely determined. However, this expansion is an asymptotic expansion and is only uniquely defined up to non-perturbative terms.

5.3 Stress tensor renormalisation

In free theories, the Green's function suffice to solve the full theory as all higher point functions can simply be obtained by Wick contraction. As such, many interesting observables, like the stress tensor vacuum expectation value, are quadratic in the fields and can be obtained through coincidence limits of the Green's functions.

However, in four dimensions, for arbitrary ξ one can easily see that the first two terms in the expansion of the Green's function (5.13) are divergent in the UV limit, i.e. when $s \to 0$ and $\sigma \to 0$. This is easy to see after performing the ds integration for the first two terms,

$$G_F^{(2)}(x,x') = -\mathrm{i}\frac{|g(x)|^{1/4}}{4\pi^2}\left[\frac{m}{\sqrt{-2\sigma}}K_1(m\sqrt{-2\sigma}) + \frac{a_1(x,x')}{2}K_0(m\sqrt{-2\sigma})\right], \tag{5.28}$$

where K_i are Bessel functions of the second kind. Higher order terms do not involve UV divergences for the two-point function but as we will see the fourth adiabatic order is necessary to tame the logarithmic divergence of the stress tensor vacuum expectation value.

In order to obtain the physical expectation value for the two-point function evaluated at coincident points, we need to subtract the divergent pieces. The adiabatic regularisation or subtraction method tells us that we can do so in $n = 4$ dimension as

$$\begin{aligned}\langle 0|\,\phi(x)\phi(x')\,|0\rangle_{\text{phys.}} = \lim_{n\to 4}\Bigg[&\langle 0|\,\phi(x)\phi(x')\,|0\rangle - \mathrm{i}\Delta^{1/2}(x,x')(4\pi)^{-n/2}\times \\ &\times \int_0^\infty \mathrm{d}s(\mathrm{i}s)^{-n/2}\mathrm{e}^{-\mathrm{i}m^2 s + \frac{\sigma(x,x')}{2\mathrm{i}s}}\big(1 + a_1(x,x')(\mathrm{i}s)\big)\Bigg].\end{aligned} \tag{5.29}$$

This quantity has a smooth, well-defined limit as $x \to x'$ in any space-time. Note that we took a limit $n \to 4$, to regularise the infinite pieces. This is an example of dimensional regularisation.

Next, let us apply this procedure to the stress tensor vacuum expectation value $\langle T \rangle$ and let us focus on a conformally coupled scalar, with $\xi = \frac{1}{6}$ in the massless limit, $m \to 0$. The trace of the stress tensor is given by

$$T^{\mu}{}_{\mu} = -\nabla^{\mu}\phi\nabla_{\mu}\phi + 2m^2\phi^2 + \frac{1}{6}R\phi^2 + \frac{1}{2}\Box\phi^2\,, \tag{5.30}$$

which after applying the equations of motion simply becomes

$$T^{\mu}{}_{\mu} = m^2\phi^2\,. \tag{5.31}$$

Formally one might therefore be tempted to identify the vacuum expectation values

$$\left\langle T^{\mu}{}_{\mu}\right\rangle = m^2\left\langle\phi^2\right\rangle\,, \tag{5.32}$$

and conclude that in the massless case the expectation value of the trace vanishes. However, note that (5.32) does not imply that $\left\langle T^{\mu}{}_{\mu}\right\rangle_{\text{phys.}} = m^2\left\langle\phi^2\right\rangle_{\text{phys.}}$! The reason is that the separate components of the stress tensor have residual logarithmic divergences after subtracting the counterterms (5.29). Indeed, in general one has to be very cautious in using equations of motion in the physical expectation values.

In order to cancel the additional logarithmic divergence we need to furthermore subtract the counterterm proportional to a_2, in order to find,

$$\begin{aligned}\langle 0|\, T^{\mu}{}_{\mu}(x)\,|0\rangle_{\text{phys.}} = \lim_{n\to 4} m^2\Big[\,&\langle 0|\,\phi(x)^2\,|0\rangle - \mathrm{i}\Delta^{1/2}(x,x')(4\pi)^{-n/2}\times\\ &\times\int_0^{\infty}\mathrm{d}s(\mathrm{i}s)^{-n/2}\mathrm{e}^{-\mathrm{i}m^2 s+\frac{\sigma(x,x')}{2\mathrm{i}s}}\left(1+a_1(x,x')(\mathrm{i}s)+a_2(x,x')(\mathrm{i}s)^2\right)\Big]\,.\end{aligned} \tag{5.33}$$

Note that in the massless limit, this expression reduces to

$$\lim_{m\to 0}\langle 0|\, T^{\mu}{}_{\mu}(x)\,|0\rangle_{\text{phys.}} = -\frac{1}{(4\pi)^2}a_2(x)\big|_{\xi=\frac{1}{6}}\,. \tag{5.34}$$

where for $\xi = \frac{1}{6}$ the second adiabatic coefficient simplifies to

$$\begin{aligned}a_2(x)\big|_{\xi=\frac{1}{6}} &= \frac{1}{180}R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - \frac{1}{180}R_{\mu\nu}R^{\mu\nu} - \frac{1}{180}\Box R\\ &= \frac{1}{120}W - \frac{1}{180}E - \frac{1}{360}\Box R\,.\end{aligned} \tag{5.35}$$

where we introduced the square of the Weyl tensor and Euler density (also known as the Gauss-Bonnet term in four dimensions),

$$\begin{aligned}W &= C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma} = R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - 2R_{\mu\nu}R^{\mu\nu} + \frac{1}{3}R^2\,,\\ E &= R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - 4R_{\mu\nu}R^{\mu\nu} + R^2\,.\end{aligned} \tag{5.36}$$

This quantum violation of the tracelessness of the stress tensor in a conformal theory is known as the trace anomaly and is characterised in four dimensions by a and c which appear as

$$\left\langle T^{\mu}{}_{\mu}\right\rangle = (a\,E - c\,W + b\,\Box R)\;. \tag{5.37}$$

Hence with our chosen normalisation we see that the conformally coupled free scalar contribute as follows to the anomalies,

$$(a,c)_{\text{conformal free scalar}} = \frac{1}{90(8\pi)^2}(1,3)\,. \tag{5.38}$$

The coefficient b on the other hand is scheme dependent and can be removed by adding appropriate UV counterterms. The constants a and c are scheme independent and therefore characterise the theory. Similar to the c trace anomaly in two dimension, the a anomaly is monotonous along the RG flow. Indeed, if we have a flow from a UV CFT to another IR CFT we have $a_{\text{UV}} > a_{\text{IR}}$ [Zam86, KS11]. In four dimensions, the constant c does not universally satisfy such monotonicity properties. See your course on CFT for more details on a-theorems and trace anomalies.

In this chapter we discussed how to renormalise the scalar two-point function and stress tensor vacuum expectation value using the adiabatic subtraction scheme. In the literature a variety of other renormalisation schemes has been used, such as Hadamard renormalisation, proper time renormalisation or dimensional regularisation to name a few. All these schemes give a different prescription to remove the infinities, however, on scheme independent quantities they should and are in many case proven to produce identical results.

Part II

APPLICATIONS

Chapter 6

The Unruh effect

A fundamental application of the formalism introduced above can already be seen in flat space. Indeed, the general principles of relativity state that it should be possible to express the laws of physics as being the same for all observers, even those undergoing acceleration. Indeed, the effects of constant acceleration are equivalent to those caused by the presence of a uniform gravitational field. Thus one can ask the question: in flat space, how does the Minkowski vacuum appear to an accelerating observer? The perhaps surprising answer to this question lies in the Unruh effect which states that this observer will perceive a thermal state!

Studying the physics of an accelerated observer will highlight many of the properties we discussed in the previous chapters. Due to its simplicity we will be able to exactly solve this problem and explicitly see the theory at work. Moreover, as we will see in the next chapter, many of the curious properties observed in this case immediately carry over to the study of an evaporating black hole.

6.1 Thermal states

Before deriving the Unruh effect, let us briefly review some aspects of thermal states in quantum field theory. Thermal states are a feature of statistical physics at a temperature T, the equilibrium state is a probability distribution of physical states. In quantum mechanics such a distribution is given in the form of a density matrix,

Definition 6.1. A density matrix is an element $\rho \in \mathcal{H} \otimes \mathcal{H}^*$ that is Hermitian, positive definite and has unit trace $\operatorname{tr} \rho = 1$.

Such matrices are always diagonalizable and an orthonormal basis of states $|n\rangle$ can be found such that they can be expressed in the form

$$\rho = \sum_n p_n |n\rangle \langle n| \tag{6.1}$$

where the coefficients are positive $p_n \geq 0$ and $\sum_n p_n = 1$. In this context, the coefficients p_n can be thought of as the probability of the ensemble to be in the state $|n\rangle$ such that the expectation value of an observable A can be computed as

$$\langle A \rangle_\rho = \operatorname{tr} \rho A = \sum_n \rho_n \langle n|A|n\rangle \,. \tag{6.2}$$

A density matrix represents a pure state when $\rho = |\psi\rangle \langle \psi|$ for some normalized state $|\psi\rangle$, i.e., ρ has rank 1, otherwise states are said to be mixed. Mixed states arise naturally when part of a quantum system is hidden. Consider for example a pure state $|\psi\rangle \in \mathcal{H} = \mathcal{H}_L \otimes \mathcal{H}_R$ where the Hilbert space $\mathcal{H}_L$ is hidden for an observer. The state observed by the observer is then obtained by the partial trace over $\mathcal{H}_L$. More precisely, consider the state $|\psi\rangle$

$$|\psi\rangle = \sum_{r,s} \psi_{r,s} |r\rangle_L \otimes |s\rangle_R \, . \tag{6.3}$$

where $\{|r\rangle\}$ represents a basis of $\mathcal{H}_L$ and $\{|s\rangle\}$ of $\mathcal{H}_R$. The partially traced density matrix perceived by the observer is then given by

$$\rho_R = \mathrm{Tr}_L |\psi\rangle \langle\psi| = \sum_r \bar{\psi}_{r,s_1} \psi_{r,s_2} |s_2\rangle \langle s_1| \, . \tag{6.4}$$

The observed state is mixed if the density matrix ρ_R has rank greater than one. In this case the systems R and L are said to be entangled. Given a Hamiltonian H, we define a thermal state as follows,[1]

Definition 6.2. A thermal state of temperature T is a state of the form

$$\rho_\beta = \frac{1}{Z_\beta} \exp(-\beta H) = \frac{1}{Z_\beta} \sum_n \mathrm{e}^{-\beta E_n} |n\rangle \langle n| \, , \qquad \beta = \frac{1}{T} \, , \tag{6.5}$$

where $\langle n|$ is a basis of energy eigenstates of energy E_n and

$$Z_\beta = \mathrm{Tr} \exp(-\beta H) = \sum_n \mathrm{e}^{-\beta E_n} \, , \tag{6.6}$$

is the partition function.

Note that we use units where the Boltzmann's constant $k_B = 1$. This is in exact analogy with the canonical ensemble in statistical mechanics.

Example 6.1 (Bose-Einstein distribution)**.** *In the context of a harmonic oscillator (i.e., a single mode of a quantum field), we can compute the thermal expectation of the number operator N using $E_n = (n + \frac{1}{2})\omega$ to obtain*

$$\mathrm{Tr}(\rho_\beta N) = \frac{\sum_n n \mathrm{e}^{-\beta n \omega}}{\sum_n \mathrm{e}^{-\beta n \omega}} = \frac{1}{\omega} \frac{d}{d\beta} \log(1 - \mathrm{e}^{-\beta\omega}) = \frac{1}{\mathrm{e}^{\beta\omega} - 1} \, . \tag{6.7}$$

A useful way to characterize thermal states in quantum mechanics and quantum field theory was described by Kubo, Martin and Schwinger [Kub57, MS59] and is called the KMS condition. Similarly a state satisfying the KMS condition is called a KMS state.

Definition 6.3 (KMS state)**.** A KMS state is a state for which the time evolution of operators $A \to A_t$ can be continued to complex time in such a way that for a time-independent operator, B, we have

$$\langle A_t B \rangle_{KMS} = \left\langle B A_{t+i\beta} \right\rangle_{KMS} \, , \tag{6.8}$$

where $\langle A_z B \rangle_{KMS}$ and $\langle B A_z \rangle_{KMS}$ are analytic functions of z in the strip $0 < \mathrm{im}\, z < \beta$.

[1] One can decorate this definition with chemical potentials when the particle number is allowed to change, i.e. in a grand canonical ensemble.

For finite systems the definition of the KMS condition is equivalent to the the definition of thermal states above. To see this recall that in the Heisenberg representation, an operator A has time dependence $A_t = \mathrm{e}^{iHt} A_0 \mathrm{e}^{-iHt}$. For our thermal density matrix above we can compute

$$\begin{aligned}
\langle A_t B\rangle_\beta &= \frac{1}{Z_\beta}\,\mathrm{tr}(\mathrm{e}^{-\beta H} A_t B) = \frac{1}{Z_\beta}\,\mathrm{tr}(\mathrm{e}^{-\beta H + iHt} A_0 \mathrm{e}^{-iHt} B) \\
&= \frac{1}{Z_\beta}\,\mathrm{tr}(A_{t+i\beta}\mathrm{e}^{-\beta H} B) = \frac{1}{Z_\beta}\,\mathrm{tr}(\mathrm{e}^{-\beta H} B A_{t+i\beta}) \\
&= \langle B A_{t+i\beta}\rangle_\beta \,,
\end{aligned}$$

where we have used the cyclic property of the trace. This is often interpreted as the property that our system can be analytically continued to Euclidean signature with periodicity in imaginary time. In infinite dimensions these manipulations are a lot more subtle as we might encounter phase transitions, spontaneous symmetry breaking, operators that are not trace class and so on. However, the content of the KMS condition is precisely that a similar relation continues to hold in the thermodynamic limit.[2]

We can apply the same analytic continuation to Green's functions. The thermal Green's function can be obtained as the analytic continuation of the Wightman function or alternatively as the Green's function for the relevant operator, i.e., the Laplacian on $\mathbb{R}^3 \times S^1_\beta$ where now S^1 is a circle of length β.

Definition 6.4. The thermal Green's function is defined as

$$G_\beta(x,y) = \frac{\mathrm{Tr}(\mathrm{e}^{-\beta H}\phi(x)\phi(y))}{\mathrm{Tr}(\mathrm{e}^{-\beta H})}\,. \tag{6.9}$$

Furthermore, when our system has a time-like symmetry, the thermal Green's function $G_\beta = G_\beta(t - t', \mathbf{x}, \mathbf{x}')$ and satisfies the KMS condition, i.e. it is periodic in imaginary time with period $i\beta$. This property follows directly as above from the KMS condition.

Example 6.2. *In flat space, the thermal propagator can be constructed by images in imaginary time of period β. Thus we can identify the thermal greens function on Minkowski space for the massless wave equation as*

$$G_\beta(x,x') = \sum_{n\in\mathbb{Z}} \frac{1}{4\pi^2((t-t'+in\beta+i\epsilon)^2 - \mathbf{x}\cdot\mathbf{x})}\,. \tag{6.10}$$

6.2 Particle detectors

Since in general curved space-times the notion of a particle is observer-dependent it will prove useful to give a coordinate independent characterisation of the temperature. A useful way to achieve this is to consider an observer equipped with a so-called Unruh detector [Unr76, GH77b].

The detector will have some internal energy states and can interact with the scalar field by exchanging energy, i.e. by emitting or absorbing scalar particles. The detector could for example be constructed so that it emits a 'ping' whenever its internal energy state changes. All observers will agree on whether or not the detector has pinged, although they may disagree on whether the ping was caused by particle

[2] If a phase transition takes place or if some symmetry is spontaneously broken, the KMS state might not be uniquely defined.

emission or absorption. Such a detector can be modelled by a coupling of the scalar field $\phi(x(\tau))$ along the world-line $x(\tau)$ of the observer to some operator $m(\tau)$ acting on the internal detector states

$$g\int_{-\infty}^{\infty} \mathrm{d}\tau m(\tau)\phi(x(\tau)), \tag{6.11}$$

where g is the strength of the coupling and τ is the proper time along the observer's world-line. Let H_0 denote the detector Hamiltonian, with energy eigenstates $\left|E_j\right\rangle$,

$$H_0\left|E_j\right\rangle = E_j\left|E_j\right\rangle, \tag{6.12}$$

and let m_{ij} be the matrix element of the operator $m(\tau)$ at $\tau = 0$,

$$m_{ij} = \langle E_i|\, m(0)\left|E_j\right\rangle. \tag{6.13}$$

We will calculate the transition amplitude from a state $|0\rangle \otimes |E_i\rangle \in \mathcal{H}_\phi \otimes \mathcal{H}_{\text{det.}}$ in the tensor product of the scalar field and detector Hilbert spaces to the state $\left\langle E_j\right| \otimes \langle\psi|$, where $|\psi\rangle$ is any state of the scalar field. To first order in perturbation theory (for small g) the desired amplitude can be computed as

$$\mathcal{A} = g\int_{-\infty}^{\infty} \mathrm{d}\tau \left\langle E_j\right| \otimes \langle\psi|\, m(\tau)\phi(x(\tau))\,|0\rangle \otimes |E_i\rangle\,. \tag{6.14}$$

Using (in the Heisenberg picture) that $m(\tau) = \mathrm{e}^{\mathrm{i}H\tau}m(0)\mathrm{e}^{-\mathrm{i}H\tau}$, this can be written as

$$\mathcal{A} = g m_{ji}\int_{-\infty}^{\infty} \mathrm{d}\tau\, \mathrm{e}^{\mathrm{i}(E_j - E_i)\tau}\,\langle\psi|\,\phi(x(\tau))\,|0\rangle\,. \tag{6.15}$$

Since we are only interested in the probability for the detector to make the transition from E_i to E_j, we should square this amplitude and sum over the final state $|\psi\rangle$ of the scalar field, which will not be measured. Using the resolution of identity $\sum_\psi |\psi\rangle\langle\psi| = 1$ we find the probability

$$P(E_i \to E_j) = g^2|m_{ij}|^2\int_{-\infty}^{\infty} \mathrm{d}\tau \mathrm{d}\tau' \mathrm{e}^{-\mathrm{i}(E_j - E_i)(\tau' - \tau)} G_+\left(x(\tau'), x(\tau)\right), \tag{6.16}$$

where G_+ is the Wightman function. Notice that the prefactor in (6.16) depends on the details of the detector, so it is useful to extract the piece which depends only on the scalar field and the world-line trajectory. For this reason we define the detector response function

$$\mathcal{F}(E) = \int_{-\infty}^{\infty} \mathrm{d}\tau \mathrm{d}\tau' \mathrm{e}^{-\mathrm{i}E(\tau' - \tau)} G_+\left(x(\tau'), x(\tau)\right), \tag{6.17}$$

When the Wightman function only depends on $\Delta\tau = \tau' - \tau$ we can change variables to $\Delta\tau$ and $\bar{\tau} = \frac{\tau' + \tau}{2}$. The detector response function is then defined by removing the diverging volume factor coming from the integration over $\bar{\tau}$,

$$f(E) = \int_{-\infty}^{\infty} \mathrm{d}\Delta\tau\, \mathrm{e}^{-\mathrm{i}E\Delta\tau} G_+(\Delta\tau). \tag{6.18}$$

Example 6.3. *First consider Minkowski space in the vacuum state. The Wightman function is then given by*

$$G_+(\Delta\tau) = \frac{1}{4\pi^2}\frac{1}{(\Delta\tau - \mathrm{i}\epsilon)^2}\,. \tag{6.19}$$

If we plug this in the formula (6.18), we can calculate the integral by residues. Since $E > 0$ we should close the contour for $\Delta\tau$ in the lower half-plane, since in this case the integral at the half-circle at infinity goes to zero due to the damping factor $\mathrm{e}^{-\mathrm{i}E\,\Delta\tau}$ (Jordan's lemma) and we conclude that

$$f(E) = 0\,. \tag{6.20}$$

So, unsurprisingly, we find that there is no particle detection in the Minkowski vacuum.

Example 6.4. *We can also arrive at the Bose-Einstein distribution from the Green's function by computing the detector responds function for a detector at $\mathbf{x} = 0$. Inserting the thermal Green's function in (6.18) we find*

$$\begin{aligned} f(E) &= \int_{-\infty}^{\infty} \mathrm{d}\Delta\tau\, \mathrm{e}^{-\mathrm{i}E\,\Delta\tau}\, G_\beta(\Delta\tau) \\ &= \int_{-\infty}^{\infty} \mathrm{d}\Delta\tau \sum_{n\in\mathbb{Z}} \frac{-\mathrm{e}^{-\mathrm{i}E\,\Delta\tau}}{4\pi^2(\Delta\tau + \mathrm{i}n\beta + i\epsilon)^2} \end{aligned} \tag{6.21}$$

The integral can be analytically continued so as to be evaluated by residues along the negative imaginary axis where it has double poles at $t = -n\beta i$. Thus we obtain the sum of residues

$$f(E) = E\sum_{n\in\mathbb{Z}} \mathrm{e}^{-En\beta} = \frac{E}{\mathrm{e}^{\beta E} - 1}\,, \tag{6.22}$$

in line with our expectation for a detector immersed in a thermal bath of temperature $T = 1/\beta$.

6.3 The Unruh effect

Having introduced these tools we are ready and well-equipped to study the Unruh effect. Here we will perform the calculation for massless modes in $1+1$ dimensions where we can use the conformal invariance of the wave equation allowing us to explicit perform calculations. All the essential features will already be present in this setup. The analysis can be performed explicitly also for massive modes and be extended to higher dimensions, but only at the expense of having to deal with Bessel function identities.[3]

Consider an observer $\mathcal{O}_a$ with constant acceleration a along the x-axis. The world-line for this observer can be parametrised as

$$X(\tau) = (t(\tau), x(\tau)) = \frac{1}{a}(\sinh a\tau, \cosh a\tau)\,, \tag{6.23}$$

where τ is the proper time of the observer. Note that these trajectories parametrise the hyperbolae, $x^2 - \tau^2 = a^2$. We will be asking the question as to how the accelerating observer sees the Minkowski

[3] Going to higher dimensions is no worse than introducing a mass, but we lose the conformal invariance that we exploited in this section.

vacuum. To do so we want to find the natural coordinates in Minkowski space-time adapted to this observer. I.e. we want the time coordinate to be its proper time, and the spatial coordinate to be characterized by the fact that the observed is at rest in it. Using the clock and radar method, the observer will set up coordinates (τ, ξ) related to inertial coordinates (t, x) by

$$(t,x) = \frac{e^{a\xi}}{a}(\sinh a\tau, \cosh a\tau), \qquad ds^2 = dt^2 - dx^2 = e^{2a\xi}(d\tau^2 - d\xi^2). \tag{6.24}$$

Indeed, in these coordinates, the path followed by the observer is given by $X(\tau) = (\tau, 0)$. The new coordinates have range $\{\tau, \xi\} \in (-\infty, \infty)$ but they only cover the part of Minkowski space with $\mathcal{R} : \{x > |t|\}$, called the Rindler wedge. This is the portion of space-time that the accelerating observer can measure and see. Minkowski space equipped with this metric will be denoted Rindler space. Note that Rindler space corresponds to the right wedge foliated by the world-lines of the accelerated observers, labelled by $\mathcal{R}$ in Figure 6.1.

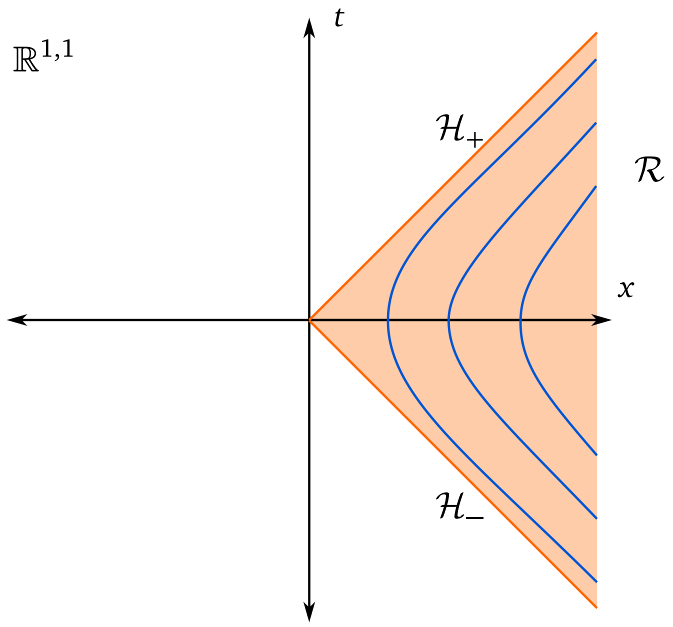


Figure 6.1: The Rindle wedge $\mathcal{R}$ in two-dimensional Minkowski space is denote by the orange region. The uniformly accelerated observer $\mathcal{O}_a$ follows the blue hyperbolic trajectories in $\mathcal{R}$. $\mathcal{H}_\pm$ denote the Killing horizons which are the boundaries of $\mathbb{R}^{1,1}$ perceivable for this observer.

The inertial light-cone coordinates $(u, v) := (t - x, t + x)$ can be rewritten in terms of the light-cone coordinates adapted to the accelerating observer $(U_R, V_R) := (\tau - \xi, \tau + \xi)$,

$$(u,v) = \frac{1}{a}\left(-e^{-aU}, e^{aV}\right), \qquad ds^2 = du dv = e^{a(V_R - U_R)} dU_R dV_R. \tag{6.25}$$

The Rindler wedge is therefore the whole plane in the (U_R, V_R) coordinates but only the quadrant with $-u, v > 0$ in the (u, v) coordinates.

More generally, we see that the lines of constant ξ in the Rindler metric describe uniformly accelerated observers with acceleration

$$\alpha = a e^{-a\xi}, \tag{6.26}$$

and proper time τ. Therefore, Rindler space can be regarded as a foliation of Minkowski space by the trajectories of uniformly accelerated observers. Near the horizon where $\xi \to -\infty$, we have $\alpha \to \infty$ such that the observers feel an infinite proper acceleration.

Similarly, we can also cover the left wedge of Minkowski space, $x < |t|$, by defining the coordinates

$$t = -\frac{e^{a\xi}}{a}\sinh a\xi\,, \qquad x = -\frac{e^{a\xi}}{a}\cosh a\xi\,. \tag{6.27}$$

Notice that the events happening in the left Rindler wedge are causally disconnected from the world-lines of a Rindler observer in the right Rindler wedge, and the line $u = 0$ effectively behaves as an event horizon. This observation will be relevant in the context of our discussion of black holes in the next chapter.

Let us now consider the quantisation of a massless scalar field in Rindler space. Since the wave equation in this case is conformal, we can trivially solve it by applying a conformal transformation to the standard Minkowski space modes. In standard Minkowski space, we have respectively the left and right moving modes[4]

$$\phi_\omega(u) = \frac{1}{\sqrt{4\pi\omega}}e^{-i\omega u}\,, \qquad \tilde{\phi}_\omega(v) = \frac{1}{\sqrt{4\pi\omega}}e^{-i\omega v}\,, \tag{6.28}$$

which constitute the positive frequency modes for $\omega > 0$. In the adapted coordinates introduced above the metric is conformal to the standard Minkowski metric so the modes seen/measured by $\mathcal{O}_a$ of frequency ω will be respectively

$$\Phi^R_\omega(U_R) = \frac{\theta(-u)}{\sqrt{4\pi\omega}}e^{-i\omega U_R}\,, \qquad \tilde{\Phi}^R_\omega(V_R) = \frac{\theta(v)}{\sqrt{4\pi\omega}}e^{-i\omega V_R}\,. \tag{6.29}$$

where we added the Heaviside functions to denote that these modes are only non-zero in the right Rindler wedge.

Similarly, considering the left Rindler wedge L with lightcone coordinates U_L and V_L we find the modes for the quantum field there as,

$$\Phi^L_\omega(U_L) = \frac{\theta(u)}{\sqrt{4\pi\omega}}e^{i\omega U_L}\,, \qquad \tilde{\Phi}^L_\omega(V_L) = \frac{\theta(-v)}{\sqrt{4\pi\omega}}e^{i\omega V_L}\,. \tag{6.30}$$

Given that the left- and right-moving modes decouple, we can focus on the right-moving modes while the results for left-moving modes will follow identically.[5] In the standard Minkowski picture, we can define a general right-moving field operator as

$$\hat{\phi} = \int_0^\infty \frac{d\omega}{\sqrt{4\pi\omega}}[e^{-i\omega u}a_\omega + e^{i\omega u}a^\dagger_\omega]\,, \tag{6.31}$$

while for an accelerating observer in the right Rindler wedge the right-moving field operators are

[4]In two dimensions we have $|\mathbf{k}| = \omega$ where $\mathbf{k}$ only has one component. The right-moving waves have $k > 0$ while the left-moving ones have $k < 0$.

[5]The left and right movers can never mix under Bogoliubov transformations, since $\Phi^{R/L}_\omega(U_{R/L})$ only depends on $U_{R/L}$ while $\tilde{\Phi}^{R/L}_\omega(V_{R/L})$ only depends on $V_{R/L}$.

defined as

$$\hat{\Phi}_R = \int_0^\infty \frac{\mathrm{d}\lambda}{\sqrt{4\pi\lambda}} [\mathrm{e}^{-i\lambda U_R} A^R_\lambda + \mathrm{e}^{i\lambda U_R} A^{R\dagger}_\lambda], \tag{6.32}$$

where a_ω and A^R_λ are the standard raising and lowering operators satisfying the commutation relations

$$[a_\omega, a^\dagger_{\omega'}] = \delta(\omega - \omega'), \qquad [A_\lambda, A^\dagger_{\lambda'}] = \delta(\lambda - \lambda'), \tag{6.33}$$

and analogous for the left Rindler wedge. Implicit in these definitions are the Minkowski vacuum $|0_M\rangle$, satisfying $a_\omega |0_M\rangle = 0$, and the Rindler vacuum satisfying $A_\lambda |0_R\rangle = 0$ for respectively all $\omega > 0$ and $\lambda > 0$.

As by now standard, we can build a Fock space $\mathcal{F}_{R/L}$ on the Rindler vacua $\left|0_{R/L}\right\rangle$ using the respective creation and annihilation operators$A^{R/L\dagger}_\lambda$ and $A^{R/L}_\lambda$. These Fock spaces are based on the Rindler modes (6.29) or (6.30) measured by an observer $\mathcal{O}^{R/L}_a$ respectively in the right or left Rindler wedge. In particular the vacua $|0\rangle_{R/L}$ is the state in which $\mathcal{O}^{R/L}_a$ sees no particles. However, As it stands, there can be no identification between the Minkowski Fock space $\mathcal{F}_M$ and either of the Rindler Fock space $\mathcal{F}_{R/L}$ separately as they only determine the Minkowski fields respectively for $u < 0$ or $u > 0$ and are not defined in the opposite patch. To determine $\hat{\phi}$ and $\mathcal{F}_M$ from Rindler type data, we have to consider both the left and right Fock spaces $\mathcal{F}_R$ and $\mathcal{F}_L$ such that the Minkowski Fock space is now a tensor product

$$\mathcal{F}_M = \mathcal{F}_L \otimes \mathcal{F}_R\,. \tag{6.34}$$

Note however, that the Minkowski vacuum $|0\rangle_M$ might be an entangled state in this product, i.e. $|0\rangle_M \neq |0\rangle_L \otimes |0\rangle_R$.

Having discussed in detail the various vacua in the picture we now wish to compute the distribution of the number of particles of frequency λ detected by the observer $\mathcal{O}_a$ in the Minkowski vacuum. To do so we need to compute the Bogoliubov coefficients relating the Minkowski and Rindler modes in the right Rindler wedge,

$$\theta)(-u)\phi_\omega(u) = \int \mathrm{d}\lambda \left[\alpha^R_{\omega\lambda} \Phi^R_\lambda(U_R) + \beta^R_{\omega\lambda} \Phi^{R*}_\lambda(U_R)\right]. \tag{6.35}$$

Inserting the modes in the definition for the Bogoliubov coefficients, (4.38), [6] we find,

$$\begin{aligned}
\alpha^R_{\omega\lambda} =& \mathrm{i} \int_{-\infty}^\infty \mathrm{d}U_R\, \Phi^{R*}_\lambda \overleftrightarrow{\partial_u}\, \xi_\omega \\
=& \frac{1}{2\pi\sqrt{\omega\lambda}} \int_{-\infty}^0 \mathrm{d}u \lambda \mathrm{e}^{-\mathrm{i}\omega u} (-au)^{-\frac{\mathrm{i}\lambda}{a} - 1} \\
=& \frac{1}{2\pi a} \sqrt{\frac{\lambda}{\omega}} \left(\frac{a}{\omega}\right)^{-\frac{\mathrm{i}\lambda}{a}} \Gamma\left(-\frac{\mathrm{i}\lambda}{a}\right) \mathrm{e}^{\frac{\pi\lambda}{2a}}\,.
\end{aligned} \tag{6.36}$$

[6]where the inner product is defined on the hypersurface $\xi =$ constant with normal vector $n^\mu = \mathrm{e}^{-a\xi}(1,0)$

Similarly, for $\beta^R_{\omega\lambda}$ we find

$$\begin{aligned}\beta^R_{\omega\lambda} =& \mathrm{i}\int_{-\infty}^{\infty} \mathrm{d}U_R\, \Phi^R_\lambda \overleftrightarrow{\partial_u}\, \xi_\omega \\ =& \frac{1}{2\pi a}\sqrt{\frac{\lambda}{\omega}}\left(\frac{a}{\omega}\right)^{\frac{\mathrm{i}\lambda}{a}}\Gamma\left(\frac{\mathrm{i}\lambda}{a}\right)\mathrm{e}^{-\frac{\pi\lambda}{2a}}\,.\end{aligned} \tag{6.37}$$

The verification of the intermediate steps are left as an exercise for the reader but mainly consist of rewriting the integral in terms of an integral representation of the Gamma function and using some Gamma function identities.

The main takeaway from this calculation is the relation

$$|\alpha_{\omega\lambda}|^2 = \mathrm{e}^{2\pi\lambda/a}|\beta_{\omega\lambda}|^2\,, \tag{6.38}$$

as this allows us to compute the expectation value of the number operator N_λ of the modes with frequency λ detected by $\mathcal{O}_a$ in the Minkowski vacuum,

$$\begin{aligned}\langle N_\lambda\rangle_M :=& \langle 0|_M A^{R\dagger}_\lambda A^R_\lambda|0_M\rangle \\ =& \int_0^\infty d\omega|\beta_{\omega\lambda}|^2\end{aligned} \tag{6.39}$$

This can be simplified by using the normalization condition for Bogoliubov coefficients which in our context reads

$$\int_0^\infty d\omega(\alpha_{\omega\lambda}\bar{\alpha}_{\omega\lambda'} - \beta_{\omega\lambda}\bar{\beta}_{\omega\lambda'}) = \delta(\lambda-\lambda')\,. \tag{6.40}$$

Evaluating at $\lambda = \lambda'$, we reinterpret the right hand side $\delta(0) = V$ as the volume of space, regularized as usual by putting the system in a finite box. This allows us to deduce the particle number density as

$$\frac{\langle N_\lambda\rangle_M}{V} = \frac{1}{\mathrm{e}^{2\pi\lambda/a}-1}\,. \tag{6.41}$$

This is the main result of this section as we now recognize this as the Bose Einstein distribution with Unruh temperature

$$T_{\text{Unruh}} = \frac{a}{2\pi}\,. \tag{6.42}$$

We conclude that an observer moving with uniform acceleration through the Minkowski vacuum observes a thermal spectrum of particles. The Unruh temperature $T = \frac{a}{2\pi}$ is the temperature that would be measured by an observer moving along the path $\xi = 0$, which feels the acceleration $\alpha = a$. Any other path with $\xi =$ constant feels an acceleration

$$\alpha = a\mathrm{e}^{-a\xi}\,, \tag{6.43}$$

and will thus measure thermal radiation at temperature $T = \frac{\alpha}{2\pi}$. As $\xi \to \infty$, the temperature approaches 0, in line with the fact that near ∞ the Rindler observer is nearly inertial. We conclude that not only does the choice of vacuum, and hence concept of particle become time-dependent, it is also observer-dependent, even in flat space-time.

Coming back to the relation between the Minkowski and Rindler vacua we have that the Unruh state ρ_U as measured by $\mathcal{O}_a$ is given by the density matrix

$$\rho_U = \mathrm{Tr}_{\mathcal{F}_L} |0\rangle_M \langle 0|_M \; . \tag{6.44}$$

A key point to analyse this state is that time translation ∂_τ for $\mathcal{O}_a$ in R is given by the boost Killing vector on $\mathcal{R}$ but in $\mathcal{L}$ it is given by minus the boost Killing vector

$$B := x\partial_t + t\partial_x = v\partial_v - u\partial_u = \frac{\partial}{\partial \tau_R} = -\frac{\partial}{\partial \tau_L} \tag{6.45}$$

However, since the Minkowski vacuum is Lorentz and hence boost invariant we must have an entangled product of the form

$$|0\rangle_M = \sum_n f_n |n\rangle_L \otimes |n\rangle_R \,, \tag{6.46}$$

for some f_n. Indeed, our calculations show that $f_n = \mathrm{e}^{-\beta E_n}$, where $\beta = 1/T_{\text{Unruh}}$.

Alternatively we can recognize the thermal nature of the Rindler vacuum by looking at the KMS condition on the Wightman function in the Rindler vacuum. Since we are working in a conformal setup we can straightforwardly obtain the Rindler space Wightman function from the Minkowski Wightman function. Indeed, in two dimensions it is unchanged so we simply have to change coordinates to obtain the Wightman function in Rindler space

$$\begin{aligned} G_+^M(x,x') &= \langle 0|_M \,\hat{\phi}(x)\hat{\phi}(x')\,|0\rangle_M = \frac{1}{(x^0 - x'^0 - i\epsilon)^2 - (\mathbf{x} - \mathbf{x}')^2} \\ &= \frac{-a^2}{\mathrm{e}^{2a\xi} + \mathrm{e}^{2a\xi'} + a^2\epsilon^2 - 2\mathrm{e}^{a(\xi+\xi')}\cosh a(\tau - \tau') + 2ai\epsilon\,(\mathrm{e}^{a\xi}\sinh a\tau + \mathrm{e}^{a\xi'}\sinh a\tau')} \\ &= \langle 0|_R \,\hat{\phi}(x)\hat{\phi}(x')\,|0\rangle_R \\ &= G_+^R(x,x') \,. \end{aligned} \tag{6.47}$$

From this expression is is clear that in Rindler coordinates, G_+^R is periodic in complex τ with period $\beta = \frac{2\pi}{a}$. Thus, for $\mathcal{O}_a$, the Minkowski propagator is a thermal Green's function of temperature T_{Unruh}.

As before, we can make the previous remark more explicit by restricting to the accelerating world-line $x = x(\tau), x' = x(\tau = 0)$ and introducing a particle detector. Restricting to the world-line we obtain the Green's function

$$\begin{aligned} G_+^R(x(\tau), x(0)) &= \frac{a^2}{(\sinh a\tau - ia\epsilon)^2 - (\cosh a\tau - 1)^2} \\ &= \frac{-a^2}{2(1 - \cosh a\tau + 2ia\epsilon \sinh a\tau + a^2\epsilon^2/2)} \,. \end{aligned} \tag{6.48}$$

With this expression at hand we can compute the detector response function $f(E)$ characterising the detection of field transitions at energy E as

$$\begin{aligned} f(E) &= \int_{-\infty}^{\infty} \mathrm{d}\Delta\tau\, \mathrm{e}^{-iE\Delta\tau}\, G_+^R(\Delta\tau) \\ &= \int_{-\infty}^{\infty} \mathrm{d}\Delta\tau\, \frac{-a^2 \mathrm{e}^{-\mathrm{i}E\Delta\tau}}{2(1-\cosh a\tau + ia\epsilon \sinh a\tau)}\,, \end{aligned} \tag{6.49}$$

The integral can be analytically continued so as to be evaluated by residues along the negative imaginary axis where it has double poles at $a\tau = 2n\pi i$. We obtain the sum of residues

$$f(E) = E\sum_{n=0}^{\infty} \mathrm{e}^{-\frac{2\pi E n}{a}} = \frac{E}{\mathrm{e}^{\frac{2\pi E}{a}} - 1}\,, \tag{6.50}$$

as expected for a detector immersed in a thermal bath at the Unruh temperature T_{Unruh}.

Before moving on let us introduce an alternative way to detect the thermal nature of space-times. In terms of the Euclidean continuation of the space-time, it turns out thermal effects can be seen as the need to periodically identify the imaginary time coordinate [GH77a, GH94]. It is easy to see that the inverse is true. In a space-time which is periodic in imaginary time on can compute the Euclidean Green's function which consequentially will also be periodic. After analytically continuing to Lorentzian signature, the Euclidean Green's function becomes the Feynman Green's function which naturally inherits the complex periodicity of its Euclidean counterpart.

As an example, consider the the analytic continuation of the Rindler wedge metric,

$$\mathrm{d}s^2 = -a^2(\mathrm{d}\rho^2 + \rho^2 \mathrm{d}\theta^2)\,, \tag{6.51}$$

where we defined the coordinate $\rho = \mathrm{e}^{a\xi}$ and $\theta = \frac{i\tau}{a}$. This is usual metric on flat Euclidean space but generically, it has a conical singularity at $\rho = 0$. In order to avoid this singularity we need to periodically identify θ with period 2π. This is essential for regularity at the horizon and gives rise to the imaginary periodicity $\tau \sim \tau + i\beta$ for $\beta = 1/T_{\mathrm{Unruh}}$. This is a theme that can be taken much further in curved space-times where similar consideration prove very useful in studying black hole backgrounds.

Remark. Let us finish this chapter with some apparent paradoxes, and their resolutions, in Rindler space. First, note that a Rindler observer with smaller constant ξ coordinate are accelerating faster to keep up. This may seem surprising because in Newtonian physics, observers who maintain constant relative distance must share the same acceleration. In relativistic physics, this is no longer true and we see that the trailing endpoint of a rod which is accelerated by some external force (parallel to its symmetry axis) must accelerate a bit faster than the leading endpoint, or else it must ultimately break. This is a manifestation of Lorentz contraction. As the rod accelerates, its velocity increases and its length decreases. Since it is getting shorter, the back end must accelerate harder than the front. Another way to look at it is: the back end must achieve the same change in velocity in a shorter period of time. This leads to a differential equation showing that, at some distance, the acceleration of the trailing end diverges, resulting in the Rindler horizon. This phenomenon is the basis of a well known "paradox", Bell's spaceship paradox. However, it is a simple consequence of relativistic kinematics. One way to see this is to observe that the magnitude of the acceleration vector is just the path curvature of the corresponding world line. But the world lines of our Rindler observers are the analogues of a family of concentric circles in the Euclidean plane, so we are simply dealing with the Lorentzian analogue of a fact familiar to speed skaters: in a family of concentric circles, inner circles must bend faster (per unit arc length) than the outer ones.

The main observation of this chapter was that an accelerated observer detects particles in the Minkowski vacuum state. An inertial observer would say that the same state is completely empty, the expectation value of the energy momentum tensor $\left\langle T_{\mu\nu}\right\rangle_M = 0$. If there is no energy momentum how can the Rindler observer detect particles? If the Rindler observer is to detect background particles, they must carry a detector. This must be coupled to the particle being detected. However, if a detector is being maintained at constant acceleration, energy is not conserved. From the point of view of the Minkowski observer the Rindler detector emits as well as absorbs particles, once the coupling is introduced the possibility of emission is unavoidable. When the detector registers a particle the inertial observer would say that it had emitted a particle and felt a radiation-reaction force in response. Ultimately the energy needed to excite the Rindler detector does not come from the background energy momentum tensor but from the energy we put into the detector to keep it accelerating.

Chapter 7

Hawking radiation

The creation of particles by black holes is necessary for maintaining the second law of thermodynamics in their presence. This process of radiation and evaporation of black holes is an important facet in the fundamental search for a microscopic explanation of the entropy of black holes; a search which appears to be leading to new and exciting physics connecting gravitation and quantum theory. In this chapter we will explore the effect of Hawking radiation and introduce some of the challenges this phenomenon generates.

7.1 Quantum fields in a black hole background

The goal of this section is to explore quantum fields in a black hole background. The simplest, and prototypical example of such a background is the Schwarzschild background, with metric,

$$\mathrm{d}s^2 = \left(1-\frac{2M}{r}\right)\mathrm{d}t^2 - \left(1-\frac{2M}{r}\right)^{-1}\mathrm{d}r^2 - r^2\mathrm{d}\Omega^2\,. \tag{7.1}$$

Black holes, such as the Schwarzschild black hole and its rotation or charged cousins were discussed at length in the course general relativity 2. In Appendix C we collect all the necessary background information for these notes to be self-contained but we refer the reader to the lecture notes of general relativity 2 for more information and references.

One might be surprised that anything interesting can happen since the Schwarzschild black hole is a static space-time. Surely one can use the Schwarzschild time-like Killing vector (at least at large distances) to define positive and negative frequency and proceed with the quantisation just like in Minkowski space. The point of this chapter is to show that interesting things do happen! We will do so in steps which will become increasingly realistic at the expense of a loss of explicit solvability.

$(1+1)$-dimensional toy model

We start our exploration with a massless scalar field in a two-dimensional "black hole" background with the same time-radial part of the metric as the Schwarzschild black hole

$$\begin{aligned}\mathrm{d}s^2 &= \left(1-\frac{2M}{r}\right)\mathrm{d}t^2 - \left(1-\frac{2M}{r}\right)^{-1}\mathrm{d}r^2 \\ &= \left(1-\frac{2M}{r}\right)(\mathrm{d}t^2 - \mathrm{d}r_*^2) \\ &= \frac{2M}{r}\mathrm{e}^{-\frac{r}{2M}}\,\mathrm{d}U\mathrm{d}V\,.\end{aligned} \tag{7.2}$$

where we introduced the tortoise coordinate $r_* = r + 2M \log\left(\frac{r}{2M} - 1\right)$ and the Kruskal-Szekeres and Eddington-Finkelstein null coordinates U, V and u, v are defined as,

$$U = -4M\mathrm{e}^{-\frac{u}{4M}}\,, \qquad V = 4M\mathrm{e}^{-\frac{v}{4M}}\,, \qquad u = t + r^*\,, \qquad v = t - r^*\,,. \tag{7.3}$$

See Appendix C for more details on the relevant coordinates and their properties. Note that this $(1+1)$-dimensional model merely serves to illustrate some properties of the $(3+1)$-dimensional black hole and should not be taken seriously on its own. In itself it is not even a solution to the (vacuum) Einstein equation.

A first hint that this picture has something to do with the Unruh effect can already be seen from the coordinate change (7.3), which is identical to the transformation between null coordinates in Rindler and Minkowski space upon substituting $a = (4M)^{-1}$. Indeed, the problem is very similar to the one in Rindler space. The two coordinate systems we consider, Eddington-Finkelstein and Kruskal-Szekeres are respectively very similar to the Rindler and Minkowski coordinates.

Consider for simplicity the massless minimally coupled scalar. We could introduce mass and a coupling to the Ricci scalar but note that in the four-dimensional the Ricci scalar vanishes so we want consider it. The addition of mass breaks conformality so for the sake of keeping things simple we will not include it here but comment on it later. The Eddington-Finkelstein coordinates are adapted to an observer sitting very far from the black hole, where the metric approaches Minkowski space $\mathrm{d}s^2 \to \mathrm{d}u\mathrm{d}v$. In these coordinates it's straightforward to solve the wave equation and find a complete set of incoming and outgoing modes

$$\psi_\omega - \frac{1}{\sqrt{4\pi\omega}}\mathrm{e}^{-\mathrm{i}\omega u}\,, \qquad \tilde{\phi}_\omega = \frac{1}{\sqrt{4\pi\omega}}\mathrm{e}_{-\mathrm{i}\omega v}\,. \tag{7.4}$$

To these modes we can associate a vacuum called the Boulware vacuum which is defined by

$$b_\omega \left|0\right\rangle_B = 0\,. \tag{7.5}$$

The Boulware vacuum contains no particles from the point of view of a distant observer. However, since the Eddington-Finkelstein coordinates do not cover the whole of space-time, similar to the Rindler coordinates, this vacuum is similar to the Rindler vacuum of an accelerated observer.

Similarly, in Kruskal-Szekeres coordinates we can solve the wave equation finding the following set of positive frequency, incoming and outgoing, modes,

$$\xi_\omega(U) = \frac{1}{\sqrt{4\pi\omega}}\mathrm{e}^{-\mathrm{i}\omega U}\,, \qquad \tilde{\xi}_\omega(V) = \frac{1}{\sqrt{4\pi\omega}}\mathrm{e}^{-\mathrm{i}\omega V}\,. \tag{7.6}$$

The corresponding Kruskal vacuum is defined as,

$$a_\omega \left|0\right\rangle_K = 0\,. \tag{7.7}$$

In Kruskal-Szekeres coordinates, the metric near the black hole horizon approaches $\mathrm{d}s^2 \to \mathrm{d}U\mathrm{d}V$, so the Kruskal vacuum is the appropriate one for an observer sitting next to the black hole horizon. Since Kruskal–Szekeres coordinates cover the whole of space-time, they are the analogue of the Minkowski vacuum that we studied in the quantization of a scalar field in Rindler space.

We can now ask the following question: if a Kruskal observer is in the vacuum state, what does the Boulware observer see? Since the relation between both systems is the same as we found before in the case of the Unruh effect, the calculation of the Bogoliubov coefficients will be the identical as the one for the Unruh effect. The only difference lies in replacing the acceleration a by the surface gravity $a \to \kappa = \frac{1}{4M}$. We conclude that the Boulware observer sees a thermal spectrum with temperature

$$T_H = \frac{\kappa}{2\pi} = \frac{1}{8\pi M} \,. \tag{7.8}$$

Far from the black hole, the factor g_{00} in Eddington–Filkenstein coordinates goes to 1, and $T_0 = T$, so (7.8) is the physical temperature observed by an observer at infinity.

$(3+1)$-dimensional Schwarzschild

The two-dimensional toy model discussed above was extremely simple but included all the necessary ingredients to observe the Hawking temperature. However, we are really interested in the four-dimensional Schwarzschild black hole. In this case we lose conformality and we will not be able to exactly solve the problem. However, using some approximations we will still be able to come to a similar conclusion as in the toy model above.

Let us consider again a massless scalar field but now in the full Schwarzschild background (7.1). As in Schwarzschild, and similarly in the Kerr black hole we have $R = 0$ we can safely ignore the coupling to the Ricci scalar.[1] As the problem is entirely symmetric on the two-sphere it will prove useful to decompose our field in spherical harmonics,

$$\phi = f_{lm}(t,r) Y_{lm}(\theta,\phi) \,, \tag{7.9}$$

where $Y_{lm}(\theta,\phi)$ are the spherical harmonics. Substituting this expansion in the wave equation results in

$$\Box^{(4)}\phi = 0 \Rightarrow \left(\Box^{(2)} + V_l(r)\right) f_{lm}(r,t) \,, \tag{7.10}$$

where $\Box^{(d+1)}$ denotes the $(d+1)$-dimensional Laplacian on respectively the full Schwarzschild space-time or the 2d time-radial slice considered in the above. The potential $V_l(r)$ is given by

$$V_l(r) = \left(1 - \frac{2M}{r}\right)\left(\frac{2M}{r^3} + \frac{l(l+1)}{r^2}\right) . \tag{7.11}$$

So we see that the massless scalar in $(3+1)$ dimensions decomposes in infinitely many massless scalars in $(1+1)$ dimensions in the presence of a potential. The only change from the story above is therefore that a wave escaping the black hole needs to propagate through the potential barrier caused by $V_l(r)$. Even though we cannot solve this problem analytically, note that the potential falls off exponentially in r^* as $r^* \to -\infty$, i.e. when one approaches the horizon, and falls off polynomially as $r \to \infty$. For this reason we can use the same asymptotic states as above. Hence, the only effect of the potential is that it decreases the intensity of the wave and changes the resulting spectrum of

[1] In the collapse picture below, in the collapsing phase the Ricci scalar might be non-vanishing. However, this will not change the late time spectrum and so we will keep ignoring this coupling.

emitted particles by a greybody factor $0 < \Gamma_l(\omega) < 1$,

$$\langle n_\omega \rangle = \frac{\Gamma_l(\omega)}{e^{\frac{\omega}{T_H}} - 1}\,. \tag{7.12}$$

The greybody factor is entirely due to the potential outside the black hole horizon. It is clear that this factor is not directly related to the quantum origin of the Hawking radiation and therefore the basic features of the derivation above survive without significant alterations. This result can be generalized for the case of a massive scalar field, and also for vector and spinor fields. The conclusion is that the black hole must emit all possible species of particles, each having the Hawking thermal spectrum corrected by the corresponding greybody factor.

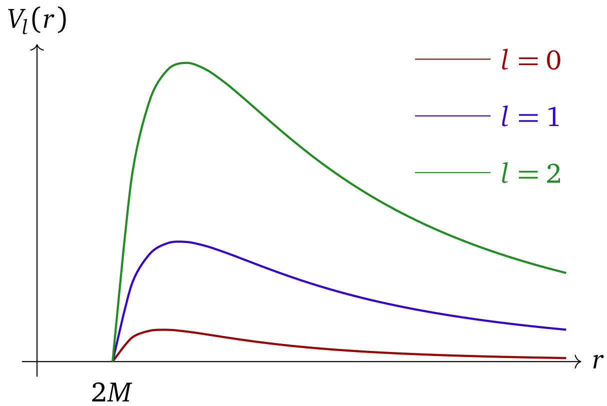


Figure 7.1: The effective potential $V_l(r)$ experienced by the spherically symmetric modes f_{lm} for the values of $l = 0, 1, 2$.

Black hole formed through collapse

The eternal black hole described above is rather unphysical and we don't expect to see the full Kruskal extension. A more physical picture would be to consider a ball of spherically symmetric dust collapsing to form a black hole. The Penrose diagram for this space-time is given in Figure 7.2.

We can now proceed as before and quantise the massless scalar field in this background. The past null hypersurface $\mathscr{I}_-$ is a Cauchy hypersurface, hence we can quantise the scalar field using this hypersurface and write

$$\phi = \int d\omega \left(a_\omega f_\omega + a^\dagger_\omega f^*_\omega \right)\,, \tag{7.13}$$

where the f_ω are a complete set of orthonormal solutions to the wave equation with associated annihilation and creation operators a_ω and $a^\dagger_\omega$. Far outside the collapsing body at early times, the definition of physical particles that would be detected by inertial observers, or equivalently of positive frequency solutions of the wave equation, is unambiguous. We choose the f_ω such that they form a complete set of incoming positive frequency solutions of energy ω. Their asymptotic form on past null infinity is

$$f_\omega \sim \frac{1}{\sqrt{16\pi^3\omega}} e^{-i\omega v} Y_{lm}(\theta, \phi)\,, \qquad \langle f_\omega, f_{\omega'} \rangle = \delta(\omega - \omega')\,, \tag{7.14}$$

where we suppress the discrete quantum numbers l and m in labelling the functions f_ω.

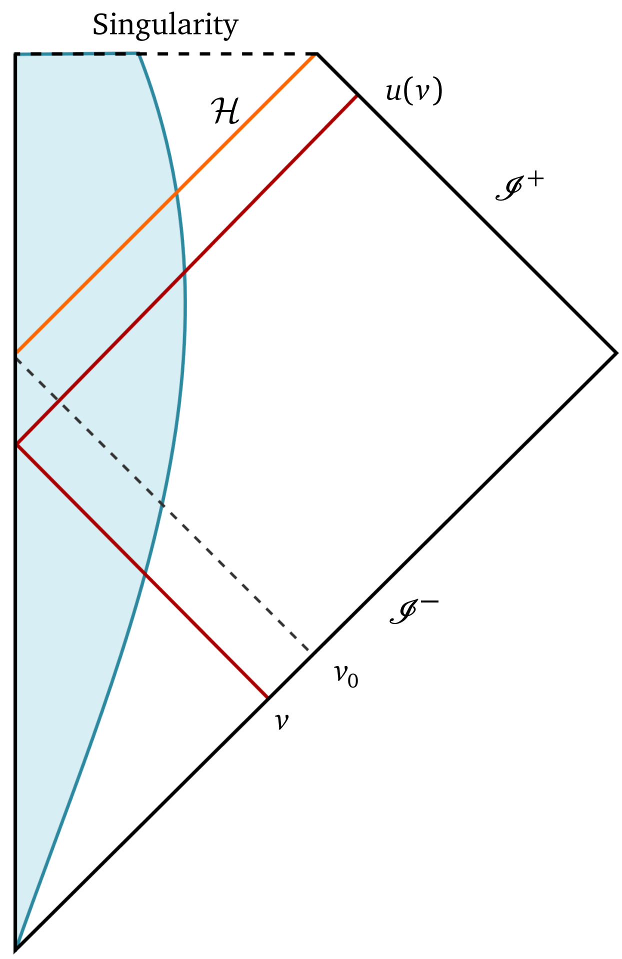


Figure 7.2: The Penrose diagram for collapse to the Schwarzschild black hole. The singularity is located on top and shielded by the horizon $\mathcal{H}$ in orange. The collapsing cloud of dust is pictured in blue. Once the cloud enters the horizon, a black hole is formed. The incoming ray with $v = v_0$ is the last one that reaches the centre of the collapsing body and makes it to $\mathscr{I}^+$. Rays with $v < v_0$ fall into the black hole.

At late times on the other hand, we know that $\mathscr{I}^+$ is not a Cauchy hypersurface. Instead we have to consider boundary data both at future null infinity and the event horizon $\mathcal{H}$. On $\mathscr{I}^+$, just like on $\mathscr{I}^-$, the definition of positive frequency modes is unambiguous and we can find a complete set $\{p_\omega, p_\omega^*\}$ of orthonormal solutions on $\mathscr{I}^+$. The asymptotic form of these functions on $\mathscr{I}^+$ is

$$p_\omega \sim \frac{1}{\sqrt{16\pi^3\omega}} \mathrm{e}^{-\mathrm{i}\omega u} Y_{lm}(\theta, \phi), \tag{7.15}$$

where u is the outgoing null coordinate at $\mathscr{I}^+$. A general solution, incoming from the past, will also have a part that is incoming at the event horizon. Therefore we must introduce a second complete basis of orthonormal functions q_ω on the horizon which have zero Cauchy data on $\mathscr{I}^+$. Since the functions p_ω and q_ω are supported in disjoint regions at late times, their (conserved) scalar product must vanish $\langle q_\omega, p_{\omega'} \rangle = 0$ and similarly for their complex conjugates. For this reason the precise form of the functions q_ω will not affect observations on $\mathscr{I}^+$. The details are therefore not important since we will trace over the modes at the horizon. We can thus expand the field ϕ in the entire space-time as

$$\phi = \mathrm{d}\omega \left(b_\omega p_\omega + c_\omega q_\omega + b_\omega^\dagger p_\omega^* + c_\omega^\dagger q_\omega^* \right) , \tag{7.16}$$

with b_ω and c_ω the annihilation operators for outgoing particles at late times. The vacuum at $\mathscr{I}_+$ defined by $b_\omega |0\rangle_B$ is the Boulware vacua as defined before, while the vacuum at past null infinity $|0\rangle_-$, defined by $a_\omega |0\rangle_-$ will take up the role of the Kruskal vacuum. The task we have to do is then clear, we want to compute the number density of particles observed by a Boulware observer in the "Kruskal" vacuum. Although conceptually clear the computation is rather involved and some details will be left to fill in by the reader.

To compute the density of emitted particles we have to compute the Bogoliubov coefficients

$$\alpha_{\omega\omega'} = \langle f_{\omega'}, p_\omega \rangle \,, \qquad \beta_{\omega\omega'} = -\left\langle f_{\omega'}^*, p_\omega \right\rangle . \tag{7.17}$$

To determine these coefficients we need to trace back in time the function p_ω along an outgoing geodesic have a large value of u, close to the horizon. Such a geodesic is illustrated in Figure 7.2 as the red line and passes through the center of the collapsing cloud just before the event horizon is formed and emerges as an incoming geodesic characterised by a value of v close to v_0. The value of u depending at $\mathscr{I}^+$ depending on v can be computed by analysing the null geodesics in this space-time, see Appendix C, and is given by

$$u(v) = -4M \log \left(\frac{v_0 - v}{K} \right) , \tag{7.18}$$

where K is some positive constant. Inserting this expression into expressions (7.15) for the functions p_ω we can compute the Bogoliubov coefficients as

$$\begin{aligned} \alpha_{\omega\omega'} &= C \int_{-\infty}^{v_0} \mathrm{d}v \left(\frac{\omega'}{\omega} \right)^{\frac{1}{2}} \mathrm{e}^{\mathrm{i}\omega' v - \mathrm{i}\omega u(v)} , \\ \alpha_{\omega\omega'} &= C \int_{-\infty}^{v_0} \mathrm{d}v \left(\frac{\omega'}{\omega} \right)^{\frac{1}{2}} \mathrm{e}^{-\mathrm{i}\omega' v - \mathrm{i}\omega u(v)} , \end{aligned} \tag{7.19}$$

where C is a constant. Substituting $s = v_0 - v = \mathrm{i}z$ we can compute

$$\begin{aligned} \alpha_{\omega\omega'} &= -C \int_{\infty}^{0} \mathrm{d}s \left(\frac{\omega'}{\omega} \right)^{\frac{1}{2}} \mathrm{e}^{-\mathrm{i}\omega'(s-v_0)} \exp\left(4\mathrm{i}\omega M \log \frac{s}{K} \right) \\ &= -\mathrm{i}C\mathrm{e}^{\mathrm{i}\omega' v_0} \int_{-\infty}^{0} \mathrm{d}z \left(\frac{\omega'}{\omega} \right)^{\frac{1}{2}} \mathrm{e}^{\omega' z} \exp\left(4\mathrm{i}\omega M \log \frac{\mathrm{i}z}{K} \right) \\ &= -\mathrm{i}C\mathrm{e}^{\mathrm{i}\omega' v_0} \mathrm{e}^{2\pi\omega M} \left(\frac{\omega'}{\omega} \right)^{\frac{1}{2}} \int_{-\infty}^{0} \mathrm{d}z \, \mathrm{e}^{\omega' z} \exp\left(4\mathrm{i}\omega M \log \frac{|z|}{K} \right) , \end{aligned} \tag{7.20}$$

and similarly,

$$\beta_{\omega\omega'} = \mathrm{i}C\mathrm{e}^{-\mathrm{i}\omega' v_0} \mathrm{e}^{-2\pi\omega M} \left(\frac{\omega'}{\omega} \right)^{\frac{1}{2}} \int_{-\infty}^{0} \mathrm{d}z \, \mathrm{e}^{\omega' z} \exp\left(4\mathrm{i}\omega M \log \frac{|z|}{K} \right) . \tag{7.21}$$

We immediately find that

$$|\alpha_{\omega\omega'}|^2 = \mathrm{e}^{8\pi M\omega} |\beta_{\omega\omega'}|^2 , \tag{7.22}$$

for the part of the wave packet that was propagated back in time through the collapsing body just before it formed a black hole.

For the components p_ω of this part of the wave packet, we have the scalar product,

$$\langle p_\omega, p_{\omega'} \rangle = \Gamma(\omega)\delta(\omega - \omega'), \tag{7.23}$$

where $\Gamma(\omega)$ is the fraction of the wave packet that would propagate back in time through the collapsing body. Indeed, we can divide the functions p_ω in two parts,

$$p_\omega = p_\omega^{(1)} + p_\omega^{(2)}. \tag{7.24}$$

The part $p_\omega^{(1)}$ propagates backwards in time outside of the collapsing body and reaches $\mathscr{I}_-$ at some value $v > v_0$. This part of the wave will interact minimally with the collapsing matter and consequentially the frequency will not change significantly from $\mathscr{I}_-$ to $\mathscr{I}_+$. For this reason we can ignore this part of the wave when asking questions about particle production. Indeed, since $p_\omega^{(1)}$ and $p_\omega^{(2)}$ have disjoint support on $\mathscr{I}_-$ (resp. $v > v_0$ and $v < v_0$), they do not interact and we can safely ignore the parts $p_\omega^{(1)}$. Their only effect is the introduction of the function Γ. From the normalisation condition (7.23) we therefore find

$$\Gamma(\omega)\delta(\omega - \omega') = \int \mathrm{d}\lambda \left(\alpha^*_{\omega\lambda}\alpha_{\omega'\lambda} - \beta^*_{\omega\lambda}\beta_{\omega'\lambda} \right) \tag{7.25}$$

where now the Bogoliubov coefficients refer to the coefficients in the expansion of $p_\omega^{(2)}$ only.

As before, this allows us to compute the density of emitted particles as

$$\langle N_\omega \rangle_- = \langle 0|_K \, b^\dagger_\omega b_\omega \, |0\rangle_- \simeq \langle 0|_- \, b^{(2)\dagger}_\omega b^{(2)}_\omega \, |0\rangle_- = \int \mathrm{d}\omega' |\beta_{\omega\omega'}|^2 . \tag{7.26}$$

The resulting integral is again divergent but can be regularised by putting the system in a box and computing the density of emitted particles instead,

$$n_\omega = \frac{1}{V} \langle N_\omega \rangle_- = \frac{\Gamma(\omega)}{2\pi \left(\mathrm{e}^{8\pi M\omega} - 1 \right)} . \tag{7.27}$$

Hence after this long computation we come to exactly the same conclusion as before and find that the collapsing black hole emits and absorbs radiation exactly like a gray body of absorptivity $\Gamma_{lm}(\omega)$ and Hawking temperature $T_H = (8\pi M)^{-1}$!

For large black holes this temperature $T_H \sim 6 \times 10^{-8} \frac{M}{M_\odot} K$ is extremely small for large black holes with $M \gg M_\odot$, where $M_\odot$ is the mass of the sun. For this reason our assumption that the background does not back-react against this radiation seems to be justified. For small black holes on the back-reaction cannot be ignored and a more sophisticated treatment is needed.

Indeed, from energy conservation one can estimate the rate of loss of mass. Stefan's law for the evaporation of a black body states that

$$\frac{dE}{dt} \propto -AT^4 , \tag{7.28}$$

where A is the area. With $E \propto M^2$ and $T \propto M^{-1}$ this leads to the rate of mass loss to be proportional to

$$\frac{\mathrm{d}M}{\mathrm{d}t} \simeq -\frac{c}{M^2}\,, \tag{7.29}$$

where c is a positive constant that depends on the number and type of quantised matter fields that couple to gravity. From this expression we it becomes indeed apparent that for large black holes $M \gg \frac{\mathrm{d}M}{\mathrm{d}t}$ justifying our assumption of ignoring back-reaction. For reference, this leads to the black-hole evaporating in a finite time of the order of $10^{71}\left(\frac{M}{M_\odot}\right)^3$ seconds.

Finally, before moving on, note that a static observer $\mathcal{O}$ at finite radius r measures a blue-shifted temperature $T_{\mathcal{O}} = \frac{T_H}{|g_{00}|}$. As $r \to \infty$ this approaches the Hawking temperature, but it diverges at the horizon where $|g_{00}| \to 0$ due to the infinite acceleration of the static observer at the horizon. This is precisely the Unruh effect we observed in the previous section. A freely falling observer however sees no divergence as they cross the horizon.

7.2 The Hawking thermal state and friends

In analogy with the Rindler case, we can easily observe the thermal nature of the $|0\rangle_-$ vacuum for a Boulware observer. The observer in the Boulware vacuum $|0\rangle_B$ has access to a Fock space $\mathcal{F}_B$ built by acting on $|0\rangle_B$ with the creation operators $b^\dagger_\omega$. As in the Rindler case however, and from the fact that $\mathscr{I}^+$ is not a Cauchy surface, we know that this is not enough to construct the full Fock space as seen by an observer at past null infinity. Indeed, the full Fock space is obtained as the tensor product $\mathcal{F}_- = \mathcal{F}_B \otimes \mathcal{F}_H$.

As before for Rindler, the late time vacuum is a complicated state of the form

$$|0\rangle_- \propto \sum_n f_n \, |n\rangle_H \, |n\rangle_B \,, \tag{7.30}$$

Considering the associated density matrices, we construct the Boulware vacuum by tracing over the horizon modes

$$\rho_B = \mathrm{Tr}_{\mathcal{F}_H} \rho_- = \mathrm{Tr}_{\mathcal{F}_H} |0\rangle_- \langle 0|_- \propto \sum_n \mathrm{e}^{-n\pi\Omega/\kappa} |n\rangle_{BB}\langle n| \tag{7.31}$$

which gives the desired thermal state.

The Hawking state was what arose from an essentially Minkowskian vacuum at $\mathscr{I}^-$ in the collapsing scenario, but other states are natural for the eternal Schwarzschild black hole where we start from the Kruskal vacuum defined near the horizon. In this case the Cauchy surface in the far past consists of two components, $\mathscr{I}_- \cup \mathcal{H}_-$, as can be seen from the Penrose diagram in Figure E.1. Having said so, it becomes clear that there are various 'natural' choices for the vacuum in the past depending on what we define as positive frequency states. The options are summarised in Table 7.1 below for positive frequencies $\omega > 0$.

The three options defined in this table each have a distinct physical interpretation.

- The Boulware vacuum corresponds to our familiar concept of an empty state defined far away from the black hole and is defined with respect to a static observer. It is pathologic in the sense

Vacuum	Positive modes on $\mathcal{H}_-$	Positive modes on $\mathscr{I}_-$
Boulware vacuum $\lvert B\rangle$	$e^{-i\omega u}$	$e^{-i\omega v}$
Unruh vacuum $\lvert U\rangle$	$e^{-i\omega U}$	$e^{-i\omega v}$
Hartle-Hawking vacuum $\lvert H^2\rangle$	$e^{-i\omega U}$	$e^{-i\omega V}$

Table 7.1: The three natural vacua in the eternal Schwarzschild black hole. The options differ by the choice of positive frequency modes in the two component of the far past Cauchy surface $\mathcal{H} \cup \mathscr{I}_-$.

that the expectation value of the stress tensor diverges at the horizon. This is similar to the Rindler vacua becoming singular at the Killing horizon seen in the previous chapter.

- The Unruh vacuum is regular on the future horizon, but not on the past horizon. At infinity, this vacuum corresponds to an outgoing flux of blackbody radiation at the black hole temperature. The black hole collapse studied in the previous section brings about the Unruh state.

- The Hartle-Hawking vacuum does not correspond to our usual notion of a vacuum. It is well-behaved both on the future and past horizon but the price we have to pay for this is that the state is not empty at infinity, but instead corresponds to a thermal distribution of quanta at the Hawking temperature. That is, the Hartle-Hawking vacuum corresponds to a black hole in (unstable) equilibrium with an infinite bath of blackbody radiation.

All these vacua are interesting in their own right and have been studied for a variety of reasons. However, from the point of view of the 'physical' collapse picture described above, it seems that the Unruh vacuum best approximates the state obtained following the gravitational collapse of a massive cloud of dust.

There are various ways to investigate the thermal nature of the various vacua above. As mentioned above, the Hartle-Hawking state is a thermal state both at $\mathscr{I}_\pm$. Moreover, it is an example of a Thermal Green's function

$$G^{H^2}(x,x') := \langle H^2 \rvert \, \hat{\phi}(x)\hat{\phi}(x') \, \lvert H^2 \rangle = G_\beta(x,x') \,. \tag{7.32}$$

We will not attempt to explicitly compute the Green's function (see for example [CJ86] for the result) but a key statement is that it analytically extends to complex time and is periodic in imaginary time with period,

$$\beta = 1/T_H = 8\pi M \,, \tag{7.33}$$

An alternative, and easier way to recognize the thermal nature of black holes is to study the Euclideanised background. Wick rotating $t \to i\tau$, we find the positive definite background with metric[2]

$$ds^2 = \left(1 - \frac{2M}{r}\right) d\tau^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 \,. \tag{7.34}$$

[2]It's even easier to see this in Kruskal coordinates, where the metric is given by $ds^2 = \frac{2M}{r} e^{-\frac{r}{2M}} dU dV$. Remembering that $U = -4M e^{\frac{r^*}{4M}} e^{-\frac{i\tau}{4M}}$ it is equally clear that unless τ has period $8\pi M$ this metric has a conical singularity at the origin.

Substituting $r = 2M + \epsilon$ and expanding in small ϵ we find

$$ds^2 \approx \frac{2M}{\epsilon} d\epsilon^2 + \frac{\epsilon}{2M} d\tau^2 + 4M^2 d\Omega^2 . \tag{7.35}$$

At $\epsilon = 0$ we see that the angular part nicely factorises out. Changing coordinates to $\rho = \sqrt{8M\epsilon}$ the metric becomes

$$ds^2 \approx d\rho^2 + \frac{\rho^2}{16M^2} d\tau^2 + 4M^2 d\Omega^2 . \tag{7.36}$$

Hence we clearly see that in order to avoid a conical singularity at the origin we need to impose the periodicity $\tau \sim \tau + 8\pi M$. This gives alternative evidence for the Hawking temperature, analogous to the Euclideanisation argument for the Unruh temperature in the previous section. In Figure 7.3 we sketch the topology of the various Euclideanisations discussed so far. This property is an important red herring for the presence of thermal states and remains valid much more general in various dimensions with various types of matter content. The Euclideanisation of a black hole has the characteristic topology of a cigar.

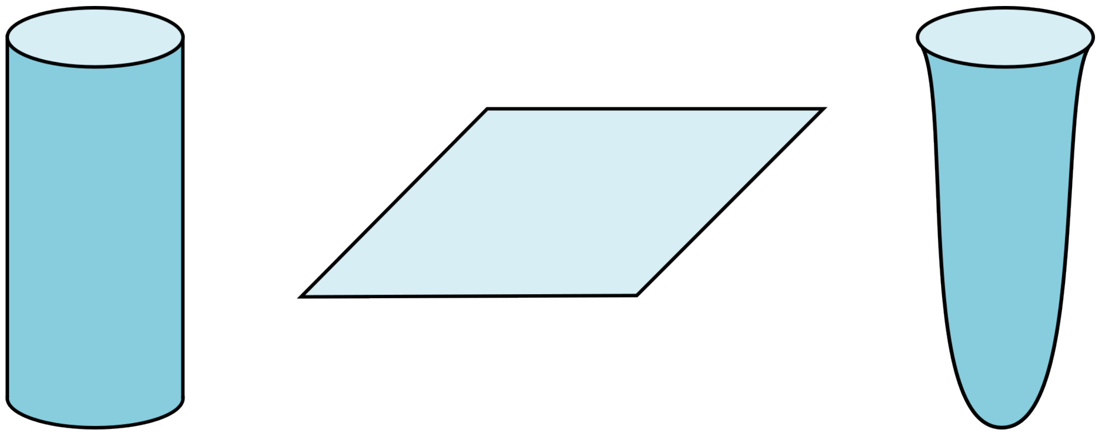


Figure 7.3: From left to right, the Euclideanised Minkowski space $\mathbb{R} \times S^1$, Rindler space $\mathbb{R}^2$ and (r, t) plane of the Schwarzschild black hole. The cigar topology background is characteristic for black hole backgrounds.

7.3 Black hole thermodynamics

Prior to the discovery of Hawking radiation of black holes Bekenstein already conjectured that black holes must have a non-vanishing intrinsic entropy [Bek73]. He came to this conclusion through the following thought experiment. Consider a black hole that absorbs matter with non-zero entropy. If the black hole entropy were vanishing then the total entropy in the system would decrease, violating the second law of thermodynamics. Based on this reasoning Bekenstein concluded that the second law can only be preserved if a black hole has an intrinsic entropy S_{BH} proportional to its surface area. However, the proportionality constant could not be fixed until the discovery of Hawking radiation.

Differentiating the expression for the surface area $A = 16\pi M^2$, we find

$$dM = \frac{1}{8\pi M} d\frac{A}{4} . \tag{7.37}$$

Recognising the coefficient on the left hand side as the Hawking temperature this looks precisely like the first law of thermodynamics

$$\mathrm{d}E = T\mathrm{d}S\,, \tag{7.38}$$

Following this analogy we conclude that the black hole (or Bekenstein-Hawking) entropy must be equal to

$$S_{BH} = \frac{A}{4} = 4\pi M^2\,. \tag{7.39}$$

In line with its thermodynamic counterpart, the first law of black hole thermodynamics can be generalised to closed systems with rotation and charge as follows,

$$\mathrm{d}E = T\mathrm{d}S + \Omega \mathrm{d}J + \Phi \mathrm{d}Q\,, \tag{7.40}$$

where we interpret Φ as the electric potential at the horizon and Q the total charge. Similarly, J is the angular momentum and Ω the angular velocity.

Exercise 7.1. *Consider the Reissner-Nordstrom solution*

$$\mathrm{d}s^2 = \frac{\Delta(r)}{r^2}\mathrm{d}t^2 - \frac{r^2}{\Delta(r)}\mathrm{d}r^2 - r^2\mathrm{d}\Omega^2\,, \qquad \Delta(r) = r^2 - 2Mr + Q^2\,, \tag{7.41}$$

This is a solution to the Einstein-Maxwell equations with electromagnetic potential

$$A = \frac{Q}{r}\mathrm{d}t\,. \tag{7.42}$$

Assuming $Q < M$, state the second law of thermodynamics by differentiating the area as a function of mass and show that the coefficient is indeed equal to the Hawking temperature. (Hint: the Hawking temperature has the same expression as for Schwarzschild when expressed in terms of the surface gravity.)

Exercise 7.2. *If you are feeling courageous, repeat the previous exercise for the Kerr black hole.*

The entropy of astrophysical black holes is extremely large, for a solar mass black hole for example one finds $S_{BH}^{\odot} \sim 10^{76}$. Interpreting this as a statistical entropy implies that a quantum mechanical black hole has an enormous number of microstates corresponding to the unique classical black hole. Finding a microscopic derivation of this entropy is an active area of modern research. In asymptotically flat space-time such a derivation has been given through string theory [SV96] but in asymptotically AdS or dS space-times this remains an open question.

Taking into account the entropy of a black hole, we can state the generalised second law of thermodynamics as follows.

$$\delta S_{\text{total}} = \delta S_{\text{matter}} + \delta S_{BH} \geq 0\,. \tag{7.43}$$

I.e. the total entropy of all black holes and matter combined can never decrease. In classical general relativity, one can prove that the combined area of all black hole horizons cannot decrease. This applies not only to adiabatic processes but also to strongly out of equilibrium processes such as collisions and mergers of black holes.

Ordinary thermodynamic systems can be in a stable equilibrium with an infinity heat reservoir. However, this is not true for black holes because they have a negative heat capacity! In other words, black holes get colder when they absorb energy. Indeed, with $E(T) = M = (8\pi T)^{-1}$, we find

$$C_{BH} = \frac{\partial E}{\partial T} = -\frac{1}{8\pi T^2} < 0\,. \tag{7.44}$$

This means that a black hole surrounded by an infinite thermal bath at temperature $T < T_H$ will emit radiation and become even hotter. The process of evaporation is not halted in an infinite thermal reservoir with constant temperature. Similarly, putting a black hole in a bath with $T > T_H$ will make the black hole colder! In either case no stable equilibrium is possible. Stable equilibrium is only possible in a finite reservoir. In this case the radiation of the black hole changes the temperature of the bath until both reach the same temperature.

7.4 The information paradox

In this chapter we have discussed quantum fields in a black hole background and discovered that black holes have a temperature. But where precisely does this radiation come from? The answer, discovered by Hawking, is that we must consider quantum processes, more precisely quantum fluctuations of the vacuum. In the vacuum pairs of particles and antiparticles are continuously being created and annihilated. Consider such fluctuations for electron-positron pairs. Suppose we apply a strong electric field in a region which is pure vacuum. When an electron–positron pair is created, the electron gets pulled one way by the field and the positron gets pulled the other way. Thus instead of annihilation of the pair, we can get creation of real (instead of virtual) electrons and positrons which can be collected on opposite ends of the vacuum region. Thus we get a current flowing through the space even though there is no material medium filling the region where the electric field is applied. This is called the 'Schwinger effect'.

A similar effect happens with the black hole, with the effect of the electric field now replaced by the gravitational field. We do not have particles that are charged in opposite ways under gravity. But the attraction of the black hole falls off with radius, so if one member of a particle–antiparticle pair is just outside the horizon it can flow off to infinity, while if the other member of the pair is just inside the horizon then it can get sucked into the hole. The particles flowing off to infinity represent the 'Hawking radiation' coming out of the black hole. Doing a detailed computation, one finds that the rate of this radiation is given by (7.28). Thus we seem to have a very nice thermodynamic physics of the black hole. The hole has entropy, energy, and temperature and radiates as a thermal body should.

So far, so good, but there is a deep problem arising out of the way in which this radiation is created by the black hole [Haw76]. The radiation which emerges from the hole is not in a 'pure quantum state'. Instead, the emitted quanta are in a 'mixed state' with excitations which stay inside the hole. There is nothing wrong with this in this by itself, but the problem comes at the next step. The black hole loses mass because of the radiation and eventually disappears. Then the quanta in the radiation outside the hole are left in a state that is 'mixed', but we cannot see anything that they are mixed with! Thus the state of the system has become a 'mixed' state in a fundamental way. This does not happen in quantum mechanics. If we start with a pure state $|\psi\rangle$ and evolve it by some Hamiltonian H to $|\psi'\rangle = e^{-iHt}\,|\psi\rangle$ we obtain another pure state at the end. Mixed states arise in usual physics when we coarse-grain over some variables and thereby discard some information about a system. This coarse-graining is done for convenience, so that we can extract the gross behaviour of a system without keeping all its fine details, and is a standard procedure in statistical mechanics. But there is always a 'fine-grained' description available with all information about the state, so that underlying the full system there is always a pure state. With black holes we seem to be getting a loss of information in a fundamental way. We are not throwing away information for convenience; rather we cannot get a pure state even if we wanted.

To make this discussion a bit more quantitative, let us introduce the von Neumann entropy, which is an extension of the Gibbs entropy from statistical mechanics to quantum statistical mechanics. For a quantum system described by a density matrix ρ, the von Neumann entropy is defined as,

$$S = -\mathrm{Tr}\, \rho \log \rho \,, \tag{7.45}$$

In a finite dimensional system we can always write the density matrix in a basis of eigenvectors $|n\rangle$ as

$$\rho = \sum_n p_n |n\rangle \langle n| \,, \tag{7.46}$$

which makes it clear that for pure state the von Neumann entropy vanishes, while its maximal value $S = \log \dim \mathcal{H}$ is reached for the maximally mixed state $\rho = \frac{1}{\dim \mathcal{H}} \sum |n\rangle \langle n|$. Now, let us consider the evaporation of a black hole à la Hawking. The black hole starts in a pure state, hence initially we have $S = 0$. After some time part of the black hole has evaporated where the radiation is in a mixed state. Hence, during the evaporation, the von Neumann entropy gradually increases until it reaches its maximum when the black hole is fully evaporated into thermal radiation. See Figure 7.4 for a graphical representation of the entropy as a function of time.

This paradox has been a guiding post for progress on quantum gravity since its discovery by Hawking in 1975. Hawking initially advocated that in the presence of gravity we should change our ideas about quantum mechanics and loosen the our demand of having purely unitary evolution. However, this is a very unsettling proposal which opens a Pandora's box of unwanted consequences and most physicists are not willing to abandon ordinary quantum mechanics when it works so well in all other contexts. Luckily, in the 90s and 2000s string theory provided various hints that information is not lost! But how can it be that we need string theory for this? The gravitational interactions at the event horizon for a large black hole are so incredibly small that we would expect that our semiclassical intuition should be valid here.

Around 2020 a new perspective emerged in the papers [Pen20, AEMM19] and many papers after that. In this paper an alternative semiclassical computation was performed that instead of Hawkings entropy curve produces the Page curve. As can be seen from Figure 7.4 this curve descends back to zero entropy at the end of the evaporation, therefore restoring unitarity! The fundamental idea behind these computations is to introduce a new tool called the 'quantum extremal surface' which takes into account the microscopic structure of the black hole as well as the coupling with the external fields. Performing the semiclassical computation using this surface, instead of the event horizon as in Hawkings computation results in a different prediction for the entropy where at the Page time a transition takes place after which the entropy starts to shrink, reproducing the Page curve. A full discussion of their formalism would lead us beyond the scope of this course so we refer the reader to

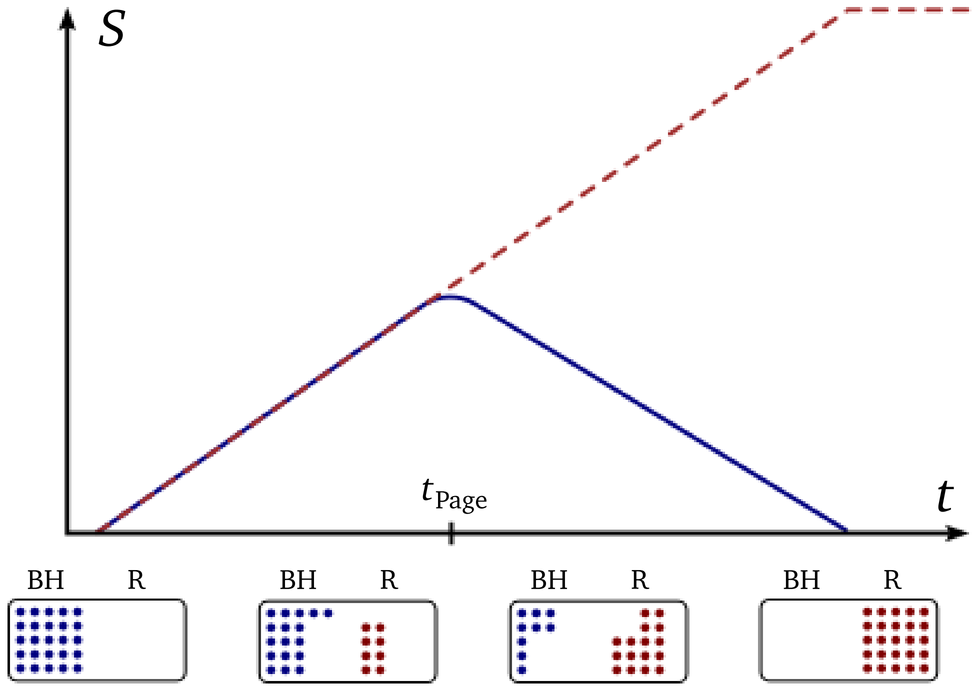


Figure 7.4: The red line represents the entropy of the radiation following Hawking's calculation, while the blue line is the page curve. The turning point at the page time occurs at the point in time where the entropy of Hawking radiation is equal to the Bekenstein-Hawking entropy of the black hole. The dots give a cartoon picture of the qubits of information transferred from the black hole to the radiation.

the original literature.

This approach immediately brings us a whole range of new questions. Why are the equations for various quantities modified by quantum gravity when a black hole is involved? And should they then also be modified when studying the sun or Mercury? The key in answering this question turns out to be complexity. Black holes are incredibly complex objects, they are maximally chaotic and pack information in the densest possible way. This characteristic sets them apart from the other astrophysical object where we find a similar curvature as at the event horizon of a black hole. The computation of the quantum extremal surface turn out to crucially depend on complexity. It turns out that usual semi-classical gravity is valid at low curvature and low complexity. However, at large complexity our semi-classical intuition has to be modified in order to predict the correct physical behaviour. Research in this direction continues until today and is an exciting area of new developments in quantum gravity.

Chapter 8

Quantum fields in de Sitter space

Our universe is expanding. Moreover, it is expanding at an accelerating pace. If this persists, we will eventually head towards a cold and lonely world. All matter will drift apart, until we are left only with our Milky Way, which by then will most likely have merged with Andromeda. After even longer time scales, all cosmic radiation will have stretched to sizes beyond the cosmic horizon, leaving us with a universe dominated by the non-diluting cosmological constant. We will end up in a de Sitter universe.

We have observational evidence for two periods of exponential expansion. The first is the early inflationary era, during which the universe dramatically expanded in the first moments after the big bang while the second is the late time era dominated by the cosmological constant. This chapter will deal with quantum fields in pure de Sitter space, while the next chapter will look at more general cosmological backgrounds. For more details and references, see [Ann12, SSV01, Har17].

8.1 de Sitter basics

The de Sitter space is the maximally symmetric spacetime of positive curvature

$$R = \frac{d(d+1)}{L^2}, \tag{8.1}$$

where L is the characteristic de Sitter length scale. This space is a solution to the vacuum Einstein equations with positive cosmological constant $\Lambda = \frac{d(d+1)}{2L^2}$. To get a better grip on the structure of the de Sitter space, a variety of coordinates systems can be used. The most common ones are summarised in Appendix E to which we refer for more details.

As we discussed in detail in Chapter 3, a lot of information about the causal structure of a spacetime is encoded in its Penrose diagram. The Penrose diagram for de Sitter space is given in Figure 8.1. Note that for dS_{d+1}, each point in this diagram is actually an S^{d-1}, except for the points on the left and right boundary, where the $(d-1)$-spheres shrink. Future and past null infinity, where all null geodesics end and start, are located at the top and bottom of the diagram and are both spacelike hypersurfaces. Moreover they, or any other horizontal slice of this diagram, are Cauchy hypersurfaces and hence, de Sitter space is clearly globally hyperbolic.

One peculiar feature of de Sitter space is that no single observer can access the entire space-time. De Sitter space describes a exponentially expanding universe and for this reason, there is an observer-dependent horizon, called the cosmological horizon beyond which space-time is expanding faster than the speed of light. This is a null hypersurface beyond which the observer can never receive or send a signal. For example, an observer sitting at the right edge of the Penrose diagram will never

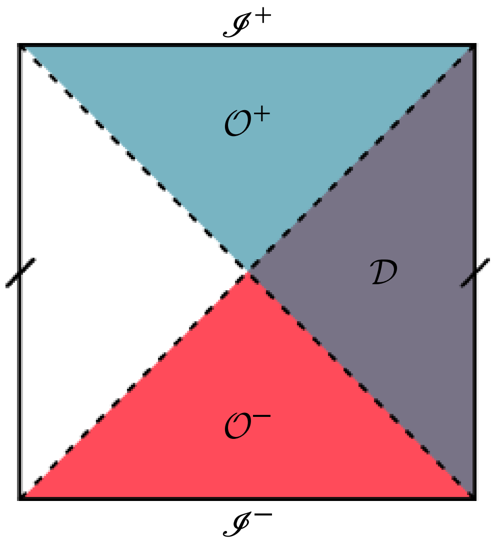


Figure 8.1: The Penrose diagram for dS_{d+1}. The left and right boundary are time-like lines which can be identified, while every point in the interior represents an S^{d-1}. The dashed lines are the (cosmological) past and future horizons for an observer at the left or right edge of the diagram. Conformal infinity $\mathscr{I}^{\pm}$ is space-like.

be able to see anything past the dashed line stretching from the bottom-left to top-right. This is qualitatively very different from Minkowski space, where a time-like observer will eventually have access to the entire history of the universe in their past lightcone.

The blue region $\mathcal{O}^+$ in the Penrose diagram is the part of spacetime that this observer can send signals to, while the red region is the part from which they can receive signals. Their intersection $\mathcal{D} = \mathcal{O}^+ \cap \mathcal{O}^-$ is called the causal diamond.

Black holes in de Sitter space

Imposing that the space-time is only asymptotically de Sitter allows us to discuss a variety of more general solutions. The simplest one of which is the Schwarzschild-de Sitter solutions. This solution describes an electrically neutral non-ratating black hole in de Sitter space and can be described by the metric,

$$\mathrm{d}s^2 = \left(1 - \frac{r^2}{L^2} - \frac{2M}{r^{d-2}}\right)\mathrm{d}t^2 - \left(1 - \frac{r^2}{L^2} - \frac{2M}{r^{d-2}}\right)^{-1}\mathrm{d}r^2 - r^2\mathrm{d}\Omega_2^2\,. \tag{8.2}$$

The g_{tt} component of the metric has two positive real zeros, $r_c(M)$ and $r_+(M)$ with $r_c > r_+$ which are the cosmological and black hole horizons. As the parameter $M > 0$ increases, r_c and r_+ tend to each other and eventually meet at a critical mass $M_c = \frac{L}{3\sqrt{3}}$. This is known as the Nariai limit. For $M > M_c$ one finds a cosmological solution with no horizons reaching all the way to $\mathcal{I}^+$ with a space-like singularity at $r = 0$, where r is now a time coordinate. Indeed, for $M > M_c$ one can identify $t \in \mathbb{R}$ without introducing closed timelike curves. For $M < 0$, one finds only one horizon and a naked timelike singularity at $r = 0$. Note that as we take $L \to \infty$ and $d \to 3$ we recover the asymptotically flat Schwarzschild black hole discussed in the previous chapter. As in asymptotically flat space, one can further generalise these solutions to include rotation and/or charge by appropriately modifying the

Kerr(-Newman) and Reissner-Nordstrom black holes. One reason to introduce the Schwarzschild-de Sitter solution is that it plays an important role in determining the entropy and temperature of pure de Sitter, as will be reviewed below.

One reason why we introduced the Schwarzschild-de Sitter black hole is that it plays an important role in the work of Gibbons and Hawking [GH77b] determining the entropy of pure de Sitter (see below for more). To make things more concrete it will be useful to focus on the $(2+1)$-dimensional case,

$$\mathrm{d}s^2 = (1-8GE-r^2)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{(1-8GE-r^2)} - r^2\mathrm{d}\phi^2\,, \tag{8.3}$$

where for convenience we set $L=1$ and normalised the energy E appropriate for a three-dimensional black hole. In three dimensions many things simplify, for example, in this case we only have one horizon at $r_H = \sqrt{1-8GE}$, and as E goes to zero this reduces to the usual horizon in empty de Sitter space. The fact that there is no black hole horizon is not so surprising given that in three-dimensions gravity is not really dynamical and there are no black holes. We can learn a bit more about this solution by looking at it near $r=0$ where we see that the metric behaves as

$$\mathrm{d}s^2 \simeq r_H^2\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{r_H^2} + r^2\mathrm{d}\phi^2\,. \tag{8.4}$$

Defining the rescaled coordinates,

$$t' = r_H t\,, \qquad r' = r/r_H\,, \qquad \phi' = r_H\phi\,, \tag{8.5}$$

this metric reduces to

$$\mathrm{d}s^2 = \mathrm{d}t'^2 - \mathrm{d}r'^2 - r'^2\mathrm{d}\phi'^2\,. \tag{8.6}$$

This looks like flat space but it is not quite. The periodicity of ϕ was 2π so that the periodicity of ϕ' becomes $2\pi r_H$. Therefore we find that the space has a conical singularity with a positive deficit angle at the origin. This might look familiar, if you remember that in three-dimensions if you put a point-like mass you get a conical deficit angle at the position of the particle. Hence, the three-dimensional Schwarzschild-de Sitter background behaves more like a particle then a black hole.

Exercise 8.1. *Show that the Schwarzschild-de Sitter black hole is locally isometric to de Sitter space. More precisely, show that it corresponds to an orbifold of the universal covering space / maximal extension of de Sitter space.*

8.2 Quantum fields in de Sitter space

In this section we discuss the quantization of a non-interacting massive scalar field in a fixed de Sitter background. As discussed in Chapter 4, when working in curved space, it is much harder to define what one means by energy. Similarly, the vacuum is no longer a non-ambiguously defined state. In de Sitter space, thanks to it being maximally symmetric, we can work things out very explicitly.

Let us for simplicity continue with the free scalar field $\phi(x)$ with mass m in a fixed de Sitter background in planar coordinates. The action of such a scalar is given by:

$$S = \frac{1}{2}\int \mathrm{d}^d x \sqrt{|g|}\left(g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + m^2\phi^2 + \xi R\phi^2\right) . \tag{8.7}$$

Using the conformal time coordinate η, the de Sitter metric becomes a FLRW metric where the function $C(\eta)$ is given by

$$C(\eta) = \frac{L^2}{\eta^2} . \tag{8.8}$$

In Chapter 4 we solved the problem of quantising the scalar field in such a background. In de Sitter space, we have

$$C(\eta)R(\eta) = \frac{d(d+1)}{\eta^2} , \tag{8.9}$$

so that the Klein-Gordon equations reduces to

$$\chi_{\mathbf{k}}'' + \left[k^2 + \frac{m^2L^2}{\eta^2} + \frac{d(d+1)(\xi - \xi(d))}{\eta^2}\right]\chi_{\mathbf{k}} = 0 . \tag{8.10}$$

Performing the change of coordinates $s = -k\eta$ and defining

$$\chi_{\mathbf{k}} = \sqrt{s} f(s) , \tag{8.11}$$

it is easy to see that $f(s)$ satisfies the Bessel equation,

$$s^2 f''(s) + s f'(s) + (s^2 - \nu^2) f(s) = 0 , \tag{8.12}$$

where

$$\nu^2 = \frac{1}{4} + (\xi(d) - \xi)\, d(d+1) - m^2L^2 = \frac{d^2}{4} - d(d+1)\xi - m^2L^2 . \tag{8.13}$$

We can then write $\chi_{\mathbf{k}}$ in terms of Bessel functions, for $\eta < 0$, as

$$\chi_{\mathbf{k}}(\eta) = \sqrt{k|\eta|}\,[A_{\mathbf{k}} J_\nu(k|\eta|) + B_{\mathbf{k}} Y_\nu(k|\eta|)] . \tag{8.14}$$

and the normalisation condition leads to

$$-\mathrm{i}k^2\eta\left(A_k B_k^* - A_k^* B_k\right) W\left[J_\nu(k|\eta|), Y_\nu(k|\eta|)\right] = 1 . \tag{8.15}$$

Here $W[f,g] = f g' - f' g$ is the Wronskian, where the derivative is with respect to the full argument $k|\eta|$. The Wronskian can be computed to be

$$W\left[J_\nu(k|\eta|), Y_\nu(k|\eta|)\right] = \frac{2}{\pi x} , \tag{8.16}$$

so that we find

$$A_k B_k^* - A_k^* B_k = -\frac{\mathrm{i}\pi}{2k} . \tag{8.17}$$

To get a better feeling for the behaviour of these functions consider the asymptotic region near $\mathscr{I}_+$, i.e. $\eta \to 0$. In this region, the modes asymptotically behave as

$$\chi_k \propto (-k\eta)^{\frac{1}{2}\pm\nu}. \tag{8.18}$$

Notice that for small mass m and ξ, ν is real and the modes either blow up or vanish asymptotically while for large masses and/or ξ it becomes an oscillatory mode with positive frequency $|\nu|$.

de Sitter vacua

De Sitter space is a static space-time, and moreover maximally symmetric. Hence, guided by the discussion in the previous sections we look for the 'preferred' vacuum which preserves all said symmetries. For this reason, let's consider the behaviour in the asymptotic past $\mathscr{I}_-$, i.e. $k|\eta| \gg 1$, where we have $\omega_k \simeq k$. In this regime we want to consider modes which behave like Minkowski modes in conformal time η,

$$\chi_k \propto \frac{1}{\sqrt{2k}} \mathrm{e}^{-ik\eta}. \tag{8.19}$$

In order to find solutions to the Klein-Gordon with this behaviour, consider the asymptotic behaviour of the Bessel functions. We have

$$\chi_k \sim \sqrt{\frac{2}{\pi}} \left(\frac{A+\mathrm{i}B}{2} \mathrm{e}^{-\mathrm{i}\lambda} + \frac{A-\mathrm{i}B}{2} \mathrm{e}^{\mathrm{i}\lambda} \right), \qquad \eta \to -\infty, \tag{8.20}$$

where

$$\lambda = k|\eta| - \frac{\nu\pi}{2} - \frac{\pi}{4}. \tag{8.21}$$

Therefore, we require that $A + \mathrm{i}B = 0$, which together with the condition (8.17) results in

$$|A|^2 = \frac{\pi}{4k}. \tag{8.22}$$

Inserting this in the expression for the modes we find,

$$\chi_k(\eta) = \frac{1}{2}(\pi|\eta|)^{\frac{1}{2}} \left(J_\nu(k|\eta|) + \mathrm{i}Y_\nu(k|\eta|)\right) = \frac{1}{2}(\pi|\eta|)^{\frac{1}{2}} H_\nu^{(1)}(k|\eta|), \tag{8.23}$$

where $H_\nu^{(1)}$ is the Hankel function of the first kind. This vacuum is called the Euclidean vacuum or Bunch-Davies vacuum $|0\rangle_{BD}$, after [BD78][1]

Positive frequency modes in the Bunch-Davies vacuum state are those which become the positive frequency modes in Minkowski space upon taking the limit $k|\eta| \to \infty$ [BD78] (see also [STY95]). We can thus expand our quantum field in terms of the creation and annihilation operators associated to $|0\rangle$,

$$\phi(\eta, \vec{x}) = \sum_{\mathbf{k}} \left[a_{\mathbf{k}} u_{E,\mathbf{k}}(\eta, \mathbf{k}) + a_{\vec{\mathbf{k}}}^\dagger u_{E,\mathbf{k}}^*(\eta, \mathbf{k}) \right], \tag{8.24}$$

[1] History has its ways so that Bunch and Davies got the honour of naming this vacuum. However, it was already described earlier in various papers such as [CT68, SS76].

where the creation and annihilation operators satisfy the usual properties:

$$a_{\mathbf{k}}|0\rangle = 0\,, \qquad \left[a_{\mathbf{k}}, a^{\dagger}_{\mathbf{k}'}\right] = \delta_{\mathbf{k}\mathbf{k}'}\,. \tag{8.25}$$

This vacuum has various interesting properties. Namely, it is invariant under the de Sitter isometry group SO(1,4). Clearly, it is invariant under rotations of the spatial coordinates $\mathbf{x}$, since $\chi_k(\eta)$ only depends on the modulus of k. It is also invariant under the dilatation,

$$\eta \to \lambda\eta\,, \qquad \mathbf{x} \to \lambda\mathbf{x}\,, \qquad \lambda \in \mathbb{R}\backslash\{0\} \tag{8.26}$$

Indeed, under this transformation the wave-vector transforms as $\mathbf{k} \to \frac{1}{\lambda}\mathbf{k}$ such that the argument of the Hankel function remains invariant. Collecting the overall factor $|\eta|^{\frac{1}{2}}$ in $\chi_k(\eta)$, together with the factor $C^{\frac{d-1}{4}}/4$, we get a total factor of $|\eta|^{\frac{d}{2}}$. However, this factor gets cancelled against the factor of $1/V^{\frac{1}{2}}$ in the wave-function u_k, where $V = L^d$ is the spatial volume. This factor combines with the η factor to produce $(|\eta|/L)^{d/2}$, showing that the modes are invariant under dilatations. Below we will see that the invariance of this vacuum manifests itself in the O(1,4) invariance of the corresponding Wightman functions.

Only demanding that the vacuum state is invariant under the de Sitter isometries does not uniquely pick out the Bunch-Davies vacuum. Indeed, the vacuum state of the quantum field could be rather different. A more general family of vacua is given by the so-called the α-vacua, $|\alpha\rangle$ [Mot85,All85], parameterised by a complex number α. These vacua and their properties are reviewed in [BMS02, SSV01] but as we'll see in a bit, these vacua have some funny properties for which reason we usually discard them. In this course we will mostly focus on the Bunch-Davies vacuum as defined above but surely at some point these funny extra vacua will turn out to have some purpose in life.

Remark. The Bunch-Davies vacuum is picked out for another reason. We can prepare a de Sitter vacuum by starting with a Euclidean sphere, cutting it in half and gluing it to 'half' of the Sitter space. The state prepared by this procedure is precisely the Bunch-Davies vacuum. This origin furthermore elucidates where the name 'Euclidean vacuum' comes from. This construction is motivated by the 'no boundary wave function' proposal of Hartle and Hawking [HH83], where a conjectural description for the wavefunction of the universe was given. According to their prescription, the universe has no origin. If we would travel back towards the beginning of the universe we would note that, similar to in the interior of the black hole, our notion of space and time changes. In this case they propose that close enough to the singularity time stops to exist and we are left with a Euclidean space which smoothly caps off.

Finally, note that when $m = 0$ and $\xi = \xi(d)$, the Bunch-Davies vacuum reduces to the conformal vacuum. Indeed, in this case the index of the Hankel function is $\nu^2 = \frac{1}{4}$ and we have

$$H^{(1)}_{\nu}(z) = -\mathrm{i}\left(\frac{2}{\pi z}\right)^{\frac{1}{2}} \mathrm{e}^{\mathrm{i}z}\,. \tag{8.27}$$

so that the modes reduce up to a phase to the Minkowski modes, $\chi_k - \frac{\mathrm{i}}{\sqrt{2k}}\mathrm{e}^{-ik\eta}$, appropriate for a

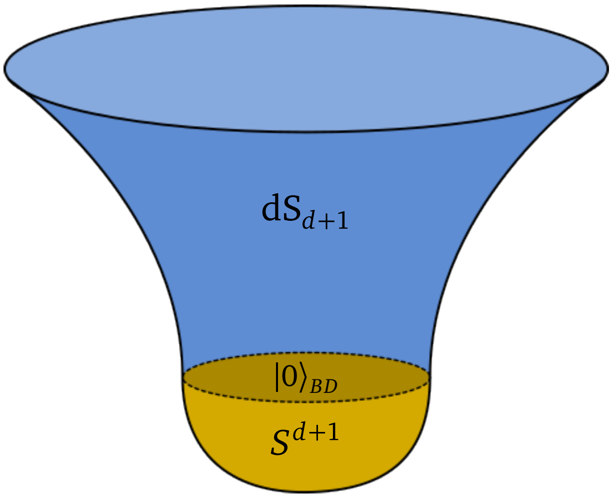


Figure 8.2: The Bunch-Davies state is prepared by gluing the de Sitter space to a sphere, i.e. Euclidean de Sitter space. This construction can be though of as a toy model for the universe.

conformal vacuum.

Green's functions

Having defined the vacuum we would like to compute the Wightman functions. This can be done in two ways. One way is the brute-force approach, by explicitly plugging the equations for the modes in (5.3) and performing the integration over momenta. This has been done in [BD78, SS76, CT68], but it's rather tedious. Another, more elegant approach is to solve the Klein-Gordon equation for $G_+(x,x')$ while assuming $\mathrm{O}(1,d+1)$ invariance along the way. This in fact gives the answer not only for the Bunch-Davies vacuum but also for the more general family of α-vacua $|\alpha\rangle$.

If the Wightman function is computed in a state invariant under the de Sitter isometry group it should only depend on the $\mathrm{O}(1,d+1)$ invariant distance between the two points. This is nothing but the geodesic distance $P(x,x')$ defined in more detail in Appendix E. we can then try to solve the homogeneous Klein-Gordon equation

$$\left(\Box_x + m^2 + \xi R\right) G^+\left(P(x,x')\right) = 0\,. \tag{8.28}$$

To do so let us note two useful properties of the geodesic distance,

$$\nabla^\mu P \nabla_\mu P = \frac{P^2-1}{L^2}\,, \qquad \nabla_\mu \nabla_\nu P = g_{\mu\nu}\frac{P}{L^2}\,. \tag{8.29}$$

Exercise 8.2. *Prove the two properties above. You can do so abstractly or by explicit computation in your favourite coordinate system. Hint: in conformal coordinates* $(\eta,\mathbf{x})$ *these take a particularly simple form. To do so you will need to compute the Christoffel symbols.*

Consider now a function $F(P)$ depending on x only through $P(x,x')$. We then have,

$$\Box_x F(P) = \frac{P^2-1}{L^2}F''(P) + \frac{d+1}{L^2}PF'(P)\,. \tag{8.30}$$

Using these properties, it follows that the Wightman function for a scalar field in de Sitter space satisfies,

$$(P^2-1)\partial_P^2 G^+ + (d+1)P\partial_P G^+ + \left(m^2L^2+\xi d(d+1)\right)G^+ = 0\,. \tag{8.31}$$

After a change of variables, $z=\frac{1+P}{2}$ this becomes a hypergeometric equation,

$$z(1-z)\partial_z^2 G^+ + \left(\frac{d+1}{2}-(d+1)z\right)\partial_z G^+ - \left(\frac{m^2}{H^2}+\xi d(d+1)\right)G^+ = 0\,. \tag{8.32}$$

Comparing this to the standard form (see Appendix F), we find that

$$\alpha+\beta = d\,, \qquad \alpha\beta = m^2L^2+\xi d(d+1)\,, \tag{8.33}$$

so that we find the general solution

$$G^+ = c_{m,d}\,{}_2F_1\left(h_+,h_-,\frac{d+1}{2},z\right), \tag{8.34}$$

where $c_{m,d}$ is a normalisation constant to be determined shortly and we defined

$$h_\pm = \frac{1}{2}\left[d\pm\sqrt{d^2-4(m^2L^2+\xi d(d+1))}\right]. \tag{8.35}$$

The hypergeometric function in (8.34) has a pole at $z=1$, or $P=1$ and a branch cut for $1<P<\infty$. The pole occurs when the points x and x' are separated by a null geodesic. At short distances, the scalar field is insensitive to the fact that it lives in a de Sitter space and the form of the singularity should be the same as that of the propagator in flat Minkowski space. We can use this fact to fix the normalisation constant. Near $z=1$, the hypergeometric function behaves as

$${}_2F_1\left(h_+,h_-,\frac{d+1}{2},z\right) \sim \frac{\zeta(P(x,x'))^{1-d}}{2^{1-d}}\frac{\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(\frac{d-1}{2}\right)}{\Gamma(h_+)\Gamma(h_-)}\,, \tag{8.36}$$

where $\zeta(P)=\cos^{-1}P$ is the geodesic separation between the two points. Comparing this expression with the usual short distance singularity,

$$G^{+,\text{flat}}(x,x') \sim (-1)^{\frac{d-1}{2}}\frac{\Gamma\left(\frac{d-1}{2}\right)}{4\pi^{\frac{d+1}{2}}}\frac{1}{(l^2)^{\frac{d-1}{2}}}\,, \qquad l^2=(x-x')^2\,. \tag{8.37}$$

Thus we find the normalisation constant

$$c_{m,d} = L^{1-d}\frac{\Gamma(h_+)\Gamma(h_-)}{(4\pi)^{\frac{d}{2}}\Gamma\left(\frac{d}{2}\right)}\,. \tag{8.38}$$

Exercise 8.3. *Use the properties of hypergeometric functions to show that near the pole the hypergeometric function behaves as in (8.36). (Hint: to do so write $P=1+\delta$, where $\delta=\frac{(\Delta\eta)^2-(\Delta\mathbf{x})^2}{2\eta\eta'}$ and expand for small δ.)*

Exercise 8.4. *Show that for $m=0$ and $\xi=\xi(d)$ this expression reduces to the expected expression for a two-point function in a conformal vacuum.*

As noted above, the hypergeometic function (8.34) has a branch cut along the semi-infinite axis running from 1 to ∞. This corresponds to points where $P(X,X')\geq 1$, i.e. points inside the lightcone. The prescription for avoiding the singularity at the lightcone is the same as in Minkowski space and simply consists of changing

$$(\eta-\eta')^2 \to (\eta-\eta'-\mathrm{i}\epsilon)^2\,. \tag{8.39}$$

Finally, note that the equation (8.31) is symmetric under interchanging $P\longleftrightarrow -P$. So if $G(P)$ is a solution, then so is $G(-P)$. We therefore find a second linearly independent solution,

$$G_+^{(2)} = c_{m,d}\,{}_2F_1\left(h_+,h_-,\frac{d+1}{2},\frac{1-P}{2}\right). \tag{8.40}$$

The singularity now lies at $P=-1$, which corresponds to X being null separated from the antipodal point to X'. This singularity sounds rather nonphysical at first, but we should recall that antipodal points in de Sitter space are always separated by a cosmological horizon. The Green's function (8.40) can thus be thought of as arising from an image source behind the horizon, and is non-singular everywhere within an observer's horizon. Hence the nonphysical singularity can never be detected by any experiment. The de Sitter space therefore has a one parameter family of de Sitter invariant Green's functions

$$G_+^{(\alpha)} = \langle\alpha|\,\phi(x)\phi(x')\,|\alpha\rangle = \sin\alpha\, G^+ + \cos\alpha\, G^{+(2)}\,, \tag{8.41}$$

corresponding to the α vacua $|\alpha\rangle$ discussed above. Putting $\alpha=0$ the antipodal singularity disappears and the vacuum reduces to the Bunch-Davies vacuum.

8.3 De Sitter temperature

Having discussed some properties of quantum fields in a fixed de Sitter background, let us proceed to give a brief account of some semi-classical aspects. In particular, we will show that an observer moving along a time-like geodesic observes a thermal bath of particles when the scalar field is in the (Bunch-Davies) vacuum state $|0\rangle$. We conclude that the de Sitter space is naturally associated with a temperature T_{dS}.

In Chapter 6 we introduced a variety of tools to diagnose the thermal nature of spacetimes. Now the time has come to put them to good use. Let us consider an observer, equipped with an Unruh detector, moving along a time-like geodesic. Along a time-like geodesic, the geodesic distance P is given by

$$P = \cosh\Delta\tau/L\,, \tag{8.42}$$

where $\Delta\tau$ is the difference in proper time between the two points under consideration. Let us now consider the detector response function in de Sitter space,

$$\begin{aligned} f(E) =& \int \mathrm{d}\Delta\tau\, \mathrm{e}^{-\mathrm{i}E\Delta\tau} G^+(\Delta\tau/L) \\ =& c_{m,d}\,{}_2F_1\left(h_+,h_-,\frac{d+1}{2},\frac{1}{2}(1+\cosh\Delta\tau/L)\right). \end{aligned} \tag{8.43}$$

As discussed above, the hypergeometric function has poles at $P = 1$ or equivalently at $\Delta\tau = 2\pi \mathrm{i} L n$, for any integer $n \in \mathbb{Z}$. To compute the integral in (8.43), consider the contour depicted in Figure 8.3. Since the contour does not contain any poles the total integrand vanishes

$$\int_{-\infty}^{\infty} d\Delta\tau\, \mathrm{e}^{-\mathrm{i}E\Delta\tau} G^{+}(\cosh\Delta\tau) + \int_{+\infty-\mathrm{i}\beta}^{-\infty-\mathrm{i}\beta} d\Delta\tau\, \mathrm{e}^{-\mathrm{i}E\tau} G^{+}(\cosh\Delta\tau) = 0\,, \tag{8.44}$$

where $\beta = 2\pi L$. After changing variables in the second integral $\Delta\tau \to -\Delta\tau - \mathrm{i}\beta$ this implies that

$$f(E) = \mathrm{e}^{-\beta E} f(-E)\,, \tag{8.45}$$

which is nothing but the principle of detailed balance in a thermal ensemble!

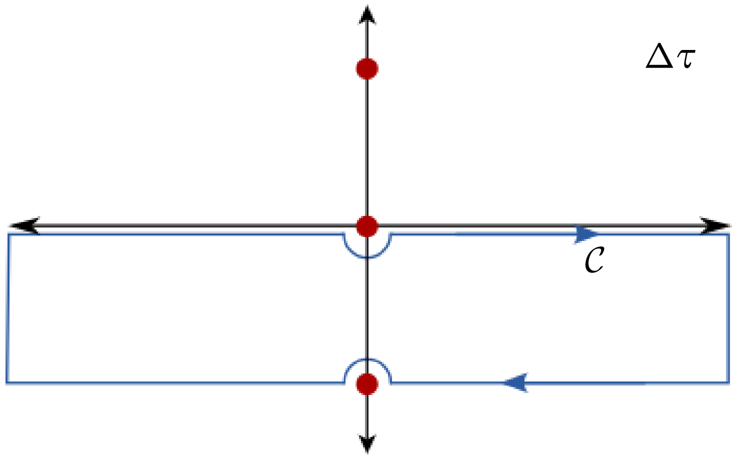


Figure 8.3: The integrand in (8.43) has singularities at $\Delta\tau = 2\pi \mathrm{i} n$ for any integer n. The contour $\mathcal{C}$ will be used to compute the integral.

Remark. Let us suppose that we are in a state where we have the following relation between probabilities,

$$P(E_i \to E_j) = P(E_j \to E_i)\mathrm{e}^{-\beta(E_j - E_i)}\,, \tag{8.46}$$

for all i and j. Furthermore suppose that the energy levels of the detector are thermally populated, i.e.

$$n_i = N\mathrm{e}^{-\beta E_i}\,, \tag{8.47}$$

for some N. Then it is clear that the total transition rate from E_i to E_j is the same as that from E_j to E_i,

$$R(E_i \to E_j) = n_i P(E_i \to E_j) = N\mathrm{e}^{-\beta E_i} P(E_j \to E_i) = R(E_j \to E_i)\,. \tag{8.48}$$

Hence the state we describe is a thermal state at equilibrium at a temperature $T = \frac{1}{\beta}$. This is the principle of detailed balance applied to a thermal state.

Since the Green function used in this calculation is invariant under all de Sitter isometries, the result for the temperature is the same for any observer moving along a time-like geodesic. We conclude that any geodesic observer in de Sitter space will feel a thermal bath of particles at temperature,

$$T_{\mathrm{dS}} = \frac{1}{2\pi L} \,. \tag{8.49}$$

As discussed in the previous chapter, the laws of black hole thermodynamics [BCH73, Bek73] were a crucial stepping stone paving the way toward a better understanding of the microscopic nature of a black hole. The basic observation was that classically any physical process can only lead to an increase in the area A_{BH} of the black hole horizon (area law) and the way the area responds to classical processes follows an equation equivalent in form to the first law of thermodynamics,

$$\delta M = T_{BH} \delta S_{BH} + \Omega_H \delta J \ , \tag{8.50}$$

where δM is the change in the ADM mass, T_{BH} is the surface gravity divided by 2π, Ω_H is the angular velocity of the horizon, δJ is the change in ADM angular momentum and $S_{BH} = A_{BH}/4G$. Hawking famously discovered that T_{BH} is in fact the temperature of the black hole as measured by a far away observer. The appearance of the first law was no coincidence and a statistical mechanics interpretation was eventually provided by string theory [SV96]. The black hole entropy is interpreted as a count of the number of microstates with macroscopic quantities equivalent to those of the black hole they constitute.

In analogy to the case of black holes, it was proposed by Gibbons and Hawking [GH77b] that there exists an entropy associated to the pure de Sitter horizon,

$$S_{dS} = \frac{\pi L^2}{G_N} \,. \tag{8.51}$$

This de Sitter entropy is proportional to the area of the horizon, as in the case of black hole entropy. Furthermore, as with all horizons, Hawking's famous result tells us that the de Sitter horizon has a temperature associated to it, given by $T_{\mathrm{dS}} = 1/2\pi L$. As before this thermal behaviour is manifested in the Euclideanised version of de Sitter. In particular, considering the static patch metric and transforming $t \to i\tau$ we find that in order to avoid conical singularities in the Euclidean geometry we have to impose the following periodicity on the Euclidean time coordinate,

$$\tau \sim \tau + 2\pi L \,. \tag{8.52}$$

Once again we see that the temperature is indeed matched by the period of the imaginary time. For the cosmological constant measured in our own universe one finds $S_{dS} \sim 10^{120}$, which far exceeds the present entropy[2] of the remaining matter content.

When studying black hole thermodynamics, one typically asks how the black hole responds to some classical process such as absorbing some mass. What is the analogous question in the case of a cosmological horizon surrounding the static patch observer? We begin with some mass M localized at the center of the static patch. From analysing the metric (8.2), we see that this has the effect of reducing the size of the cosmological horizon. One can indeed check that if the mass falls outside the

[2]The entropy of the present day universe, including the Bekenstein-Hawking entropy of supermassive black holes, is estimated around 10^{104} (see [EL10] and references therein).

cosmological horizon

$$-\delta M = T_{\rm dS}\delta S_{cos}\ . \tag{8.53}$$

This is the analogue law of thermodynamics for a de Sitter horizon. It can be naturally generalized to include angular momentum and other conserved quantities. The point is that the de Sitter horizon responds to physical processes very much like any other horizon. Notice however the minus sign in front of δM. The entropy increases when we throw mass *outside* the cosmological horizon surrounding us. Indeed, the less information we have about the interior of the cosmological horizon, the higher its entropy will be.

We end with an important remark. Though it seems that much of the black hole picture carries forward to the case of the de Sitter horizon, there are some crucial differences. For instance, an observer can never approach her own cosmological horizon and probe it. In other words, there is no sense in which the cosmological horizon is an object localized in space as would be the case for a black hole. Also, black holes in flat space decay when emitting Hawking radiation. On the other hand, at least naively, the Hawking radiation of a de Sitter horizon is reabsorbed by the de Sitter horizon itself, leading to no overall evaporation of the horizon.

8.4 De Sitter entropy

To finish this chapter, let us discuss the entropy associated to de Sitter space in some more detail. For simplicity, let us again restrict our attention to dS_3, where the analysis simplifies considerably.

For the case of black holes one can use similar methods as those in the previous chapter to calculate the temperature $T_{\rm BH}$ of the black hole. The black hole entropy $S_{\rm BH}$ can then be found by integrating the thermodynamic relation

$$\frac{dS_{\rm BH}}{dE_{\rm BH}} = \frac{1}{T_{\rm BH}}, \tag{8.54}$$

where $E_{\rm BH}$ is the energy or mass of the black hole. So if you know the value of the temperature just for one value of $E_{\rm BH}$ you will not be able to get the entropy, but if you know it as a function of the black hole mass then you can simply integrate (8.54) to find the entropy. The constant of integration is determined by requiring that a black hole of zero mass has zero entropy.

So for de Sitter space one would expect to use the relation

$$\frac{dS_{\rm dS}}{dE_{\rm dS}} = \frac{1}{T_{\rm dS}} \tag{8.55}$$

to find the entropy $S_{\rm dS}$. The problem in de Sitter space is that once the coupling constant of the theory is chosen there is just one de Sitter solution, whereas in the black hole case there is a whole one parameter family of solutions labeled by the mass of the black hole, for fixed coupling constant. In other words, what is $E_{\rm dS}$ in (8.55)? One might try to vary the cosmological constant, but that is rather unphysical as it is the coupling constant. One would be going from one theory to another instead of from one configuration in the theory to another configuration in the same theory.

Let us instead follow Gibbons and Hawking [GH77b] and use the one parameter family of Schwarzschild-de Sitter solutions to see how the temperature varies as a function of the parameter E

labeling the mass of the black hole.

Exercise 8.5. *The three-dimensional Schwarzschild-de Sitter in static coordinates is*

$$ds^2 = (1-8GE-r^2)dt^2 - \frac{dr^2}{(1-8GE-r^2)} - r^2 d\phi^2 . \tag{8.56}$$

Find a Green function by analytic continuation from the smooth Euclidean solution. Show that this Green function is periodic in imaginary time with periodicity

$$t \sim t + \frac{2\pi i}{\sqrt{1-8GE}}. \tag{8.57}$$

From this exercise and the discussion above we conclude that the temperature associated with the Schwarzschild-de Sitter solution is

$$T_{\text{SdS}} = \frac{\sqrt{1-8GE}}{2\pi}. \tag{8.58}$$

Using the formula

$$\frac{dS_{\text{SdS}}}{dE} = \frac{1}{T_{\text{SdS}}}, \tag{8.59}$$

and writing the result in terms of the area A_H of the de Sitter horizon at $r_H = \sqrt{1-8GE}$ which is given by

$$\sqrt{1-8GE} = \frac{A_H}{2\pi}, \tag{8.60}$$

one finds that the entropy is equal to

$$S_{\text{SdS}} = -\frac{A_H}{4G} . \tag{8.61}$$

This differs by a minus sign from the famous Bekenstein-Hawking formula! What did we do wrong? Gibbons and Hawking suggested that to get the de Sitter entropy we should use not (8.59) but instead

$$\frac{dS_{\text{SdS}}}{d(-E_{\text{dS}})} = \frac{1}{T_{\text{SdS}}} . \tag{8.62}$$

This looks funny but in fact there is a very good reason for using this new formula.

The de Sitter entropy, although we don't know exactly how to think about it, is supposed to correspond to the entropy of the stuff behind the horizon which we can't observe. Now in general relativity the expression for the energy on a surface is the integral of a total derivative, which reduces to a surface integral on the boundary of the surface, and hence vanishes on any closed surface. Consider a closed surface in de Sitter space such as the one shown in figure 8.4. If we put something with positive energy on the south pole, then necessarily there will be some negative energy on the north pole.

This can be seen quite explicitly in the Schwarzschild-de Sitter solution. With no black hole, the space-like slice in figure 8.4 is an S^2, but we saw that in the Schwarzschild-de Sitter solution there is a positive deficit angle at both the north and south poles. If we ascribe positive energy to the positive deficit angle at the south pole, then because the Killing vector $\partial/\partial t$ used to define the energy changes direction across the horizon, we are forced to ascribe negative energy to the positive deficit angle at the north pole.

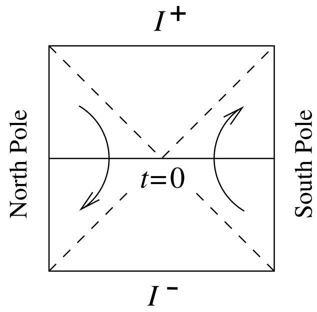


Figure 8.4: The energy associated to the Killing vector $\partial/\partial t$ (indicated by the arrows) along the spacelike slice $t = 0$ (solid line) must vanish. If we ascribe positive energy to a positive deficit angle at the south pole, then we must ascribe negative energy to a positive deficit angle at the south pole since the Killing vector $\partial/\partial t$ runs in the opposite direction behind the horizon.

Therefore the northern singularity of Schwarzschild-de Sitter behind the horizon actually carries negative energy. In (8.59) we varied with respect to the energy at the south pole, and ended up with the wrong sign in (8.61), but if we more sensibly vary with respect to the energy at the north pole, then we should use the formula (8.62). Then we arrive at the entropy for Schwarzschild-de Sitter

$$S_{\text{SdS}} = \frac{A_H}{4G} = \frac{\pi}{2G}\sqrt{1-8GE}\,. \tag{8.63}$$

The integration constant has been chosen so that the entropy vanishes for the maximal energy $E = \frac{1}{8G}$ at which value the deficit angle is 2π and the space has closed up. In conclusion, we see that the Bekenstein-Hawking area-entropy law indeed continues to apply in three dimensional Schwarzschild-de Sitter. Moreover, we see that the de Sitter entropy is maximal for an empty, pure de Sitter space. After a moment of thought this is very logical though. Our universe is expanding, so eventually all matter will dilute and end up behind the cosmological horizon. In order to preserve the second law of thermodynamics we are automatically lead to the conclusion that empty de Sitter has to have maximal entropy.

This is in stark contrast to flat space or anti-de Sitter where if we throw more matter into the space, or build a bigger black hole the entropy continues to increase. Indeed, in these spaces the universe is not expanding and eventually the gravitational force will win and pull all the matter together into one huge black hole, which according to the second law of thermodynamics is then the most entropic state in (asymptotically) flat or anti-de Sitter spaces.

Chapter 9

Cosmological perturbations

In a maximally symmetric universe such as Minkowski or (A)dS, there cannot be a beginning or end of time. There cannot be any 'history' because each time is equivalent to any other time. While such eternal universes might be appealing from a philosophical or aesthetic point of view, it is in contradiction with the last century of cosmological observations. This simplest possible non-trivial cosmological toy models are FLRW universes.

The same process that creates particles in an expanding universe is responsible for creating, in the context of an early inflationary expansion, the primordial fluctuations that are now observed with astonishing accuracy in the cosmic microwave background (CMB) radiation. These same primordial fluctuations also appear responsible for the large-scale structure of the universe.

9.1 Cosmological models

The FLRW models are models in which we assume a collection of co-moving observers for whom the universe is homogeneous (the same for each observer) and isotropic (the same in every direction). In Appendix E we introduce some basic facts about FLRW spaces as well as discuss the Friedmann equations in general dimensions. In this chapter we will mostly restrict ourselves to $d=3$ spatial dimensions. For more details we refer the reader to one of the standard textbooks on modern cosmology, such as [Dod03].

The FLRW metric has been introduced before and can be written as

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a(t)^2 \mathrm{d}s^2_{\mathcal{M}_{k,3}}\,, \tag{9.1}$$

where $\mathcal{M}_{k,3}$ is the maximally symmetric spatial manifold with curvature $k=0,\pm 1$. For any such spatial manifold, the FLRW metric is conformally flat, as can immediately be seen by going to conformal coordinates

$$\mathrm{d}s^2 = C(\eta)^2\left(\mathrm{d}\eta^2 - \mathrm{d}s^2_{\mathcal{M}_{k,3}}\right), \qquad \eta(t) = \int^t \frac{\mathrm{d}t'}{a(t')}\,, \tag{9.2}$$

where $C(\eta) = a(t(\eta))$. In the context of these cosmological models, the conformal coordinate is particularly useful to understand the causal structure of the model. In our own universe, the spatial curvature is tantalisingly close to $k=0$, but so far experiments are not yet able to determine whether the spatial curvature is exactly vanishing or not. A particularly important quantity in such models is the Hubble parameter,

$$H(t) = \frac{\dot{a}(t)}{a(t)} = \frac{\dot{C}(\eta)}{C(\eta)^2}\,, \tag{9.3}$$

which captures the time dependence of the scale factor. All the cosmological phenomena we will encounter will have an crucial dependence on H. As long as the so-called null energy condition (NEC) is met, H will be decreasing during the expansion of the universe. A purely (positive) cosmological constant saturates the NEC and consequentially, de Sitter space has a constant hubble parameter,

$$H_{\rm dS} = \frac{1}{L} \,. \tag{9.4}$$

A natural assumption for the stress tensor, respecting the symmetries of the problem, is that of a perfect fluid,

$$T_{\mu\nu} = \mathrm{diag}(\rho, -p, -p, -p)\,, \tag{9.5}$$

The Einstein equations for an FLRW metric are usually referred to as the Friedman equations and in terms of the conformal coordinate η take the form,

$$\left(\frac{\dot{C}}{C}\right)^2 = -\frac{8\pi}{3}\rho - \frac{k}{C^2}\,, \qquad 6\frac{\ddot{C}+C}{C^3} = 8\pi G_N(\rho - 3p)C + \frac{k}{C}\,. \tag{9.6}$$

Together with the continuity equation,

$$\dot{\rho} + 3\frac{\dot{C}}{C}(\rho + p) = 0\,, \tag{9.7}$$

these determine the full evolution of the FLRW universe. However, the form of the equation of state $p(\rho)$ is not determined by these equations and has to be put in by hand. For non-relativistic matter (dust), relativistic matter (radiation) or a cosmological constant the equation of state takes the form $p = w\rho$, where $w \sim 0$ for non-relativistic matter, $w = \frac{1}{3}$ for relativistic matter and $w = -1$ for the cosmological constant.

Example 9.1. *As an example, let us consider dust in an FLRW universe with $k = 1$. After solving the Friedman equations we find,*

$$C(\eta) = c(1 - \cos\eta)\,, \qquad t = c(\eta - \sin\eta)\,. \tag{9.8}$$

Hence, we find a big bang ($C = 0$) at $\eta = 0$, followed by a big crunch at $\eta = \pi$.

Why inflation?

To finish the discussion of cosmological models, let us recall a particular problem for universes undergoing a decelerated expansion. In the current standard cosmological model (known as ΛCDM) an accelerated expansion is induced at late times by the cosmological constant. At earlier times, when the universe is radiation or matter dominated, we expect a period of decelerated expansion. We refer to such models as hot big bang models, where the adjective hot refers to the temperature of radiation.

In general, in order to justify the homogeneity of the spatial slices, we would like to have that the distance between regions of space that look the same is much smaller then the maximal distance travelled by light since the beginning of time. Otherwise it is hard to explain why the two region can look similar, since their causal past is disconnected. However, in hot big bang models, this desired inequality is dramatically violated. More precisely, cosmological observations of far away objects allow us to see regions in the past that are separated by much more that than the particle horizon at the time, which is the furthest a signal can travel. Any mechanism attempting to explain homogeneity across these regions then necessarily violates causality and/or locality, leading to the horizon problem.

This apparent tension is resolved by inflation. This scenario posits that there is a surface of last scattering at some time t_s soon after the big bang, before which we cannot clearly see what was going on. This surface is where radiation decouples from matter and so after this time, we can see what is going on, whereas before, we just have what we see from the cosmic microwave radiation. Inflationary models are then obtained by gluing in an exponentially expanding region of de Sitter space, before the surface of last scattering. This inflationary phase gives the past of distant regions time to mix and homogenise so as to explain the homogeneity and isotropy of the universe.

Guided by experimental evidence, it is by now firmly believed that the cosmological constant of our universe is positive. This implies that at large times the scale factor will diverge. At such time the contributions from the cosmological constant will dominate the Friedmann equations which in this regime (for $k=1$) we can approximate by,

$$\dot{C} = \sqrt{\frac{\Lambda}{3}} C^2 , \qquad C \sim \frac{1}{\eta_{\mathscr{I}} - \eta} . \tag{9.9}$$

We find that $C \sim -\sqrt{\frac{\Lambda}{3}} \frac{1}{\eta - \eta_{\mathscr{I}}} + \cdots$ such that the scale factor has a pole at conformal infinity $\eta = \eta_{\mathscr{I}}$. This implies that near $\mathscr{I}^+$, conformal infinity looks like that of de Sitter, i.e. we have an asymptotically de Sitter universe. For $k = 0, -1$ the result is similar. This exponential expansion arises because a positive cosmological constant has a stress-energy tensor $T_{\mu\nu} = \text{diag}(\Lambda, -\Lambda, -\Lambda, -\Lambda)$, so that although the effective energy density is positive, the pressure is negative. At the current age of the universe, the contribution of Λ to the energy density is thought to be of the same order as that of the matter including dark matter (visible matter being thought to be 3%, dark matter 30% and cosmological constant about 67% of the critical mass of the universe). Such a ratio of matter to cosmological constant is extremely high at early times, and extremely low at late times, and this sometimes leads to the 'why are we alive now?' question. The later periods are however, very cold and boring, and the early periods rather hot, and too early for structure to form, leading to the answer in the form of an anthropic principle.

In this section we introduced just one motivation for considering inflation, i.e. the horizon problem. However, there are various other problems arising in hot big bang models, such as

- The curvature related to the approximate spatial flatness of our current universe,
- The particle horizon problem related to the statistical isotropy of our universe,
- The phase coherence problem related to the homogeneity of the CMB,
- The scale invariance problem related to the scale invariance of the CMB.

Discussing all these problems in more depth would take us too far, but importantly inflation offers a way out for each of them and provides us with a plausible explanation for the current state of the universe. For more details we refer to [Dod03, Paj20].

9.2 QFT on a cosmological background

Having briefly introduced the cosmological FLRW models, let us apply the methods introduced in this course to such space-times. In chapter 4 we already consider the quantum scalar field in such backgrounds and we refer the reader to that chapter for details on the quantisation. Here we only repeat the results we need for the analysis at hand.

Let us for simplicity consider the FLRW metric with flat spatial slices

$$ds^2 = dt^2 - a(t)^2 d\mathbf{x}\cdot d\mathbf{x} = C^2(d\eta^2 - d\mathbf{x}\cdot d\mathbf{x}), \tag{9.10}$$

where t is the proper time of the co-moving observers, a the scale factor, and η the conformal time. Writing the modes as

$$u_{\mathbf{k}}(\eta,\mathbf{x}) = \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{(2\pi)^{3/2}C(\eta)^{1/2}}\chi_{\mathbf{k}}(\eta), \tag{9.11}$$

The Klein-Gordon reduces to the following equation for

$$\chi''_{\mathbf{k}}(\eta) + \omega^2(\eta)\chi_{\mathbf{k}}(\eta), \tag{9.12}$$

where

$$\omega(\eta)^2 = \mathbf{k}^2 + m^2 C(\eta)^2 - (1-6\xi)\frac{\ddot{C}}{C}. \tag{9.13}$$

The inner produce, or equivalently the symplectic form on the phase space, is conserved in time. However, now that the space-time is explicitly time-dependent, energy is not conserved and in the absence of asymptotically static patches, particles will be created at any time and there is no fixed or preferred vacuum.

However, at each time η there is some notion of vacuum, i.e. the state without particles, or equivalently the instantaneous low energy state. Expanding the quantum scalar field as

$$\phi = \int \frac{d^3\mathbf{k}}{\sqrt{(2\pi)^3 C(\eta)}}\left(a_{\mathbf{k}}\chi_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}} + a^\dagger_{\mathbf{k}}\chi^*_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{x}}\right), \tag{9.14}$$

we find the Hamiltonian

$$H = \frac{1}{4}\int d^3\mathbf{k}\left(F_{\mathbf{k}}a_{\mathbf{k}}a_{-\mathbf{k}} + F^*_{\mathbf{k}}a^\dagger_{\mathbf{k}}a^\dagger_{-\mathbf{k}} + (a^\dagger_{\mathbf{k}}a_{\mathbf{k}} + a_{\mathbf{k}}a^\dagger_{\mathbf{k}})E_k\right), \tag{9.15}$$

where

$$E_{\mathbf{k}} = |\dot{\chi}|^2 + \omega^2|\chi|^2, \qquad F_{\mathbf{k}} = \dot{\chi}^2_{\mathbf{k}} + \omega^2\chi^2. \tag{9.16}$$

With this implicit choice of vacuum we find

$$\frac{E}{V} = \langle 0|H|0\rangle = \frac{1}{4}\int d^3\mathbf{k}E_{\mathbf{k}}, \tag{9.17}$$

where as usual we took out a diverging volume factor $V \sim \delta^{(3)}(0)$ to get the local energy density. Minimising $E_{\mathbf{k}}$ for each mode subject to the appropriate normalisation condition a brief computation shows that we obtain the following initial data to impose at the hypersurface $\eta = \eta_0$

$$(\chi_{\mathbf{k}}, \dot{\chi}_{\mathbf{k}})\big|_{\eta=\eta_0} = \frac{1}{\sqrt{2\omega(\eta_0)}}(1, -\mathrm{i}\omega(\eta_0)) . \tag{9.18}$$

The particle content measured in this vacuum is that of the instantaneous 'static' observer.

Exercise 9.1. *Derive the initial conditions* (9.18) *starting from the reduced Klein-Gordon equation* (9.12) *by demanding that the energy is minimized for each mode.*

When the frequencies $\omega(\eta)$ are varying slowly enough, we can use an adiabatic approximation to define also an evolution of the vacuum. This arises as the WKB approximation at leading order; more generally, the adiabatic approximation takes the WKB approximation beyond leading order and then there will be particle creation. [1]

Having defined the instantaneous vacuum at each time η and ideally also have an evolution equation, we can define the Bogoliubov transformations between any two times η_1 and η_2 and compute the particle creation just like we did in the previous chapters. As we discussed above, the $E_{\mathbf{k}}$ determines the energy at some time η. The $F_{\mathbf{k}}$ on the other hand determines the instantaneous particle creation at each time. If we are in the lowest energy vacuum at some time η_0 the $F_{\mathbf{k}}$ at that time vanishes.

A novel feature of having the time dependent frequencies is that for large length scales and small enough k and m, the frequency ω_k^2 can become negative. This, as we will see now, leads to the corresponding modes ceasing to oscillate and essentially freezing out.

To see this in more detail, let us consider the simplest possible cosmological model, namely de Sitter. Looking at the planar coordinates (see Appendix E) we see that de Sitter is an FLRW space with scale factor $a = L\mathrm{e}^{t/L}$ and constant Hubble parameter $H = 1/L$. In particular, the Hubble parameter is proportional to the radius of the cosmological horizon.

The de Sitter mode functions, discussed in the previous chapter, behave very different from their Minkowski counterparts when the wave-number $\mathbf{k}$ becomes smaller that the co-moving Hubble parameter

$$k < k_{\mathrm{HC}} = aH = \frac{1}{|\eta|} , \tag{9.19}$$

where HC stands for Hubble crossing, sometimes also called horizon crossing. In physical length scales this means that the physical wavelength $\lambda = a/k$ is stretched by the expansion to become larger than the Hubble radius $1/H$. Since k and H are constant, while $a = \mathrm{e}^{t/L}$ grows with time, all modes cross the Hubble radius as time proceeds and eventually become "super-Hubble" modes. Unlike "sub-Hubble" modes with $k \gg aH$, which oscillate, super-Hubble modes freeze out and asymptote to a constant.

[1]This requires a mechanism for the dissipation of the particles created and otherwise we are working in the full quantum theory in the Heisenberg representation where the state is fixed and the evolution is carried by the operators.

9.3 Cosmological correlators

Having discussed some general properties of quantum fields in cosmological backgrounds, a natural question is: What are the observables for these theories. As familiar from the above, these are given by the expectation values of operators. However, in cosmology, we have observational access only to expectation values in the infinite future $\eta \to 0$. In this limit, observables become approximately constant and so we will only be interested in the expectation values of products of local operators inserted at equal times. Such operators are often called cosmological correlators

$$\lim_{\eta\to 0} \langle \mathcal{O}(\mathbf{x}_1,\eta)\mathcal{O}(\mathbf{x}_2,\eta)\cdots\mathcal{O}(\mathbf{x}_n,\eta)\rangle \,. \tag{9.20}$$

Since we are studying free theories, all information is contained in the two-point correlators of ϕ and its conjugate momentum π. All odd point correlators vanish by the symmetry $\phi \longleftrightarrow = \phi$ and all higher even-point correlators can be reduced to two-point correlation functions using Wick's theorem. Let us therefore consider the two-point correlator in a generic cosmological background,

$$\begin{aligned}\lim_{\eta\to 0}\left\langle \phi(\mathbf{k})\phi(\mathbf{k}')\right\rangle =&|u_{\mathbf{k}}|^2\left\langle a_{\mathbf{k}}a^{\dagger}_{-\mathbf{k}'}\right\rangle\\ =&(2\pi)^3\delta(\mathbf{k}-\mathbf{k}')P(k)\,,\end{aligned} \tag{9.21}$$

where we Fourier transformed to momentum space and introduced the power spectrum

$$P(k) = \frac{H^2}{2k^3}\,, \tag{9.22}$$

for a massless scalar. Note that the Dirac delta simply encodes the momentum conservation and that the power spectrum does not depend on the direction of $\mathbf{k}$, due to the isotropy of the background. Furthermore, the fact that the power spectrum asymptotes to some non-vanishing constant at $\eta \to 0$ is related to the absence of mass. Similarly, the k-dependence $P \propto k^{-3}$ is corresponding to the fact that the massless scalar is scale invariant. Indeed this can be seen by Fourier transforming back to real space where the correlation function is constant and independent of the separation between the two points.

Introducing mass, we find the power spectrum to be

$$P(k) = |u_{\mathbf{k}}|^2 = \frac{H^2}{\pi 2^{2(\nu-1)}\Gamma(\nu)^2}\frac{(-k\eta)^{3-2\nu}}{k^3}\,, \tag{9.23}$$

for $m^2 < \frac{9}{4}H^2$, where $\nu = \sqrt{\frac{9}{4}-\frac{m^2}{H^2}}$, for a minimally coupled scalar (see previous chapter). Because of the mass, the power spectrum is not scale invariant anymore. Indeed, $P \propto k^{-2\nu}$. Also, P has acquired a time dependence. For positive $m^2 > 0$, one finds $3-2\nu > 0$ and the power spectrum decays with time and vanish at future infinity. This is to be expected because the quadratic potential pushes the field towards $\phi = 0$. For negative m^2 we would expect an instability and indeed the power spectrum grows with time and diverges at future infinity. The second case is when the mass square is large and positive, $m^2 > \frac{3}{4}H^2$, then ν becomes complex and the the power spectrum oscillates while decaying as η^3. In cosmology, we are mostly interested in massless or almost massless fields, which do not create large instability and whose perturbations survive long enough to be observable at late times.

9.4 Fluctuating gravitons

By following a very similar procedure for the scalar field, it is straightforward to quantize metric fluctuations as well. To do so, we divide the metric into a classical background and small quantum fluctuations,

$$g_{\mu\nu}(t,\mathbf{x}) = \bar{g}_{\mu\nu}(t,\mathbf{x}) + h_{\mu\nu}(t,\mathbf{x})\,. \tag{9.24}$$

A priori, there are ten independent components of the fluctuations. Four of them however obey constraint equations which are at most first order in time derivatives and therefore not dynamical. To see this recall the Bianchi identity

$$\nabla^{\mu} G_{\mu\nu} = 0\,. \tag{9.25}$$

Writing this out, we obtain

$$\partial_t G^{t\nu} = -\partial_m G^{m\nu} - \Gamma^{\alpha}_{\alpha\gamma} G^{\nu\gamma} + \Gamma^{\nu}_{\alpha\gamma} G^{\alpha\gamma}\,. \tag{9.26}$$

Since the right hand side has at most second order derivatives of the metric, we see that $G^{t\nu}$ has at most one time derivative. Hence the metric must appear with just one time derivative in four of the Einstein equations and we must have some constraint equations that limit the set of consistent initial data for the fluctuations. It takes a bit more work to which has been done in GRII to fully linearise the fluctuations and Einstein equations but let us here just summarise the result for spatially flat FLRW metrics.

A convenient gauge choice is given by

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(\delta_{mn} + \gamma_{mn})\mathrm{d}x^m \mathrm{d}x^n\,, \tag{9.27}$$

where γ is transverse, i.e. $\partial^m \gamma_{mn} = 0$ and traceless $\gamma^m_m = 0$. So we are left with two independent components. Expanding the Einstein Hilbert action to quadratic order in the fluctuations we find

$$S_2 = \frac{M_{\mathrm{Pl}}^2}{8} \int \mathrm{d}^3\mathbf{x}\,\mathrm{d}t\,a^2 \left(\gamma'_{mn}\gamma'^{mn} - \partial_m \gamma_{nk} \partial^m \gamma^{nk}\right)\,. \tag{9.28}$$

This action can be derived by rigorously linearising Einsteins equations, but alternatively it could easily have been guessed by writing the simplest action consistent with the symmetries of the problem. As we did for the scalar we can expand the graviton in plane waves by writing

$$\gamma_{mn}(x) = \int \mathrm{d}^3\mathbf{k} \sum_{s=\pm} \epsilon^s_{mn}(\mathbf{k}) \gamma_s(\mathbf{k}) \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\,, \tag{9.29}$$

where $\epsilon^s_{mn}(\mathbf{k})$ are the polarisation tensors, which are generally complex and satisfy

$$\begin{aligned} \epsilon^s_{mm}(\mathbf{k}) =& \mathbf{k}^m \epsilon^s_{mn}(\mathbf{k}) = 0\,, && \text{Transverse and traceless}\,, \\ \epsilon^s_{mn}(\mathbf{k}) =& \epsilon^s_{nm}(\mathbf{k})\,, && \text{Symmetric}\,, \\ \epsilon^s_{mn}(\mathbf{k})\epsilon^s_{nk}(\mathbf{k}) =& 0\,, && \text{Light-like}\,, \\ \epsilon^s_{mn}(\mathbf{k})\epsilon^{s'mn}(\mathbf{k}) =& \delta^{ss'}\,, && \text{Unit normalised}\,, \\ \epsilon^s_{mn}(\mathbf{k})^* =& \epsilon^s_{mn}(-\mathbf{k})\,, && \gamma_{ij} \text{ is real}\,. \end{aligned} \tag{9.30}$$

From these properties we can derive explicit expressions for the polarisation vector and rewrite the action as

$$S_2 = \frac{M_{\text{Pl}}^2}{4} \int \mathrm{d}^3\mathbf{k}\,\mathrm{d}t\, a^2 \sum_{s=\pm} \left(\gamma_s'(\mathbf{k})\gamma_s'(-\mathbf{k}) - \frac{k^2}{a^2}\gamma_s(\mathbf{k})\gamma_s(-\mathbf{k}) \right). \tag{9.31}$$

Now this action consists of two independent copies of the action for a (canonically normalised) massless scalar field, up to a normalisation factor $\frac{M_{\text{Pl}}^2}{2}$. To quantise the gravitons we therefore can proceed exactly as above. We can promote $\gamma_s(\mathbf{k})$ to an operator and expand it in creation and annihilation operators,

$$\gamma_s(\mathbf{k}) = \frac{\sqrt{2}}{M_{\text{Pl}}} \left(f_{\mathbf{k}} a^s_{\mathbf{k}} + f^*_{\mathbf{k}} a^{s\dagger}_{\mathbf{k}} \right), \tag{9.32}$$

where for both signs $s = \pm$ the creation and annihilation operators satisfy the canonical commutation relations. If we now assume a de Sitter background we can explicitly compute the graviton power spectrum in the same way as we did for the massless scalar field. We find

$$\begin{aligned} \left\langle \gamma_{mn}(\mathbf{k})\gamma_{mn}(\mathbf{k}') \right\rangle =& \sum_{s,s'} \epsilon^s_{mn}(\mathbf{k})\epsilon^{s'}_{mn}(\mathbf{k}') \left\langle \gamma_s(\mathbf{k})\gamma_{s'}(\mathbf{k}') \right\rangle \\ =& \frac{2}{M_{\text{Pl}}^2} \sum_{s,s'} \epsilon^s_{mn}(\mathbf{k})\epsilon^{s'}_{mn}(\mathbf{k}')(2\pi)^3\delta(\mathbf{k}+\mathbf{k}')|f_{\mathbf{k}}|^2 \\ =& \frac{2}{M_{\text{Pl}}^2}\frac{H^2}{2k^3} \sum_{s,s'} \delta_{ss'}(2\pi)^3\delta(\mathbf{k}+\mathbf{k}')|f_{\mathbf{k}}|^2 \\ =& (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')P_T\,, \end{aligned} \tag{9.33}$$

with

$$P_T = \frac{4}{k^3}\frac{H^2}{M_{\text{Pl}}^2}\,. \tag{9.34}$$

This power spectrum provides us with a clear prediction for the CMB spectrum after a period of inflation starting from the Bunch-Davies vacuum. This spectrum can then be compared to the observational data from cosmological experiments and gives an excellent match. However, the match is not exact. This is to be expected, since in this course we take the coarse approximation that all fields are free. It turns out that this is an excellent approximation to predict the CMB radiation but recent experiment have nonetheless found tiny non-Gaussianities hidden in the observed radiation. To properly account for such effects one has to introduce interactions. This goes beyond the scope of these lectures, but we refer the interested reader to [Paj20] for more details.

Chapter 10

Quantum fields in AdS

de Sitter space is important because it describes the early and late time behaviour of the real universe. Anti de Sitter space is important for an entirely different reason. It is the background in which the holographic principle is best understood. Thanks to the isometries of AdS, the observables in these theories are constrained by the $SO(2,d)$ conformal group, even in the presence of mass deformations, allowing us to make progress in explicitly solving them. In addition, the AdS length scale provides a convenient IR regulator for interacting quantum field theories.

In many ways putting a quantum theory in anti de Sitter space makes things behave better. We can think of AdS as putting the system in a box. However, this will come at the cost of introducing some subtleties mostly resulting from the fact that AdS is not a globally hyperbolic space and hence we need to impose appropriate boundary conditions at the conformal boundary.

10.1 Geometry of AdS spaces

Anti de Sitter space is the maximally symmetric, negatively curved space. An easy way to construct it is as the isometric immersion of an $(d+1)$-dimensional hyperboloid in a $(d+2)$-dimensional ambient space. More concretely it is defined as the universal cover of the manifold

$$-\left(X^0\right)^2+\left(X^1\right)^2+\cdots+\left(X^d\right)^2-\left(X^{d+1}\right)^2=-L^2\,, \tag{10.1}$$

embedded in $\mathbb{R}^{d,2}$. The various coordinate systems are summarised in Appendix E, to which we refer the reader for more details. For concreteness let us here consider AdS space in global coordinates. In this case we choose the embedding coordinates as

$$\begin{aligned} X^0 &= L\cos t\cosh\rho \\ X^\mu &= L\,\omega^\mu\sinh\rho \\ X^{d+1} &= L\sin t\cosh\rho \end{aligned} \tag{10.2}$$

where ω^μ $(\mu=1,\ldots,d)$ parameterises a unit $(d-1)$—dimensional sphere. Doing so we obtain the metric with metric,

$$\mathrm{d}s^2=L_{\mathrm{AdS}}^2\left(\cosh^2\rho\,\mathrm{d}t^2-\mathrm{d}\rho^2-\sinh^2\rho\,\mathrm{d}\Omega_{d-1}^2\right)\,. \tag{10.3}$$

By taking $\rho\geq 0$ and $0\leq t<2\pi$, we cover the entire hyperboloid once. However, since the t direction has the topology of an S^1, this space however contains closed time-like curves. To restore the causality we can simply unwrap the S^1, i.e. take $-\infty<t<\infty$ in (10.3) to obtain the universal covering of the hyperboloid without closed time-like curves. This coordinate system can teach us several interesting facts about the geometry of anti de-Sitter space. For example, we can see that any light ray will reach spatial infinity in finite time t. Indeed, if we consider a light ray at constant position on the $(d-1)$-sphere we find

$$\Delta t = L_{\mathrm{AdS}} \int \mathrm{d}t = L_{\mathrm{AdS}} \int_0^\infty \frac{\mathrm{d}\rho}{\cosh\rho} = \frac{\pi L_{\mathrm{AdS}}}{2} \,. \tag{10.4}$$

This has the important implication that all observers in AdS space can communicate with each other from any point in space within finite (proper) time.

Causality and the conformal boundary

Notice that as we approach $\rho \to \infty$, the metric blows up. The locus $\rho = \infty$ is strictly not a part of AdS. However, as discussed in Chapter 3 we can consider the conformal compactification of AdS including the hypersurface $\mathscr{I}_{\mathrm{AdS}}$ at infinity. $\mathscr{I}_{\mathrm{AdS}}$ is also called the conformal boundary. More precisely, the conformal boundary is defined as the conformal equivalence class of metrics $\mathrm{d}\tilde{s}^2 = \mathrm{e}^{-2\rho}\mathrm{d}s^2$ with boundary $\mathbf{R}^{1,d-1}$ at $\rho = \infty$.

Particularly interesting for our purposes is the relation between the conformal compactification of AdS and flat space. It is well-known that Euclidean flat space can be compactified to the d-sphere, S^d by adding a point at infinity. On the other hand, Euclidean AdS$_{d+1}$, which is simply the hyperbolic space, can be conformally mapped to the $(d+1)$-dimensional disk. Therefore the boundary of the compactified Euclidean AdS space is the compactified Euclidean plane.

Similarly, in Lorentzian signature, by changing to hyperspherical coordinates, i.e. performing the coordinate change $\sinh\rho = \tan\theta$ from global coordinates, we obtain

$$\mathrm{d}s^2 = \frac{L^2_{\mathrm{AdS}}}{\cos^2\theta}\left(-\mathrm{d}t^2 + \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\Omega^2_{d-1}\right), \tag{10.5}$$

which after a conformal rescaling becomes the metric of the Einstein static universe. However, it is only half of the Einstein static universe since θ is restricted to the range $[0, \pi/2)$ rather then $[0, \pi)$. The boundary of this space is at $\theta = \pi/2$ and is given by $\mathbb{R} \times S^{d-1}$. This is identical to the conformal compactification of d-dimensional Minkowski space. This identification will play an essential role in the AdS/CFT correspondence.

Before moving on to considering quantum fields in AdS there are a few causal issues that need our attention. A peculiarity of AdS spaces is that an initially radially outward trajectory from $\rho = 0$ will begin to re-converge to its starting point following a period of $\frac{\pi}{2}$. Remembering the discussion in Chapter 3 a Lorentzian spacetime is globally hyperbolic if and only if it contains a Cauchy hypersurface, i.e. a hypersurface whose domain of dependence covers all of the spacetime. Famously, AdS is not globally hyperbolic. This becomes clear by looking at Figure 10.2, where the domain of dependence for the spatial hypersurface Σ is indicated by the blue diamonds.

Therefore, there are regions for which a knowledge of events on Σ does not allow any prediction. The underlying cause of these issues is the fact that conformal infinity in AdS is a time-like hypersurface. Indeed, as a massless particle can reach spatial infinity in finite time, it can then propagate along $\mathscr{I}$

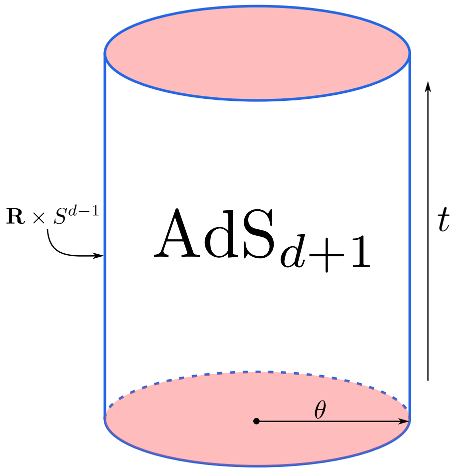


Figure 10.1: AdS_{d+1} can be conformally mapped into one half the Einstein static universe. This space has boundary $\mathbf{R}\times S^{d-1}$ which is exactly the conformal compactification of Minkowski space.

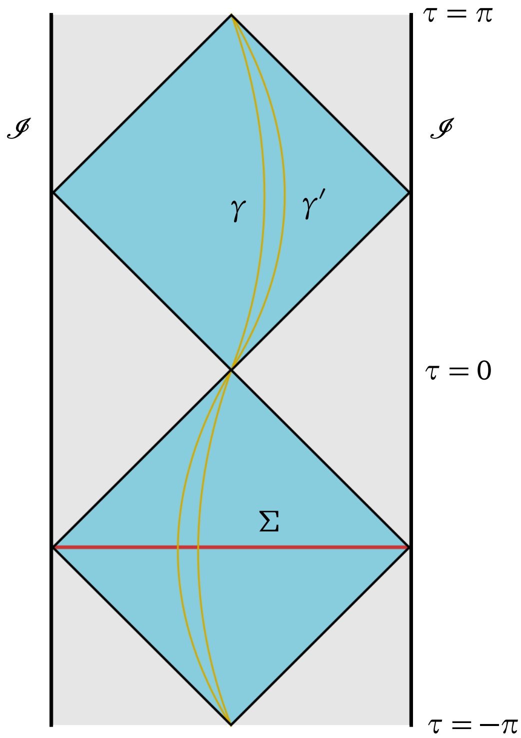


Figure 10.2: The Penrose diagram of AdS. The blue domain blue diamonds denote the domain of dependence of the spatial hypersurface Σ. Two geodesics, γ and γ' are denoted in yellow. In the universal covering the diagram continues indefinitely in the vertical direction.

and move outside of $D(\Sigma)$. In this way information is 'lost' from Σ to spatial infinity. Similarly AdS space allows for information to be introduced from spatial infinity.

To resolve these issues, and restore the predictability in the entire space, we have to provide boundary conditions at spatial infinity. In these lectures we want to obtain a closed system hence reflective boundary conditions are the natural choice [AIS78]. In the next section, when discussing the solutions to the wave equation we will discuss these boundary conditions at length. Note that such boundary conditions amount to requiring that there is no net flux across spatial infinity.

Asymptotically locally AdS spaces

When we consider matter fields coupled to gravity they will back-react on the metric and we will no longer have an exact AdS space. However, a lot of the machinery developed for AdS spaces will still be valid. By adding matter to the theory the bulk of AdS will change but the conformal boundary is a rather robust characteristic of spaces with a negative cosmological constant. It takes an infinite energy to change the asymptotics of such spaces. Therefore we will be able to extend the analysis of pure AdS to the class of asymptotically locally AdS (AlAdS) spaces. In particular they all have the same conformal boundary.

In Poincaré coordinates this statement implies that the metric of any AlAdS space near the conformal boundary is of the form

$$\mathrm{d}s^2 = \frac{L_{\mathrm{AdS}}^2}{z^2}\left(\mathrm{d}z^2 + g_{mn}(z,x)\mathrm{d}x^m\mathrm{d}x^n\right), \tag{10.6}$$

where $g_{mn}(z,x)$ is smooth and finite as $z \to 0$. This function can be expanded in powers of the radial coordinate near the conformal boundary as

$$g_{ab}(z,x) = \sum_{n=0}^{\infty} z^n g_{ab}^{(n)}(x). \tag{10.7}$$

Similarly, the matter fields coupled to gravity can be expanded near the conformal boundary. This expansion is called the Fefferman-Graham expansion.

10.2 Quantum fields in AdS

We have seen that AdS acts like a box for classical massive particles. Quantum mechanically, this confining potential gives rise to a discrete energy spectrum. Consider the Klein-Gordon equation

$$\left(\Box - m_\xi^2\right)\phi = 0, \tag{10.8}$$

in global coordinates. Since the Ricci scalar is a constant in AdS we define the effective mass m_ξ as

$$m_\xi^2 = m^2 + \xi L. \tag{10.9}$$

To avoid excessive notation we will mostly suppress the subscript ξ. To solve this problem we will use an indirect route which has the advantage that it makes the correspondence with holography more explicit. Consider the action of the quadratic Casimir of the AdS isometry group on a scalar field[1]

$$\frac{1}{2}J_{AB}J^{BA}\phi = \left[-X^2\partial_X^2 + X\cdot\partial_X\left(d + X\cdot\partial_X\right)\right]\phi. \tag{10.10}$$

[1]Formally, we are extending the function ϕ from AdS, defined by the hypersurface $X^2 = -L^2$, to the embedding space. However, the action of the quadratic Casimir is independent of this extension because the generators J_{AB} are interior to AdS, i.e. $\left[J_{AB}, X^2 + L^2\right] = 0$.

By foliating the embedding space $\mathbb{R}^{2,d}$ with AdS surfaces of different AdS radii L, we can obtain the Laplacian in the embedding space as

$$\partial_X^2 = -\frac{1}{L^{d+1}}\frac{\partial}{\partial L}L^{d+1}\frac{\partial}{\partial L} + \Box_{\text{AdS}} \ . \tag{10.11}$$

Substituting this in (10.10) and noticing that $X \cdot \partial_X = L\partial_L$ we conclude that

$$\frac{1}{2}J_{AB}J^{BA}\phi = L^2\Box_{\text{AdS}}\phi \ . \tag{10.12}$$

Therefore, we should identify $m_\xi^2 L^2$ with the quadratic Casimir of the conformal group. The Lorentzian version of the conformal generators is

$$D = -J_{0,d+1} \ , \qquad P_\mu = J_{\mu 0} + i\,J_{\mu,d+1} \ , \tag{10.13}$$

$$M_{\mu\nu} = J_{\mu\nu} \ , \qquad K_\mu = J_{\mu 0} - i\,J_{\mu,d+1} \ . \tag{10.14}$$

Exercise 10.1. *Show that, in global coordinates, the conformal generators take the form*

$$\begin{aligned}
D &= i\frac{\partial}{\partial t} \ , \\
M_{\mu\nu} &= -i\left(\omega_\mu \frac{\partial}{\partial \omega^\nu} - \omega_\nu \frac{\partial}{\partial \omega^\mu}\right) \ , \\
P_\mu &= -ie^{-it}\left[\omega_\mu\left(\partial_\rho - i\tanh\rho\,\partial_t\right) + \frac{1}{\tanh\rho}\nabla_\mu\right] \ , \\
K_\mu &= ie^{it}\left[\omega_\mu\left(-\partial_\rho - i\tanh\rho\,\partial_t\right) - \frac{1}{\tanh\rho}\nabla_\mu\right] \ ,
\end{aligned}$$

where $\nabla_\mu = \frac{\partial}{\partial \omega^\mu} - \omega_\mu \omega^\nu \frac{\partial}{\partial \omega^\nu}$ *is the covariant derivative on the unit sphere* S^{d-1}.

In analogy with CFT construction we can look for primary states, which are annihilated by K_μ and are eigenstates of the Hamiltonian, $D\phi = \Delta\phi$. The condition $K_\mu\phi = 0$ splits in one term proportional to ω_μ and one term orthogonal to ω_μ. The second term implies that ϕ is independent of the angular variables ω^μ while the first term reduces to $\left(\partial_\rho + \Delta\tanh\rho\right)\phi = 0$. This implies that

$$\phi \propto \left(\frac{e^{-it}}{\cosh\rho}\right)^\Delta = \left(\frac{R}{X^0 - X^{d+1}}\right)^\Delta \ . \tag{10.15}$$

This is the lowest energy state with eigenvalue Δ. Starting from this state one can construct excited states by acting on it with the generator P_μ. Notice that all such states have the same value for the quadratic Casimir

$$\frac{1}{2}J_{AB}J^{BA}\phi = \Delta(\Delta - d)\phi \ . \tag{10.16}$$

Hence in this way we can generate all normalisable solutions of the Klein-Gordon equation with $m^2L^2 = \Delta(\Delta - d)$. This shows that the one-particle energy spectrum is given by $\omega = \Delta + l + 2n$ where $l = 0, 1, 2, \ldots$ is the spin, generated by acting with the traceless generators, $P_{\mu_1} \ldots P_{\mu_l}$ − traces, and similarly, the quantum number $n = 0, 1, 2, \ldots$ is generated by acting with $\left(P^2\right)^n$.

Note that, for a given (effective) mass we can solve the equation $m^2L^2 = \Delta(\Delta - d)$ as

$$\Delta = \Delta_\pm = \frac{d}{2} \pm \sqrt{\frac{d^2}{4} + m^2 L^2} \, . \tag{10.17}$$

Demanding that $\Delta_\pm \in \mathbb{R}$ requires having $m \geq -\frac{d}{2L}$. Hence, we see that a range of tachyonic masses is allowed. In Minkowski space this would lead to an instability of the perturbative vacuum. In AdS space, whenever the mass-squared lies above this bound the free energy of the field is bounded from below and no instabilities arise. This bound is called the Breitenlohner-Freedman bound after [BF82].

Exercise 10.2. *Given the symmetry of the metric (E.46) we can look for solutions of the form*

$$\phi = e^{i\omega t} Y_l(\Omega) F(r) \, , \tag{10.18}$$

where $Y_l(\Omega)$ is a spherical harmonic with eigenvalue $-l(l+d-2)$ of the Laplacian on the unit sphere S^{d-1}. Derive a differential equation for $F(r)$ and show that it is solved by

$$F(r) = (\cos r)^\Delta (\sin r)^l \; {}_2F_1\left(\frac{l+\Delta-\omega}{2}, \frac{l+\Delta+\omega}{2}, l + \frac{d}{2}, \sin r\right) \, , \tag{10.19}$$

with $2\Delta = d + \sqrt{d^2 + 4(mL)^2}$. We chose this solution because it is smooth at $r = 0$. Now we also need to impose another boundary condition at the boundary of AdS*, i.e. $r = \frac{\pi}{2}$. Imposing that there is no energy flux through the boundary leads to the quantization of the energies $|\omega| = \Delta + l + 2n$ with $n = 0, 1, 2, \ldots$ (see for example [AGM+00]).*

If there are no interactions between the particles in AdS, then the Hilbert space has a Fock representation and the energy of a multi-particle state is just the sum of the energies of particles. Turning on small interactions leads to small energy shifts of the multi-particle states.

Green's functions

For the computation of the Green's function, let us now return to Euclidean signature and afterwards analytically continue the result to obtain the Lorentzian Feynman Green's function. For simplicity, we consider a free scalar field with action

$$S = \frac{1}{2} \int_{AdS} \mathrm{d}^{d+1}x \left[(\mathrm{d}\phi)^2 + m^2 \phi^2 \right] \, . \tag{10.20}$$

Similar as for the de Sitter space, we can obtain the two-point function $\langle \phi(X)\phi(Y) \rangle$ by exploiting the conformal symmetry of AdS. The Euclidean Green's function denoted by the propagator $\Pi(X,Y)$, has to obey,

$$\left[\Box_X - m^2\right] \Pi(X,Y) = -\delta(X,Y) \, . \tag{10.21}$$

From the symmetry of the problem it is clear that the propagator can only depend on the invariant $X \cdot Y$ or equivalently on the chordal distance $\zeta = (X-Y)^2/L^2$. From now on we will set $L = 1$ and all lengths will be expressed in units of the AdS radius.

Exercise 10.3. *Use (10.10) and (10.12) to show that*

$$\Pi(X,Y) = \frac{\mathcal{C}_\Delta}{\zeta^\Delta}\, {}_2F_1\left(\Delta, \Delta - \frac{d}{2} + \frac{1}{2}, 2\Delta - d + 1, -\frac{4}{\zeta}\right) , \tag{10.22}$$

where $2\Delta = d + \sqrt{d^2 + (2m)^2}$ *and*

$$\mathcal{C}_\Delta = \frac{\Gamma(\Delta)}{2\pi^{\frac{d}{2}}\Gamma\left(\Delta - \frac{d}{2} + 1\right)} . \tag{10.23}$$

For a free field, higher point functions are simply given by Wick contractions. For example,

$$\begin{aligned}\left\langle \phi(X_1)\phi(X_2)\phi(X_3)\phi(X_4)\right\rangle = \Pi(X_1,X_2)\Pi(X_3,X_4) + \Pi(X_1,X_3)\Pi(X_2,X_4) \\ + \Pi(X_1,X_4)\Pi(X_2,X_3) .\end{aligned} \tag{10.24}$$

Weak interactions of ϕ can be treated perturbatively. Suppose the action includes a cubic term,

$$S = \int_{AdS} dX \left[\frac{1}{2}(\nabla\phi)^2 + \frac{1}{2}m^2\phi^2 + \frac{1}{3!}g\phi^3\right] . \tag{10.25}$$

Then, there is a non-vanishing three-point function

$$\langle\phi(X_1)\phi(X_2)\phi(X_3)\rangle = -g\int_{AdS} dY\, \Pi(X_1,Y)\Pi(X_2,Y)\Pi(X_3,Y) + O(g^3) ,$$

and a connected part of the four-point function of order g^2. This is very similar to perturbative QFT in flat space.

10.3 Towards a conformal theory on the boundary

Given a correlation function in AdS we can consider the limit where we send all points to infinity. More precisely, we introduce

$$\mathcal{O}(P) = \frac{1}{\sqrt{C_\Delta}} \lim_{\lambda\to\infty} \lambda^\Delta\, \phi\,(X = \lambda P + \dots), \tag{10.26}$$

where P is a future directed null vector in $\mathbb{R}^{d+1,1}$ and the $\dots$ denote terms that do not grow with λ whose only purpose is to enforce the AdS condition $X^2 = -1$. In other words, the operator $\mathcal{O}(P)$ is the limit of the field $\phi(X)$ when X approaches the boundary point P of AdS. Correlation functions of $\mathcal{O}$ are naturally defined by the limit of correlation functions of ϕ in AdS. For example, the two-point function is given by

$$\langle\mathcal{O}(P_1)\mathcal{O}(P_2)\rangle = \frac{1}{(-2P_1\cdot P_2)^\Delta} + O(g^2) , \tag{10.27}$$

which is exactly the CFT two-point function of a primary operator of dimension Δ. The three-point function is given by

$$\langle \mathcal{O}(P_1)\mathcal{O}(P_2)\mathcal{O}(P_3)\rangle = -g\, C_\Delta^{-\frac{3}{2}} \int_{AdS} dX\, \Pi(X,P_1)\Pi(X,P_2)\Pi(X,P_3) + O(g^3)\,, \tag{10.28}$$

where

$$\Pi(X,P) = \lim_{\lambda\to\infty} \lambda^\Delta\, \Pi\,(X, Y=\lambda P+\dots) = \frac{C_\Delta}{(-2P\cdot X)^\Delta} \tag{10.29}$$

is the bulk to boundary propagator.

Exercise 10.4. *Write the bulk to boundary propagator in Poincaré coordinates.*

Exercise 10.5. *Compute the following generalization of the integral in (10.28),*

$$\int_{AdS} dX \prod_{i=1}^{3} \frac{1}{(-2P_i\cdot X)^{\Delta_i}}\,, \tag{10.30}$$

and show that it reproduces the expected formula for the CFT three-point function $\langle \mathcal{O}_1(P_1)\mathcal{O}_2(P_2)\mathcal{O}_3(P_3)\rangle$. It is helpful to use the integral representation

$$\frac{1}{(-2P\cdot X)^\Delta} = \frac{1}{\Gamma(\Delta)} \int_0^\infty \frac{ds}{s} s^\Delta e^{2sP\cdot X} \tag{10.31}$$

to bring the AdS integral to the form

$$\int_{AdS} dX e^{2X\cdot Q} \tag{10.32}$$

with Q a future directed timelike vector. Choosing the X^0 direction along Q and using the Poincaré coordinates (E.42) it is easy to show that

$$\int_{AdS} dX e^{2X\cdot Q} = \pi^{\frac{d}{2}} \int_0^\infty \frac{dz}{z} z^{-\frac{d}{2}} e^{-z+Q^2/z}\,. \tag{10.33}$$

To factorize the remaining integrals over s_1, s_2, s_3 and z it is helpful to change to the variables t_1, t_2, t_3 and z using

$$s_i = \frac{\sqrt{z\, t_1 t_2 t_3}}{t_i}\,. \tag{10.34}$$

State-Operator Map

We have seen that the correlation functions of the boundary operator (10.26) have the correct homogeneity property and $SO(d+1,1)$ invariance expected of CFT correlators of a primary scalar operator with scaling dimension Δ. We will now argue that they also obey an associative OPE. The argument is very similar to the one used in CFT. We think of the correlation functions as vacuum expectation values of time ordered products

$$\langle \phi(X_1)\phi(X_2)\phi(X_3)\dots\rangle = \langle 0|\dots\hat{\phi}(\tau_3,\rho_3,\Omega_3)\hat{\phi}(\tau_2,\rho_2,\Omega_2)\hat{\phi}(\tau_1,\rho_1,\Omega_1)\,|0\rangle\;,$$

where we assumed $\tau_1 < \tau_2 < 0 < \tau_3 < \dots$. We then insert a complete basis of states at $\tau = 0$,

$$\begin{aligned} &\langle \phi(X_1)\phi(X_2)\phi(X_3)\dots\rangle \\ &= \sum_{\psi} \langle 0| \dots \hat{\phi}(\tau_3,\rho_3,\Omega_3)|\psi\rangle \langle\psi| \hat{\phi}(\tau_2,\rho_2,\Omega_2)\hat{\phi}(\tau_1,\rho_1,\Omega_1)|0\rangle \ . \end{aligned} \tag{10.35}$$

Using $\hat{\phi}(\tau,\rho,\Omega) = e^{\tau D}\hat{\phi}(0,\rho,\Omega)e^{-\tau D}$ and choosing an eigenbasis of the Hamiltonian $D = -\frac{\partial}{\partial\tau}$ it is clear that the sum converges for the assumed time ordering. The next step, is to establish a one-to-one map between the states $|\psi\rangle$ and boundary operators. It is clear that every boundary operator (10.26) defines a state. Inserting the boundary operator at $P^A = \left(P^0, P^\mu, P^{d+1}\right) = \left(\frac{1}{2}, 0, \frac{1}{2}\right)$, which is the boundary point defined by $\tau \to -\infty$ in global coordinates, we can write

$$\langle \dots \phi(X_3)\mathcal{O}(P)\rangle = \langle 0|\dots\hat{\phi}(\tau_3,\rho_3,\Omega_3)|\mathcal{O}\rangle \ , \tag{10.36}$$

where

$$\begin{aligned} |\mathcal{O}\rangle &= \lim_{\tau\to-\infty} \left(e^{-\tau}\cosh\rho\right)^{\Delta} \hat{\phi}(\tau,\rho,\Omega)|0\rangle \\ &= \sum_{\psi} |\psi\rangle (\cosh\rho)^{\Delta} \lim_{\tau\to-\infty} \langle\psi| e^{\tau(D-\Delta)}\hat{\phi}(0,\rho,\Omega)|0\rangle \ . \end{aligned} \tag{10.37}$$

The limit $\tau \to -\infty$ projects onto the primary state with wave function (10.15).

The map from states to boundary operators can be established using global time translation invariance,

$$\begin{aligned} &\langle 0|\dots\hat{\phi}(\tau_3,\rho_3,\Omega_3)|\psi(0)\rangle \\ &= \lim_{\tau\to-\infty} \langle 0|\dots\hat{\phi}(\tau_3,\rho_3,\Omega_3)e^{\tau D}|\psi(\tau)\rangle \equiv \left\langle \dots\phi(X_3)\mathcal{O}_\psi(P)\right\rangle \end{aligned} \tag{10.38}$$

where $|\psi(\tau)\rangle = e^{-\tau D}|\psi\rangle$ and $P^A = \left(\frac{1}{2}, 0, \frac{1}{2}\right)$ is again the boundary point defined by $\tau \to -\infty$ in global coordinates. The idea is that $|\psi(\tau)\rangle$ prepares a boundary condition for the path integral on a surface of constant τ and this surface converges to a small cap around the boundary point $P^A = \left(\frac{1}{2}, 0, \frac{1}{2}\right)$ when $\tau \to -\infty$. This is depicted in figure 10.3.

The Hilbert space of the bulk theory can be decomposed in irreducible representations of the isometry group $SO(d+1,1)$. These are the highest weight representations of the conformal group, labelled by the scaling dimension and $SO(d)$ irrep of the the primary state. Therefore, the CFT conformal block decomposition of correlators follows from the partial wave decomposition in AdS, i.e. the decomposition in intermediate eigenstates of the Hamiltonian organized in irreps of the isometry group $SO(d+1,1)$. For example, the conformal block decomposition of the disconnected part of the four-point function,

$$\left\langle \mathcal{O}(P_1)\dots\mathcal{O}(P_4)\right\rangle = \frac{1}{\left(P_{12}P_{34}\right)^{\Delta}} + \frac{1}{\left(P_{13}P_{24}\right)^{\Delta}} + \frac{1}{\left(P_{14}P_{23}\right)^{\Delta}} \ , \tag{10.39}$$

where $P_{ij} = -2P_i \cdot P_j$, is given by a sum of conformal blocks associated with the vacuum and

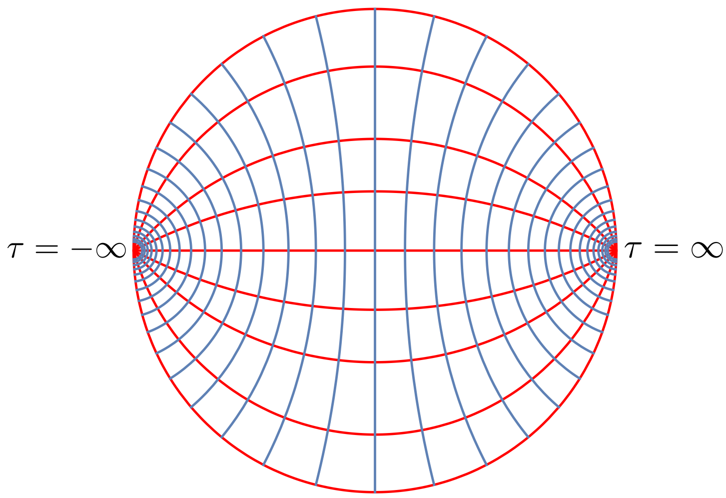


Figure 10.3: Curves of constant τ (in blue) and constant ρ (in red) for AdS_2 stereographically projected to the unit disk (Poincaré disk). This shows how surfaces of constant τ converge to a boundary bound when $\tau \to -\infty$. The cartesian coordinates in the plane of the figure are given by $\vec{w} = \frac{(\cosh\rho \sinh\tau, \sinh\rho)}{1+\cosh\rho\cosh\tau}$ which puts the AdS_2 metric in the form $ds^2 = \frac{4d\vec{w}^2}{1-\vec{w}^2}$.

two-particle intermediate states

$$\left\langle \mathcal{O}(P_1)\dots\mathcal{O}(P_4)\right\rangle = G_{0,0}(P_1,\dots,P_4) + \sum_{\substack{l=0\\ even}}^{\infty}\sum_{n=0}^{\infty} c_{n,l}\, G_{2\Delta+2n+l,l}(P_1,\dots,P_4)\,.$$

Exercise 10.6. *Check this statement in $d=2$ using the formula [DO01]*

$$G_{E,l}(P_1,P_2,P_3,P_4) = \frac{k(E+l,z)k(E-l,\bar{z}) + k(E-l,z)k(E+l,\bar{z})}{(-2P_1\cdot P_2)^{\Delta}\left(-2P_3\cdot P_4\right)^{\Delta}\left(1+\delta_{l,0}\right)} \tag{10.40}$$

where

$$k(2\beta,z) = (-z)^{\beta}\ {}_2F_1(\beta,\beta,2\beta,z)\,. \tag{10.41}$$

Determine the coefficients $c_{n,l}$ for $n \leq 1$ by matching the Taylor series expansion around $z=\bar{z}=0$.

Extra: using a computer you can compute many coefficients and guess the general formula.

Generating function

There is an equivalent way of defining CFT correlation functions from QFT in AdS. We introduce the generating function

$$W[\phi_b] = \left\langle e^{\int_{\partial AdS} dP \phi_b(P)\mathcal{O}(P)}\right\rangle\,, \tag{10.42}$$

where the integral over ∂AdS denotes an integral over a chosen section of the null cone in $\mathbb{R}^{d+1,1}$ with its induced measure. We impose that the source obeys $\phi_b(\lambda P) = \lambda^{\Delta-d}\phi_b(P)$ so that the integral is invariant under a change of section, i.e. conformal invariant. For example, in the Poincaré section the integral reduces to $\int d^d x \phi_b(x)\mathcal{O}(x)$. Correlation functions are easily obtained with functional derivatives

$$\langle \mathcal{O}(P_1)\dots\mathcal{O}(P_n)\rangle = \frac{\delta}{\delta\phi_b(P_1)}\cdots\frac{\delta}{\delta\phi_b(P_n)}W[\phi_b]\bigg|_{\phi_b=0} . \tag{10.43}$$

If we set the generating function to be equal to the path integral over the field ϕ in AdS

$$W[\phi_b] = \frac{\int_{\phi\to\phi_b}[d\phi]\,e^{-S[\phi]}}{\int_{\phi\to 0}[d\phi]\,e^{-S[\phi]}} , \tag{10.44}$$

with the boundary condition that it approaches the source ϕ_b at the boundary,

$$\lim_{\lambda\to\infty}\lambda^{d-\Delta}\phi(X=\lambda P+\dots) = \frac{1}{2\Delta-d}\frac{1}{\sqrt{C_\Delta}}\phi_b(P) , \tag{10.45}$$

then we recover the correlation functions of $\mathcal{O}$ defined above as limits of the correlation functions of ϕ.

For a quadratic bulk action, the ratio of path integrals in (10.44) is given e^{-S} computed on the classical solution obeying the required boundary conditions. A natural guess for this solution is

$$\phi(X) = \sqrt{C_\Delta}\int_{\partial AdS} dP\,\frac{\phi_b(P)}{(-2P\cdot X)^\Delta} . \tag{10.46}$$

This automatically solves the AdS equation of motion $\nabla^2\phi = m^2\phi$, because it is an homogeneous function of weight $-\Delta$ and it obeys $\partial_A\partial^A\phi = 0$ in the embedding space (see equations (10.10) and (10.12)). To see that it also obeys the boundary condition (10.45) it is convenient to use the Poincaré section.

Exercise 10.7. *In the Poincaré section (G.31) and using Poincaré coordinates (E.42), formula (10.46) reads*

$$\phi(z,x) = \sqrt{C_\Delta}\int d^d y\,\frac{z^\Delta\phi_b(y)}{(z^2+(x-y)^2)^\Delta} \tag{10.47}$$

and (10.45) reads

$$\lim_{z\to 0} z^{\Delta-d}\phi(z,x) = \frac{1}{2\Delta-d}\frac{1}{\sqrt{C_\Delta}}\phi_b(x) . \tag{10.48}$$

Show that (10.48) follows from (10.47). You can assume $2\Delta > d$.

The cubic term $\frac{1}{3!}g\phi^3$ in the action will lead to (calculable) corrections of order g in the classical solution (10.46). To determine the generating function $W[\phi_b]$ in the classical limit we just have to compute the value of the bulk action (10.25) on the classical solution. However, before doing that, we have to address a small subtlety. We need to add a boundary term to the action (10.25) in order to have a well posed variational problem.

Exercise 10.8. *The coefficient β should be chosen such that the quadratic action* [2]

$$S_2 = \int_{AdS} dw\sqrt{G}\left[\frac{1}{2}(\nabla\phi)^2 + \frac{1}{2}m^2\phi^2\right] + \beta\int_{AdS} dw\sqrt{G}\,\nabla_\alpha(\phi\nabla^\alpha\phi) \tag{10.49}$$

[2] Here w stands for a generic coordinate in AdS and the index α runs over the $d+1$ dimensions of AdS.

is stationary around a classical solution obeying (10.48) for any variation $\delta\phi$ that preserves the boundary condition, i.e.

$$\delta\phi(z,x) = z^{\Delta}[f(x) + O(z)] \ . \tag{10.50}$$

Show that $\beta = \frac{\Delta-d}{d}$ and that the on-shell action is given by a boundary term

$$S_2 = \frac{2\Delta - d}{2d} \int_{AdS} dw \sqrt{g}\, \nabla_\alpha (\phi \nabla^\alpha \phi) \ . \tag{10.51}$$

Finally, show that for the classical solution (10.47) this action is given by [3]

$$S_2 = -\frac{1}{2} \int d^d y_1 d^d y_2 \phi_b(y_1) \phi_b(y_2) K(y_1, y_2) \,, \tag{10.52}$$

where

$$\begin{aligned} K(y_1,y_2) &= C_\Delta \frac{2\Delta - d}{d} \lim_{z\to 0} \int \frac{d^d x}{z^{d-1}} \frac{z^\Delta}{(z^2 + (x-y_1)^2)^\Delta} \partial_z \frac{z^\Delta}{(z^2 + (x-y_2)^2)^\Delta} \\ &= \frac{1}{(y_1 - y_2)^{2\Delta}} \end{aligned} \tag{10.53}$$

is the CFT two point function (10.27).

Exercise 10.9. *Using $\phi = \phi_0 + O(g)$ with ϕ_0 given by (10.46), show that the complete on-shell action is given by*

$$S = -\frac{1}{2} \int d^d y_1 d^d y_2 \phi_b(y_1) \phi_b(y_2) K(y_1, y_2) + \frac{1}{3!} g \int_{AdS} dX \, [\phi_0(X)]^3 + O(g^2) \,,$$

and that this leads to the three-point function (10.28).

Extra: Compute the terms of $O(g^2)$ in the on-shell action.

10.4 The AdS/CFT correspondence

We have seen that QFT on an AdS background naturally defines conformal correlation functions living on the boundary of AdS. Moreover, we saw that a weakly coupled theory in AdS gives rise to factorization of CFT correlators like in a large N expansion. However, there is one missing ingredient to obtain a full-fledged CFT: a stress-energy tensor.

To add this will radically change the theory though, as it turns out this requires us to make gravity dynamical in the bulk. Indeed, if we have some conserved current J_μ for a global current in the boundary theory this couples to bulk fields as

$$S = S + \int_{\partial \text{ AdS}} J^\mu A_\mu \,, \tag{10.54}$$

[3]This integral is divergent if the source ϕ_b is a smooth function and $\Delta > \frac{d}{2}$. The divergence comes from the short distance limit $y_1 \to y_2$ and does not affect the value of correlation functions at separate points. Notice that a small value of $z > 0$ provides a UV regulator.

where A_μ becomes a dynamic gauge field in the bulk. Repeating this for the stress tensor we find that it couples as

$$S = S + \int_{\partial\,\mathrm{AdS}} T^{\mu\nu} g_{\mu\nu}\,, \tag{10.55}$$

where the symmetric tensor $g_{\mu\nu}$ is the dynamic bulk metric! Having a stress tensor on the boundary therefore automatically induces dynamical gravity in the bulk!

The next exercise also shows that a free QFT in AdS$_{d+1}$ can not be dual to a local CFT$_d$.

Exercise 10.10. *Compute the free-energy of a gas of free scalar particles in AdS. Since particles are free and bosonic one can create multi-particle states by populating each single particle state an arbitrary number of times. That means that the total partition function is a product over all single particle states and it is entirely determined by the single particle partition function. More precisely, show that*

$$F = -T\log Z = -T\log\prod_{\psi_{sp}}\left(\sum_{k=0}^{\infty} q^{kE_{\psi_{sp}}}\right) = -T\sum_{n=1}^{\infty}\frac{1}{n}Z_1(q^n)\ , \tag{10.56}$$

$$Z_1(q) = \sum_{\psi_{sp}} q^{E_{\psi_{sp}}} = \frac{q^\Delta}{(1-q)^d}\ , \tag{10.57}$$

where $q = e^{-\frac{1}{RT}}$ and we have used the single-particle spectrum of the hamiltonian $D = -\frac{\partial}{\partial\tau}$ of AdS in global coordinates. Show that

$$F \approx -\zeta(d+1)R^d T^{d+1} \tag{10.58}$$

in the high temperature regime and compute the entropy using the thermodynamic relation $S = -\frac{\partial F}{\partial T}$. Compare this result with the expectation

$$S \sim (RT)^{d-1}\,, \tag{10.59}$$

for the high temperature behaviour of the entropy of a CFT on a sphere S^{d-1} of radius R. See section 4.3 of reference [ESP12] for more details on this point.

Gravity with AdS boundary conditions

Consider general relativity in the presence of a negative cosmological constant

$$I[G] = \frac{1}{\ell_P^{d-1}}\int d^{d+1}w\sqrt{G}\,[\mathcal{R} - 2\Lambda]\ . \tag{10.60}$$

The AdS geometry

$$ds^2 = R^2\frac{dz^2 + dx_\mu dx^\mu}{z^2}\ , \tag{10.61}$$

is a maximally symmetric classical solution with $\Lambda = -\frac{d(d-1)}{2R^2}$. When the AdS radius R is much larger than the Planck length ℓ_P the metric fluctuations are weakly coupled and form an approximate Fock space of graviton states. One can compute the single graviton states and verify that they are in one-to-one correspondence with the CFT stress-tensor operator and its descendants (with AdS energies matching scaling dimensions). One can also obtain CFT correlation functions of the stress-energy tensor using Witten diagrams in AdS. The new ingredients are the bulk to boundary and bulk to bulk graviton propagators [LT98, LT99, DFM+99a, DFM+99b, CGP14].

In the gravitational context, it is nicer to use the partition function formulation

$$Z[g_{\mu\nu},\phi_b]=\int_{\substack{G\to g\\ \phi\to\phi_b}}[dG][d\phi]\,e^{-I[G,\phi]} \tag{10.62}$$

where

$$I[G,\phi]=\frac{1}{\ell_P^{d-1}}\int d^{d+1}w\sqrt{G}\left[\mathcal{R}-2\Lambda+\frac{1}{2}(\nabla\phi)^2+\frac{1}{2}m^2\phi^2\right] \tag{10.63}$$

and the boundary condition are

$$\begin{aligned} ds^2 &= G_{\alpha\beta}dw^\alpha dw^\beta=R^2\frac{dz^2+dx^\mu dx^\nu\left[g_{\mu\nu}(x)+O(z)\right]}{z^2}\,, \\ \phi &= \frac{z^{d-\Delta}}{2\Delta-d}\left[\phi_b(x)+O(z)\right]\,. \end{aligned} \tag{10.64}$$

By construction the partition function is invariant under diffeomorphisms of the boundary metric $g_{\mu\nu}$. Therefore, this definition implies the Ward identity (G.19). The generating function is also invariant under Weyl transformations

$$Z\left[\Omega^2 g_{\mu\nu},\Omega^{\Delta-d}\phi_b\right]=Z\left[g_{\mu\nu},\phi_b\right] \qquad (naive) \tag{10.65}$$

This follows from the fact that the boundary condition

$$\begin{aligned} ds^2 &= R^2\frac{dz^2+dx^\mu dx^\nu\left[\Omega^2(x)g_{\mu\nu}(x)+O(z)\right]}{z^2} \\ \phi &= \frac{z^{d-\Delta}}{2\Delta-d}\left[\Omega^{\Delta-d}(x)\phi_b(x)+O(z)\right] \end{aligned} \tag{10.66}$$

can be mapped to (10.64) by the following coordinate transformation

$$\begin{aligned} z &\to z\,\Omega-\frac{1}{4}z^3\Omega\left(\partial_\mu\log\Omega\right)^2+O(z^5) \\ x^\mu &\to x^\mu-\frac{1}{2}z^2\partial^\mu\log\Omega+O(z^4) \end{aligned} \tag{10.67}$$

where indices are raised and contracted using the metric $g_{\mu\nu}$ and its inverse. In other words, a bulk geometry that satisfies (10.64) also satisfies (10.66) with an appropriate choice of coordinates. If the partition function (10.62) was a finite quantity this would be the end of the story. However, even in the classical limit, where $Z\approx e^{-I}$, the partition function needs to be regulated. The divergences originate from the $z\to 0$ region and can be regulated by cutting off the bulk integrals at $z=\epsilon$ (as it happened for the scalar case discussed above). Since the coordinate transformation (10.67) does not preserve the cutoff, the regulated partition function is not obviously Weyl invariant. This has been studied in great detail in the context of holographic renormalisation [?, ?]. In particular, it leads to the Weyl anomaly $g^{\mu\nu}T_{\mu\nu}\neq 0$ when d is even. The crucial point is that this is a UV effect that does not affect the connected correlation functions of operators at separate points. In particular, the integrated form (G.22)=(G.24) of the conformal Ward identity is valid.

We do not now how to define the quantum gravity path integral in (10.62). The best we can do is a semiclassical expansion when $\ell_P \ll R$. This semiclassical expansion gives rise to connected correlators of the stress tensor $T_{\mu\nu}$ that scale as

$$\left\langle T_{\mu_1\nu_1}(x_1)\dots T_{\mu_n\nu_n}(x_n)\right\rangle_c \sim \left(\frac{R}{\ell_P}\right)^{d-1} . \tag{10.68}$$

This is exactly the scaling (G.47) we found from large N factorization if we identify $N^2 \sim \left(\frac{R}{\ell_P}\right)^{d-1}$. This suggests that CFTs related to semiclassical Einstein gravity in AdS, should have a large number of local degrees of freedom. This can be made more precise. The two-point function of the stress tensor in a CFT is given by

$$\left\langle T_{\mu\nu}(x)T_{\sigma\rho}(0)\right\rangle = \frac{C_T}{S_d^2}\frac{1}{x^{2d}}\left[\frac{1}{2}I_{\mu\sigma}I_{\nu\rho}+\frac{1}{2}I_{\mu\rho}I_{\nu\sigma}-\frac{1}{d}\delta_{\mu\nu}\delta_{\sigma\rho}\right] , \tag{10.69}$$

where $S_d=\frac{2\pi^{d/2}}{\Gamma(d/2)}$ is the volume of a $(d-1)$-dimensional unit sphere and

$$I_{\mu\nu}=\delta_{\mu\nu}-2\frac{x_\mu x_\nu}{x^2} . \tag{10.70}$$

The constant C_T provides an (approximate) measure of the number of degrees of freedom.[4] For instance, for n_φ free scalar fields and n_ψ free Dirac fields we find [?]

$$C_T = n_\varphi\frac{d}{d-1}+n_\psi 2^{\left[\frac{d}{2}\right]-1}d , \tag{10.71}$$

where $[x]$ is the integer part of x. If the CFT is described by Einstein gravity in AdS, we find [?]

$$C_T = 8\frac{d+1}{d-1}\frac{\pi^{\frac{d}{2}}\Gamma(d+1)}{\Gamma^3\left(\frac{d}{2}\right)}\frac{R^{d-1}}{\ell_P^{d-1}} , \tag{10.72}$$

which shows that the CFT dual of a semiclassical gravitational theory with $R \gg \ell_P$, must have a very large number of degrees of freedom.

In summary, semiclassical gravity with AdS boundary conditions gives rise to a set of correlation functions that have all the properties (conformal invariance, Ward identities, large N factorization) expected for the correlation functions of the stress tensor of a large N CFT. Therefore, it is natural to ask if a CFT with finite N is a quantum theory of gravity.

The AdS/CFT dictionary

Work in progress...

[4]However, for $d > 2$, C_T is not a c-function that always decreases under Renormalization Group flow.

10.5 **Hawking radiation in** AdS

In this last section, which is left entirely as an exercise, we derive the Hawking effect for black holes in AdS space.

Exercise 10.11. *Consider the following metric for a black hole in* AdS_{d+1} *space,*

$$\mathrm{d}s^2 = f(r)\mathrm{d}t^2 - \frac{\mathrm{d}r}{f(r)} + r^2 \mathrm{d}\Omega_{d-1}^2 \,, \tag{10.73}$$

where $f(r) = r^2 + 1 - \frac{\mu}{r^{d-2}}$*. The mass is related to the parameter* μ *as*

$$(d-1)\mu = 8\pi^{\frac{2-d}{2}} \Gamma\left(\frac{d}{2}\right) M \,. \tag{10.74}$$

Argue why this metric represents a black hole and show that asymptotically it reduces to a metric on standard AdS_{d+1}.

To analyse the Hawking effect in this background, define the tortoise coordinate satisfying $\mathrm{d}r^* = \frac{\mathrm{d}r}{f(r)}$ *which tends to* $-\infty$ *near the horizon and approaches a constant value near the asymptotic boundary.*

In AdS*, it is natural to impose normalizable boundary conditions on the field. Given a scalar field,* ϕ*, of mass* m*, we demand that, as* $r \to \infty$*, the field dies off as* r^{Δ_+}*, where* $2\Delta_+ = d + \sqrt{d^2 + 4m^2}$*. The relevance of this boundary condition is that it relates the left- and right-moving modes near the boundary. Show that this implies that at late times we can expand the field as*

$$\phi = \sum_l \int \mathrm{d}\omega a_{\omega l} f(\omega, l, r^*) \mathrm{e}^{-i\omega t} Y_l(\Omega) + h.c.. \tag{10.75}$$

Notice that the boundary condition at conformal infinity reduces the number independent waves to one 'standing wave'.

How does the function f *behave near the horizon and near the asymptotic boundary?*

The analysis near the horizon is identically as in flat space. Define Kruskal coordinates and expand the scalar field near the horizon.

Use all this to compute the Bogoliubov coefficients and argue that, just like in the asymptotically flat Schwarzschild black hole, the modes outside of an AdS black hole must be populated thermally. Note however that, due to the reflective boundary conditions in AdS*, this time there is no flux near the boundary since the global modes multiply standing waves. For this reason black holes in* AdS *do not evaporate and can be in thermal equilibrium with a thermal bath.*

Exercise 10.12. *An alternative way to detect the thermal nature of black hole backgrounds is to study the periodicity in Euclidean time. Show that for a generic black hole background with metric (10.73), the absence of conical singularities predicts a Hawking temperature* $T = \frac{1}{\beta}$ *with*

$$\beta = \frac{4\pi}{V'(r_h)} \,. \tag{10.76}$$

Hint: expand the metric around the horizon and rewrite in the standard metric on $\mathbb{R}^{1,1} \times S^{d-1}$.

Use this to verify your computation of the Hawking temperature of the AdS *Schwarzschild black hole in the previous exercise.*

Part III

APPENDICES

Appendix A

Conventions

In these notes we use natural (or Planck) units in which $\hbar = c = G_N = k_B = 1$. In these units the natural scales are given by

Quantity	Expression	Metric value
Length	$\ell_P = \sqrt{\frac{\hbar G_N}{c^3}}$	$1.616 \cdot 10^{-35}$ m
Mass	$m_P = \sqrt{\frac{\hbar c}{G_N}}$	$2.176 \cdot 10^{-8}$ kg
Time	$t_P = \sqrt{\frac{\hbar G_N}{c^5}}$	$5.391 \cdot 10^{-44}$ s
Temperature	$T_P = \sqrt{\frac{\hbar c^5}{G_N k_B^2}}$	$1.417 \cdot 10^{32}$ K

Using these normalised units, the cosmological constant of our observable universe is $\Lambda \sim 2.888 \cdot 10^{-122} \ell_P^{-2}$.

These notes we employ a plethora of indices, each with its own meaning. The various uses of indices are summarised in Table A.1 below.

Index	Range	Meaning
$\mu, \nu, \ldots$	$0, \ldots, d$	Curved spacetime indices
$m, n, \ldots$	$1, \ldots, d$	Space-like curved spacetime indices
$M, N, \ldots$	$0, \ldots, d+1$	Embedding space curved indices
$a, b, \ldots$	$1, \ldots, d$	Tangent space indices
$\alpha, \beta, \ldots$	$1, 2$	$\mathrm{SU}(2)_L$ indices
$\dot{\alpha}, \dot{\beta}, \ldots$	$1, 2$	$\mathrm{SU}(2)_R$ indices

Table A.1: The various indices used in these lecture notes. d is the dimension of spacetime. When considering four-dimensional spacetime we sometimes employ the exceptional isomorphism $\mathfrak{so}(4) = \mathfrak{su}(2)_L \times \mathfrak{su}(2)_R$.

A.1 Signs, signatures and curvature

Let $(\mathcal{M}, g_{\mu\nu})$ be our space-time. For the most part, we will take to be a four-dimensional manifold with metric $g_{\mu\nu}$ and here we restrict to this situation only. As usual curved indices are raised and lowered with respectively the metric and its inverse $g^{\mu\nu}$ while flat tangent space indices are raised and lowered with the Minkowski metric η_{ab} and its inverse.

A second set of conventions and possible source of confusion is related to the signature of spacetime and the curvature tensor. In order to easily compare with the literature we keep all the signs explicit in this appendix while in the main text we fix all signs to be one,

$$s_1 = s_2 = s_3 = s_4 = s_5 = 1\,. \tag{A.1}$$

The first choice of sign comes from the signature of the metric, which can be either mostly plus or mostly minus,

$$\eta_{ab} = s_1\, \mathrm{diag}\,(+,-,-,-)\,. \tag{A.2}$$

A second sign appears in the definition of the Riemann tensor,

$$R_{\mu\nu}{}^{\rho}{}_{\sigma} = -s_2 \left(\partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\tau}\Gamma^\tau_{\nu\sigma} - \Gamma^\rho_{\nu\tau}\Gamma^\tau_{\mu\sigma} \right). \tag{A.3}$$

A third sign appears in the definition of the Ricci tensor

$$s_2 s_3 R_{\mu\nu} = R^\rho{}_{\nu\rho\mu}\,, \qquad R = g^{\mu\nu} R_{\mu\nu}\,. \tag{A.4}$$

This sign in turn gives rise to a sign in the Einstein equation

$$s_3 \left(R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \right) = -\kappa^2 T_{\mu\nu}\,, \tag{A.5}$$

where by definition, T_{00} is always positive and $\kappa^2 = 8\pi G_N$. The signs s_1 and s_3 determine the signs of the kinetic terms of scalars and gravitons

$$\mathcal{L} = s_1 \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + s_1 s_3 \frac{1}{2\kappa^2} R\,. \tag{A.6}$$

Hence, looking at the Lagrangian one can easily recognize the values of these signs.

When working with frame fields the curvature can also be obtained from the spin connection $\omega_\mu{}^{ab}$. The usual convention is that it is related to the curvature defined above in terms of the Christoffel symbols as

$$R_{\rho\sigma}{}^{\mu}{}_{\nu}(\Gamma) = R_{\rho\sigma}{}^{ab}(\omega) e^\mu_a e_{\nu b}\,. \tag{A.7}$$

However, there is an independent sign in

$$R_{\mu\nu}{}^{ab} = -s_4 \left(\partial_\mu \omega^{ab}_\nu - \partial_\nu \omega^{ab}_\mu + \omega^{ac}_\mu \omega^{\ b}_{\nu c} - \omega^{ac}_\nu \omega^{\ b}_{\mu c} \right). \tag{A.8}$$

This sign is relevant when considering the covariant derivatives of vectors and spinors, which are given by

$$\nabla_\mu \psi = \left(\partial_\mu + s_2 s_4 \frac{1}{4} \omega_\mu{}^{ab} \gamma_{ab} \right) \psi\,, \qquad \nabla_\mu V^a = \partial_\mu V^a + s_2 s_4 \omega_\mu{}^{ab} V_b\,. \tag{A.9}$$

Furthermore we always (anti-)symmetrize with weight one, i.e.

$$A_{[ab]} = \frac{1}{2}(A_{ab} - A_{ba})\,, \qquad A_{(ab)} = \frac{1}{2}(A_{ab} + A_{ba})\,. \tag{A.10}$$

In some references the factor of 2 is omitted. Finally, the Levi-Civita tensor is defined as $\epsilon_{0123} = s_5$ and $\epsilon^{0123} = -s_5$.

To illustrate all these conventions above here are some useful formulae which depend on the choices of sign

$$\nabla_\mu e^a_\nu = \partial_\mu e^a_\nu + s_2 s_4 \omega_\mu{}^{ab} e_{\nu b} - \Gamma^\rho_{\mu\nu} e^a_\rho = 0\,, \tag{A.11}$$

$$\omega_\mu{}^{ab} = s_2 s_4 \left(2e^{\nu[a} \partial_{[\mu} e^{b]}_{\nu]} - e^{\nu[a} e^{b]\sigma} e_{\mu c} \partial_\nu e^c_\sigma \right)\,, \tag{A.12}$$

$$\left[\nabla_\mu, \nabla_\nu \right] V_\rho = s_2 R^\sigma{}_{\rho\mu\nu} V_\sigma \tag{A.13}$$

Appendix B

Differential forms

Differential forms often simplify formulae both computationally and conceptually. In this appendix we briefly review the essentials of this framework, for a more comprehensive treatment, see for example [Nak90, BT13, Nab10].

On any manifold we can define the formal objects, $\mathrm{d}x^\mu$, called differentials. The composition of such differential forms is done through the exterior product and denoted by a wedge, $\wedge$. This product is associative and anti-symmetric,

$$\mathrm{d}x^{\mu_1} \wedge \ldots \wedge \mathrm{d}x^{\mu_p} = \mathrm{d}x^{[\mu_1} \wedge \ldots \wedge \mathrm{d}x^{\mu_p]} \equiv \frac{1}{p!} \sum_{\sigma \in \mathcal{S}_p} (-1)^{|\sigma|} \mathrm{d}x^{\mu_{\sigma(1)}} \wedge \ldots \wedge \mathrm{d}x^{\mu_{\sigma(p)}} \tag{B.1}$$

where $|\sigma|$ denotes the signature of the permutation σ. We define a p-form as an element of the linear vector space $\wedge^p(\mathcal{M})$ spanned by the the external composition of p differentials. Any p-form can thus be represented as a homogeneous polynomial of degree p in the exterior product of differentials,

$$\alpha = \alpha_{\mu_1\mu_2\ldots\mu_p} \mathrm{d}x^{\mu_1} \wedge \ldots \mathrm{d}x^{\mu_p} = \alpha_{[\mu_1\mu_2\ldots\mu_p]} \mathrm{d}x^{\mu_1} \wedge \ldots \mathrm{d}x^{\mu_p} \in \wedge^p(\mathcal{M}) \,. \tag{B.2}$$

In a d-dimensional manifold, the direct sum of vector spaces $\wedge(\mathcal{M}) = \bigoplus_{p=0}^{d} \wedge^p(\mathcal{M})$ is called the exterior algebra. In the exterior algebra, the exterior product is a map $\wedge(\mathcal{M}) \times \wedge(\mathcal{M}) \to \wedge(\mathcal{M})$ defined as

$$\alpha \wedge \beta \equiv \alpha_{[\mu_1\ldots\mu_p} \beta_{\mu_{p+1}\ldots\mu_{p+q}]} \mathrm{d}x^{\mu_1} \wedge \ldots \wedge \mathrm{d}x^{\mu_{p+q}} \in \wedge^{p+q}(\mathcal{M}) \,, \tag{B.3}$$

where α is a p-form and β a q-form. This product is graded commutative

$$\alpha \wedge \beta = (-1)^{pq} \beta \wedge \alpha \,. \tag{B.4}$$

We also have the exterior derivative defined by

$$\mathrm{d}\alpha \equiv \partial_{[\mu_1} \alpha_{\mu_2\ldots\mu_{p+1}]} \mathrm{d}x^{\mu_1} \wedge \ldots \wedge \mathrm{d}x^{\mu_{p+1}} \in \wedge^{p+1}(\mathcal{M}) \,. \tag{B.5}$$

Key features of the exterior derivative are:

Lemma B.1. *The exterior derivative does not depend on the choice of torsion-free covariant derivative. We have $\mathrm{d}^2\alpha = 0$ for all α as a consequence of the commutation of partial derivatives (or symmetry of a torsion-free connection).*

Thus it is metric independent and can be defined just using the coordinate derivative in any coordinate system. The fact that $d^2 = 0$ allows us to define cohomology groups

$$H^p(M) = \{\alpha \in \Omega^p | d\alpha = 0\}/\{\alpha = d\beta\}\,, \tag{B.6}$$

since the exact forms, i.e. those that can be expressed as $d\beta$, are a subset of the closed forms, those that satisfy $d\alpha = 0$. Such cohomology groups encode important information about the topology of M because $d\alpha = 0$ implies that locally there exists a β with $\alpha = d\beta$ (Poincaré lemma).

Example B.1. *As an example, consider the circle S^1. Since the circle is connected, every two points are connected by a segment and are cohomologically equivalent. Indeed, this implies that $H^0(S^1) = \mathbb{R}$ which remains to be true for any connected manifold. Next, let us compute $H^1(S^1)$. Consider a generic one-form $\omega = f(\theta)\mathrm{d}\theta \in \Omega^1(S^1)$. This form is clearly closed so we are left to investigate whether it is exact, i.e. if we can find a globally well-defined function F such that $\omega = \mathrm{d}F$. Locally it is easy to see that we can find such a function,*

$$F(\theta) = \int_0^\theta f(\theta')\mathrm{d}\theta'\,. \tag{B.7}$$

In order for F to be globally well-defined we need to impose that $F(2\pi) = 0$. Defining the function

$$\lambda : \Omega^1(S^1) \to \mathbb{R} : \omega = f(\theta)\mathrm{d}\theta \mapsto \int_0^{2\pi} f(\theta')\mathrm{d}\theta'\,, \tag{B.8}$$

it is easy to see that the first cohomology group is given by

$$H^1(S^1) = \Omega^1(S^1)/\ker\lambda = \mathrm{im}\,\lambda = \mathbb{R}\,. \tag{B.9}$$

The exterior derivative satisfies the graded Leibniz rule

$$d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^p \alpha \wedge d\beta\,. \tag{B.10}$$

Furthermore, we can also define the interior product with a vector V^a that takes a p-form α to the $p-1$-form[1]

$$(V\lrcorner\alpha)_{a_2a_3\ldots a_p} = pV^{a_1}\alpha_{a_1\ldots a_p}\,. \tag{B.11}$$

This also satisfies a graded Leibniz property,

$$V\lrcorner(\alpha \wedge \beta) = (V\lrcorner\alpha) \wedge \beta + (-1)^p \alpha \wedge (V\lrcorner\beta)\,. \tag{B.12}$$

It plays a role in the Cartan formula for the Lie derivative of a form

$$\mathcal{L}_V\alpha = V\lrcorner d\alpha + d(V\lrcorner\alpha)\,. \tag{B.13}$$

When we have a metric, we can define Hodge duality: in d dimensions a p-form α is dualized to a $d-p$ form ${}^*\alpha$ by

$$({}^*\alpha)_{a_{p+1}\ldots a_d} := \frac{1}{p!}\varepsilon_{a_1\ldots a_d}\alpha^{a_1\ldots a_p} \tag{B.14}$$

[1]Another common notation for the inner product is given by $\iota_V\alpha = V\lrcorner\alpha$.

where $\varepsilon_{a_1\ldots a_d} = \varepsilon_{[a_1\ldots a_d]}$ and $\varepsilon_{0\,1\ldots d-1} = \sqrt{-g}$ is the metric volume form.

A key application is to integration. Being a covariant tensor, a p-form naturally 'pulls back' under a map, and restricts to provide a p-form on a submanifold. On a p-dimensional submanifold, it can naturally be integrated subject to the choice of an orientation on the surface.

Definition B.1. A p-surface Σ^p is said to be orientable if it is possible to choose a non-vanishing p-form. Such a choice provides an *orientation* on Σ^p.

The key point is that under a change of coordinates on the p-surface Σ^p, a p-form transforms with the determinant of the Jacobian of the coordinate transformation, whereas the change of variables formula for integration requires the modulus of the determinant which can introduce additional signs, and so we must restrict the coordinate transformations to those that preserve the sign of the chosen form making sure that the sign in question is positive.[2] The standard example of a non-orientable manifold is $\mathbb{RP}^{2n} = S^{2n}/\mathbb{Z}_2$ where the $\mathbb{Z}_2$ acts by the antipodal map which reverses the sign of the volume form.

The main theorem concerning integration on manifolds is Stoke's theorem:

Theorem B.2 (Stokes)**.** *Let Σ be a p-surface with boundary S with compatible orientations (i.e., the orientation on S is obtained from that on Σ by use of an outward pointing normal vector), and let α be a $p-1$-form on Σ, then*

$$\int_\Sigma d\alpha = \int_S \alpha\,. \tag{B.15}$$

Another application is the Cartan formulation of connections and curvature.

B.1 Connections and curvature

Instead of working with the metric, it is often useful to define a orthonormal frame of one-forms, or vielbeine, $e^a = e^a_\mu \mathrm{d}x^\mu$ satisfying

$$g_{\mu\nu} = \eta_{ab} e^a_\mu e^b_\nu\,, \tag{B.16}$$

where $\eta_{ab} = \mathrm{diag}(1, -1, \cdots, -1)$ is the flat Lorentz metric. The vielbeine e^a_μ and their inverses e^μ_a can be used to freely convert curved spacetime indices to flat tangent space indices. Note that the global structure of spacetime manifolds does not always allow the vielbeine to be chosen globally. In other words, generic spacetimes do not admit a global framing. In general this description is only valid locally. However, for globally hyperbolic spacetimes with orientable spatial slices, it is valid globally.

The connection acting on this frame can be obtained from the Cartan structural equation

$$\mathrm{d}e^a + \omega^a{}_b \wedge e^b = 0\,, \tag{B.17}$$

[2]The issue is seen in one dimension: under the transformation $y = -x$,

$$\int_a^b f(x)dx = \int_{-a}^{-b} -f(-y)dy = \int_{-b}^{-a} f(-y)dy\,,$$

so that there is no sign change if we are to integrate from the lower limit to the upper in each case.

where

$$\omega^{ab} = \omega^{[ab]} = \mathrm{d}x^\mu \omega_\mu{}^{ab}\,, \tag{B.18}$$

is the 1-form spin connection.[3] In terms of the spin connection, we can define the curvature 2-form,

$$R_a{}^b = \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu R_{\mu\nu a}{}^b = \mathrm{d}\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b\,. \tag{B.19}$$

Consistency then requires that this form satisfies the Bianchi identities

$$R^a{}_b \wedge e^b = 0\,, \qquad \mathrm{d}R^a{}_b + \omega^a{}_c \wedge R^c{}_b - R^a{}_c \wedge \omega^c{}_b = 0\,. \tag{B.20}$$

For a general 1-form we can then write the covariant derivative as

$$\nabla_a A_b = (\partial_a A_b - \omega_a{}^c{}_b A_c)\,, \tag{B.21}$$

and similar for higher forms. In addition, this formulation allows us to consider spinors in general spacetimes. For more on this see Appendix **??**.

[3]Note that here, as always in this course, we assume the connection to be torsionless. In the presence of torsion, (B.17) has to be modified to $\mathrm{d}e^a + \omega^a{}_b \wedge e^b = \Theta$, where Θ is the torsion 2-form. Similarly, in the presence of torsion the Bianchi identities have to be modified.

Appendix C

Hypersurfaces

This appendix reviews some useful facts on hypersurfaces which will be useful in the computations done in this course. More details and examples can be found in the book [Poi04].

Let $\mathcal{M}$ be a $(d+1)$-dimensional Lorentzian manifold. A hypersurface Σ can be defined by parametric equations of the form

$$x^\mu = x^\mu(y^p)\,, \tag{C.1}$$

where x are coordinates on $\mathcal{M}$ and y^p, $p = 1, \dots d$ are intrinsic coordinates on Σ. Alternatively, the hypersurface can be defined by implicit equations

$$\Phi(x^\mu) = 0\,. \tag{C.2}$$

Exercise C.1. *As in standard Euclidean geometry, show that the vector $\partial_\mu \Phi$ is always normal to the hypersurface.*

A hypersurface is null if $g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi = 0$ and space/time-like if the vectors in the tangent space at each point are space/time-like. When the hypersurface is not null, we can introduce the unit normal vector, defined by

$$n^\mu n_\mu = \epsilon = \begin{cases} +1\,, & \text{if } \Sigma \text{ is space-like,} \\ -1 & \text{if } \Sigma \text{ is time-like.} \end{cases} \tag{C.3}$$

When the hypersurface is defined implicitly, the normal is proportional to $\partial_\mu\Phi$. By definition, the normal is pointed in the direction of increasing Φ, i.e. $n^\mu\partial_\mu\Phi > 0$.

Exercise C.2. *Show that the normal can be written as*

$$n_\mu = \frac{\epsilon\,\partial_\mu\Phi}{|g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi|}\,. \tag{C.4}$$

Define now the tangent vectors

$$E^\mu{}_p = \frac{x^\mu}{\partial y^a}\,, \qquad n_\mu E^\mu{}_p = 0\,. \tag{C.5}$$

The pull-back of the metric to Σ is then given by

$$\mathrm{d}s^2\big|_\Sigma = g_{\mu\nu}E^\mu{}_p E^\nu{}_q \mathrm{d}y^p\mathrm{d}y^q = -\epsilon h_{pq}\mathrm{d}y^p\mathrm{d}y^q\,. \tag{C.6}$$

This defines the so-called induced metric, or first fundamental form, h_{pq}. For non-null surfaces we can then define the surface element as

$$d\Sigma = |h|^{1/2} d^d y\,, \qquad d\Sigma_\mu = \epsilon n_\mu d\Sigma\,. \tag{C.7}$$

One then has the following Lorentzian (or Pseudo-Riemannian) version of Stokes's theorem,

$$\int_{\mathcal{M}} d^{d+1}x \sqrt{|g|}\nabla_\mu A^\mu = \int_{\partial\mathcal{M}} d\Sigma_\mu A^\mu\,. \tag{C.8}$$

Exercise C.3. *Show that the ambient metric $g^{\mu\nu}$, when restricted to Σ, can be decomposed as*

$$g^{\mu\mu} = \epsilon\left(n^\mu n^\nu - h^{pq} E^\mu{}_p E^\nu{}_q\right)\,. \tag{C.9}$$

The second fundamental form, or extrinsic curvature, is defined as

$$K_{pq} = \nabla_\nu n_\mu E^\mu{}_p E^\nu{}_q\,. \tag{C.10}$$

Exercise C.4. *Show that K_{pq} defined in* (C.10) *is a symmetric tensor and can furthermore be written as*

$$K_{pq} = \frac{1}{2}\mathcal{L}_n g_{\mu\nu} E^\mu{}_p E^\nu{}_q\,, \tag{C.11}$$

where $\mathcal{L}_n$ is the Lie derivative along the normal vector n.

The trace of the extrinsic curvature is given by

$$K = h^{pq} K_{pq} = (n^\mu n^\nu - \epsilon g^{\mu\nu})\nabla_\nu n_\mu = h^{pq} E^\mu{}_p E^\nu{}_q \nabla_\nu n_\mu\,. \tag{C.12}$$

Exercise C.5. *As an example, and to get familiar with the concepts introduced above, consider the spacetime $\mathcal{M}$ with metric,*

$$ds^2 = V(r)dt^2 - V(r)^{-1}dr^2 - r^2 d\Omega^2_{d-1}\,. \tag{C.13}$$

and consider the hypersurface defined by $r =$ constant.

1. *Compute the tangent and normal vectors.*

2. *Compute the induced metric and extrinsic curvature.*

Appendix D

Variational calculus

To keep the discussion in these notes self-contained, this appendix includes a short discussion of variational calculus, in particular as applied to the Einstein-Hilbert action. Even though we do not discuss dynamical gravity in these notes this will be useful to revisit the general principles.

The Einstein–Hilbert action for the gravitational field in $d+1$ dimensions is

$$S_{\rm EH} = \frac{1}{16\pi G_N}\int {\rm d}^{d+1}x\sqrt{|g|}R\,. \tag{D.1}$$

Its variation leads to the vacuum Einstein equations. After including additional matter fields these give rise to a non-trivial stress-tensor. The derivation of Einstein's equations is standard and can be found in many textbooks. However, the standard derivations often do not carefully include the contributions due to boundary terms. Such contributions are not so important if your main interest is the study of the solutions to Einstein's equations in classical GR. However, in a quantum theory, the action becomes a crucial object since it gives the weight of a field configuration to the path integral. For this reason let us revisit this derivation with particular care paid to the boundary terms.

Let us consider the gravitational action in a region $\mathcal{M}$ of space-time, with boundary $\partial\mathcal{M}$ and analyze the variation of the action as we vary the metric, with the condition that the metric variation vanishes at the boundary:

$$\delta g_{\mu\nu}\Big|_{\partial\mathcal{M}} = 0\,. \tag{D.2}$$

To study the variation of the Einstein Hilbert action, note that

$$\delta\sqrt{|g|} = -\frac{1}{2}\sqrt{|g|}g_{\mu\nu}\delta g^{\mu\nu}\,, \tag{D.3}$$

such that

$$\delta\left(\sqrt{|g|}R\right) = \sqrt{|g|}G_{\mu\nu}\delta g^{\mu\nu} + \sqrt{|g|}g^{\mu\nu}\delta R_{\mu\nu}\,, \tag{D.4}$$

where $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$ is the Einstein tensor. The vanishing of the first term leads to Einstein's equations in empty space, while the last term is usually neglected by arguing that it vanishes on a boundary at infinity. However, let us have a closer look at this term. Using a local frame where the Christoffel symbols vanish, i.e. using Gauss's normal coordinates, we have

$$\delta R_{\mu\nu} = \delta R^{\rho}{}_{\mu\rho\nu} = \delta\left(\partial_\nu\Gamma^\rho_{\nu\rho} - \partial_\rho\Gamma^\rho_{\mu\nu}\right) = \nabla_\nu\delta\Gamma^\rho_{\mu\rho} - \nabla_\rho\delta\Gamma^\rho_{\mu\nu}\,. \tag{D.5}$$

This last expression is covariant and hence applies in any coordinate system. We can therefore write

$$g^{\mu\nu}\delta R_{\mu\nu} = \nabla_\mu\delta v^\mu\,, \qquad \delta v^\mu = g^{\mu\rho}\delta\Gamma^\nu_{\rho\nu} - g^{\rho\nu}\delta\Gamma^\mu_{\rho\nu}\,. \tag{D.6}$$

Using Stokes' theorem we can then write

$$\int_{\mathcal{M}} \mathrm{d}^{d+1}x\sqrt{|g|}g^{\mu\nu}\delta R_{\mu\nu} = \int_{\mathcal{M}} \mathrm{d}^{d+1}x\sqrt{|g|}\nabla_\mu \delta v^\mu = \int_{\partial\mathcal{M}} \mathrm{d}\Sigma_\mu \delta v^\mu . \tag{D.7}$$

Let us now proceed to compute the variation of v. Since the variation of the metric, but not of its derivatives, vanishes at the boundary, we find

$$\delta\Gamma^\mu_{\rho\sigma} = \frac{1}{2}g^{\mu\nu}\left(\delta\partial_\sigma g_{\nu\rho} + \delta\partial_\rho g_{\nu\sigma} - \delta\partial_\nu g_{\rho\sigma}\right), \tag{D.8}$$

and therefore we have

$$\delta v_\mu = g^{\rho\sigma}\left(\delta\partial_\mu g_{\rho\sigma} - \delta\partial_\rho g_{\mu\sigma}\right). \tag{D.9}$$

Finally, before we can plug this into (D.7), we need to contract this with the (ortho)normal vector n to the boundary,

$$n^\mu \delta v_\mu = -n^\mu \epsilon\left(n^\rho n^\sigma - h^{pq}E^\rho{}_p E^\sigma{}_q\right)\left(\delta\partial_\sigma g_{\mu\rho} - \delta\partial_\mu g_{\sigma\rho}\right). \tag{D.10}$$

Now notice that the last bracket is anti-symmetric in μ and σ, while $n^\mu n^\sigma$. Therefore, the part of the metric which involves the normal vectors drops out. Next, we note that $\delta\partial_\sigma g_{\nu\rho}E^\sigma{}_p = 0$, since the variation of the metric vanishes everywhere on the boundary and therefore the variation on its tangential derivatives has to vanish as well. Finally we have

$$n^\mu \delta v_\mu = -\epsilon h^{\rho\sigma}\delta\partial_\mu g_{\rho\sigma} n^\mu . \tag{D.11}$$

This involves the derivative of the metric variation along the normal direction to the boundary so it is in general non-vanishing.

Putting everything together, we find

$$16\pi G_N \delta S_{EH} = \int_{\mathcal{M}} \mathrm{d}^{d+1}x\sqrt{|g|}G_{\mu\nu}\delta g^{\mu\nu} - \int_{\partial\mathcal{M}} \mathrm{d}^d y\sqrt{|h|}h^{\mu\nu}\delta\partial_\rho g_{\mu\nu}n^\rho . \tag{D.12}$$

Therefore, in the presence of a boundary, Einstein's equations are not sufficient to guarantee the vanishing of the variation of the action, due to the second term in (D.12). To remedy this, we must add an explicit boundary term to the action,

$$S_{\mathrm{GH}} = \frac{1}{8\pi G_N}\int_{\partial\mathcal{M}} \mathrm{d}^d y\sqrt{|h|}K , \tag{D.13}$$

called the Gibbons-Hawking counterterm. To see that this counterterm has the correct properties, let's have a quick look at its variation. Since the metric is fixed at the boundary the variation of h vanishes and the only variation comes from the extrinsic curvature K. Computing its variation we find

$$\delta K = h^{\mu\nu}\delta\partial_\nu n_\mu = \frac{1}{2}h^{\mu\nu}\delta\partial_\rho g_{\mu\nu}n^\mu . \tag{D.14}$$

Varying the Gibbons-Hawking counterterm we then find

$$16\pi G_N \delta S_{\text{GH}} = \int_{\mathcal{M}} \mathrm{d}^d y \sqrt{|h|} h^{\mu\nu} \delta \partial_\rho g_{\mu\nu} n^\rho \,, \tag{D.15}$$

which exactly cancels the boundary term in the original variation and thus rendering the variational principle well-defined.

In the context of black hole physics and AdS/CFT this is not always enough. In general, the action, supplied with the Gibbons-Hawking counterterm, will lead to divergences when evaluated, even in flat space. In order to obtain finite values for the on-shell action, it is usually necessary to add additional counterterms removing the divergences. In the context of AdS/CFT the prescription to remove said divergences goes under the name of holographic renormalisation.

Appendix E

Ingredients from general relativity

In this appendix we review a variety of useful facts from general relativity that will come in handy in this course.

E.1 Maximally symmetric spaces

Let us start with the maximally symmetric manifolds. In Euclidean signature these are given by the sphere S^d, flat Euclidean space $\mathbb{R}^d$ and hyperbolic space H^d, which are respectively positively curved, flat or negatively curved. We will sometimes collectively denote them by $M_{k,d}$, where $k = 0, \pm 1$ and write their metrics as

$$\mathrm{d}s^2_{M_{k,d}} = \begin{cases} \mathrm{d}\Omega_d^2\,, & k = 1\,, \\ \sum_{i=1}^d \mathrm{d}x_i^2\,, & k = 0\,, \\ \mathrm{d}\Xi_d^2\,, & k = -1\,, \end{cases} \tag{E.1}$$

where $\mathrm{d}\Omega_d^2$ and $\mathrm{d}\Xi_d^2$ are the standard metrics on the d-sphere and d-dimensional hyperbolic space respectively,

$$\mathrm{d}\Omega_d^2 = \sum_{i=1}^d \left(\prod_{j=1}^{i-1} \sin^2\phi_j \right) \mathrm{d}\phi_i^2\,, \qquad \mathrm{d}\Xi_d^2 = \frac{1}{x_d^2} \sum_{i=1}^d \mathrm{d}x_i^2\,. \tag{E.2}$$

There are many other coordinate choices but unless explicitly stated otherwise we will always use the metrics above.

In most of this course we are interested in spaces with Minkowskian signature. The maximally symmetric Lorentzian spacetimes are de Sitter space, Minkowski space and anti-de Sitter space, which are respectively positively curved, flat or negatively curved.

de Sitter space

The de Sitter space is the maximally symmetric spacetime of positive curvature

$$R = \frac{d(d+1)}{L^2}\,, \tag{E.3}$$

where L is the characteristic length scale of the space. This space describes a exponentially expanding universe. For this reason, there is an observer-dependent horizon, called the 'cosmological horizon' beyond which spacetime is expanding faster than the speed of light. This is a null surface beyond which the observer can never receive a signal.

In $d+1$ dimensions de Sitter space can be described by a hypersurface in $d+2$ dimensional Minkowski space. consider the embedding space $\mathbb{R}^{1,d+1}$ with metric,

$$\mathrm{d}s^2 = \mathrm{d}X_0^2 - \sum_{i=1}^{d+1} \mathrm{d}X_i^2 \,. \tag{E.4}$$

The de Sitter space of radius L is then defined as the hyperboloid

$$X_\mu X^\mu = X_0^2 - \sum_{i=1}^{d+1} X_i^2 = -L^2 \,, \tag{E.5}$$

In the same way that the two sphere embedded in $\mathbb{R}^3$ inherits the O(3) symmetry of its ambient space, de Sitter space inherits an $\mathrm{O}(1,d)$ symmetry from the ambient Minkowski space. There are various useful coordinate systems to describe the de Sitter space of which we list a few below:

- **Global coordinates.** These coordinates $\{\tau, \omega_i\}$ are defined by

  $$X_0 = L \sinh \frac{\tau}{L} \,, \qquad X_i = L\omega_i \cosh \frac{\tau}{L} \,, \tag{E.6}$$

  where $i = 1, \dots d+1$ and the ω_i are constrained coordinates on a round unit sphere S^d such that $\sum_i \omega_i^2 = 1$. In these coordinates, the metric on the de Sitter space reads,

  $$\mathrm{d}s^2 = \mathrm{d}\tau^2 - L^2 \cosh^2 \frac{\tau}{L} \, \mathrm{d}\Omega_d^2 \,. \tag{E.7}$$

  Global coordinates are sometimes also called the closed slicing of de Sitter, especially in the context of cosmology. This terminology comes from the FLRW metric (see below) since for these coordinates the space-like slices are closed.

- **Planar coordinates.** These coordinates $\{t, \mathbf{x}\}$ are defined as

  $$\begin{aligned} X_0 &= L \sinh \frac{t}{L} + \frac{\mathbf{x}^2}{2L} \mathrm{e}^{\frac{t}{L}} \,, \\ X_{d+1} &= L \cosh \frac{t}{L} - \frac{\mathbf{x}^2}{2L} \mathrm{e}^{\frac{t}{L}} \,, \\ X_i &= x_i \mathrm{e}^{\frac{t}{L}} \,, \end{aligned} \tag{E.8}$$

  where $i = 1, \dots, d$. These coordinates do not cover the full de Sitter space but only the patch

  $$X_0 + X_{d+1} = L\mathrm{e}^{\frac{t}{L}} > 0 \,. \tag{E.9}$$

  In these coordinates, the metric reads,

  $$\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{e}^{\frac{2t}{L}} \sum_{i=1}^{d} \mathrm{d}x_i^2 \,. \tag{E.10}$$

- **Static coordinates.** de Sitter space enjoys various time-like isometries inherited from the boosts in embedding space. Yet, the metrics considered so far are time dependent. Since there is a time-like Killing vector, there must exist coordinates such that time does not appear explicitly in the metric. These are static coordinates and are defined as

$$
\begin{aligned}
X_0 =& \sqrt{L^2 - r^2} \sinh \frac{t}{L}, \\
X_{d+1} =& \sqrt{L^2 - r^2} \cosh \frac{t}{L}, \\
X_0 =& r \omega_i,
\end{aligned}
\tag{E.11}
$$

with $i = 1, \ldots, d$ and $0 \leq r < L$. They only cover the region

$$
X_{d+1} > 0, \qquad \sum_i X_i < L^2. \tag{E.12}
$$

The resulting metric reads

$$
\mathrm{d}s^2 = \left(1 - \frac{r^2}{L^2}\right) \mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - \frac{r^2}{L^2}} - r^2 \mathrm{d}\Omega_{d-1}^2. \tag{E.13}
$$

This metric is manifestly static. In static coordinates the cosmological horizon is located at $r = L$ Therefore these coordinates cover precisely the patch that is accessible to a single observer, in the sense that the observer can both send and receive signals to/from this entire region.

- **Hyperbolic coordinates.** Global coordinates foliate de Sitter with spheres, while planar coordinates foliate with planes. To cover the third possibility we introduce hyperbolic coordinates which foliate de Sitter by hyperbolic spaces. The embedding coordinates are defined by

$$
\begin{aligned}
X_0 =& \sinh \tau \cosh \psi, \\
X_{d+1} =& \cosh \tau, \\
X_i =& \sinh \tau \sinh \psi \, \omega_i,
\end{aligned}
\tag{E.14}
$$

such that the metric takes the form

$$
\mathrm{d}s^2 = \mathrm{d}\tau^2 - \sinh^2 \tau \mathrm{d}\Xi_d^2. \tag{E.15}
$$

- **Conformal coordinates.** Finally, to obtain the conformal coordinates we start from planar coordinates and perform a coordinate transformation to conformal time

$$
\eta = \int_{\infty}^{t} \frac{\mathrm{d}t'}{a(t')} = -L \mathrm{e}^{-\frac{t}{L}}, \tag{E.16}
$$

so that $-\infty < \eta < 0$ and $\eta \to -\infty$ corresponds to the infinite past $t \to -\infty$. However, one can extend this spacetime to almost all of the de Sitter by extending the range of η to the full real line. These coordinates can be parameterised as

$$
\begin{aligned}
X_0 =& \frac{1}{2\eta}\left(\eta^2 - \mathbf{x}^2 - L^2\right), \\
X_{d+1} =& -\frac{1}{2\eta}\left(\eta^2 - \mathbf{x}^2 + L^2\right), \\
X_0 =& -\frac{L}{\eta} x_i ,
\end{aligned}
\tag{E.17}
$$

so that the metric is given by

$$
\mathrm{d}s^2 = \frac{L^2}{\eta^2}\left(\mathrm{d}\eta^2 - \sum_{i=1}^{d} \mathrm{d}x_i^2\right) \tag{E.18}
$$

These coordinates cover only the submanifold $X_0 + X_{d+1} = 0$. When written in these coordinates, the metric of the de Sitter space, is manifestly conformally flat.

An important quantity in the geometry of the de Sitter space is the geodesic distance between two points $\zeta(X,X')$. As in the case of the sphere or the hyperbolic plane in two dimensions, the distance between two points of the de Sitter space is closely related to the distance defined in the embedding space. Therefore, let us define

$$
P(X,X') = -H^2 \eta_{ab} X^a X^b . \tag{E.19}
$$

Notice that, if $X = X'$ are identical, we have $P = 1$. However, if X and X' are antipodal, i.e. $X' = -X$, one has $P = \frac{1}{L^2}\eta_{ab}X^a X^b = -1$. Plugging in the explicit parameterisation, we find an expression of the geodesic distance in conformal coordinates:

$$
P(X,X') = 1 + \frac{(\eta - \eta')^2 - (\mathbf{x} - \mathbf{x}')^2}{2\eta\eta'} . \tag{E.20}
$$

One important property of $P(X,X')$ is that it is a manifestly $\mathrm{O}(1,d)$ invariant function on de Sitter space, since it is constructed out of the Lorentz invariant product in $\mathbb{R}^{1,d+1}$. Depending on the causal relationship between X and X', we have the following behaviour for $\zeta(X,X')$:

- If X and X' are joined by a time-like geodesic $P(X,X') > 1$ and the geodesic distance is given by

$$
\zeta(X,X') = \frac{1}{H}\cosh^{-1}(P) . \tag{E.21}
$$

- If X and X' are space-like separated, $|P(X,X')| < 1$ and

$$
\zeta(X,X') = \frac{1}{H}\cos^{-1}(P) . \tag{E.22}
$$

- If X and X' are light-like separated, $P(X,X') = 1$ and $\zeta(X,X') = 0$.

Notice that there are points of de Sitter space which cannot be joined by geodesics to a given point X. These are the points in the interior of the past and future light cones of $-X$, the antipodal point of X. For these points, we have that $P(X,X') < -1$. The results listed above can be obtained by an explicit analysis of geodesics in de Sitter space.

Exercise E.1. *The geodesics in the de Sitter space can be obtained by minimizing the distance in the embedding space subject to the constraint* (E.5). *Employ an appropriate Lagrange multiplier to solve this minimisation problem and explicitly find the geodesics in the de Sitter space.*

Anti-de Sitter space

The anti-de Sitter space is the maximally symmetric spacetime with negative curvature. Its scalar curvature is given by

$$R = -\frac{d(d+1)}{L^2}\,. \tag{E.23}$$

Analogous to the de Sitter space we can describe Anti-de Sitter space in $d+1$ dimensions by the hyperboloid

$$X_0^2 + X_1^2 - \sum_{i=1}^{d} X_i^2 = L^2\,, \tag{E.24}$$

embedded in a $(d+2)$ dimensional ambient space with metric

$$\mathrm{d}s^2 = \mathrm{d}X_0^2 + \mathrm{d}X_1^2 - \sum_{i=2}^{d+1} \mathrm{d}X_i^2\,. \tag{E.25}$$

The constant L is the AdS length parameterising the characteristic scale of the anti-de Sitter space. The anti-de Sitter space inherits an $O(2,d)$ symmetry from the ambient space. There are a variety of useful coordinates on the Anti-de Sitter space, similar to the de Sitter space.

- **Global coordinates.** These coordinates $\{\tau, \rho, \omega_i\}$ are defined by

$$\begin{aligned} X_0 =& L\cosh\rho\cos\tau\,,\\ X_1 =& L\cosh\rho\sin\tau\,,\\ X_i =& L\sinh\rho\,\omega_i\,, \end{aligned} \tag{E.26}$$

  where $i = 2, \ldots d+1$ and the ω_i are embedding coordinates on a round unit sphere S^{d-1} such that $\sum_i \omega_i^2 = 1$. In these coordinates, the metric on the Anti-de Sitter space reads,

$$\mathrm{d}s^2 = L^2\left(\cosh^2\rho\,\mathrm{d}\tau^2 - \mathrm{d}\rho^2 - \sinh^2\rho\,\mathrm{d}\Omega_{d-1}^2\right)\,. \tag{E.27}$$

  In the limit $\rho \to \infty$ one approaches the conformal boundary which in global coordinates is given by $\mathbb{R} \times S^{d-1}$.

- **Poincaré coordinates.** These coordinates $\{z, x\}$ are defined as

$$\begin{aligned} X_0 &= \frac{L^2 - t^2 + \mathbf{x}^2 + z^2}{2z}\,,\\ X_1 &= \frac{Lt}{z}\,,\\ X_i &= \frac{Lx_i}{z}\,,\\ x_{d+1} &= \frac{-L^2 - t^2 + \mathbf{x}^2 + z^2}{2z}\,, \end{aligned} \tag{E.28}$$

where $i = 2, \ldots, d$. These coordinates do not cover the full anti-de Sitter space but only the patch

$$X_0 - X_{d+1} = \frac{L^2}{z} > 0\,. \tag{E.29}$$

In these coordinates, the metric reads,

$$\mathrm{d}s^2 = \frac{L^2}{z^2}\left(\mathrm{d}z^2 - \mathrm{d}t^2 + \sum_i \mathrm{d}x_i^2\right). \tag{E.30}$$

In these coordinates the conformal boundary is located at $z \to 0$ and the geometry of the boundary is that of d-dimensional Minkowski space.

- **Static coordinates.** These coordinates are given by

$$\begin{aligned} X_0 &= L\sqrt{1+\frac{r^2}{L^2}}\sin\frac{t}{L}\,, \\ X_1 &= L\sqrt{1+\frac{r^2}{L^2}}\cos\frac{t}{L}\,, \\ X_i &= r\,\omega_i\,, \end{aligned} \tag{E.31}$$

with $i = 2, \ldots, d+1$. In these coordinates the metric reads

$$\mathrm{d}s^2 = \left(1+\frac{r^2}{L^2}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1+\frac{r^2}{L^2}} - r^2\mathrm{d}\Omega_{d-1}^2\,. \tag{E.32}$$

The conformal boundary is located at $r \to 0$ and its geometry is given by $\mathbb{R}\times S^{d-1}$.

- **de Sitter slicing.** Finally, we can slice the anti-de Sitter space with de Sitter slices. The embedding coordinates are given by

$$\begin{aligned} X_0 =& L\sinh\rho\sinh t\cosh\xi\,, \\ X_1 =& L\cosh\rho\,, \\ X_2 =& L\sinh\rho\cosh t\,, \\ X_i =& L\sinh\rho\sinh t\sinh\xi\,\omega_i \end{aligned} \tag{E.33}$$

so that the metric is given by

$$\mathrm{d}s^2 = L^2\left(\mathrm{d}\rho^2 + \sinh^2\rho\,\mathrm{d}s^2_{\mathrm{dS}_{d-1}}\right). \tag{E.34}$$

where $\mathrm{d}s^2_{\mathrm{dS}_{d-1}}$ is a metric on de Sitter space with Hubble scale $H = 1$.

- **Hyperspherical coordinates.** In these coordinates the embedding is parametrised as

$$\begin{aligned} X_0 =& L\sec\rho\cos\tau\,, \\ X_1 =& L\sec\rho\sin\tau\,, \\ X_i =& L\tan\rho\,\omega_i\,, \end{aligned} \tag{E.35}$$

with ω_i again parametrising a unit $d-1$ sphere. The metric in these coordinates is given by

$$ds^2 = L^2 \sec^2 \rho \left(d\tau^2 - d\rho^2 - \sin^2 \rho \, d\Omega_{d-1}^2\right) . \tag{E.36}$$

In these coordinates the boundary of (the universal covering of) AdS is the Einstein static universe.

An important quantity in the geometry of the anti-de Sitter space is the geodesic distance between two points,

$$P(X,X') = -\frac{1}{L^2}\eta_{ab}X^a X'^b . \tag{E.37}$$

where η_{ab} is the metric in the ambient space. If $X = X'$ are identical, we have $P = 1$. However, if X and X' are antipodal one has $P = -1$.

One important property of $P(X,X')$ is that it is a manifestly O$(2,d)$ invariant function on anti-de Sitter space, since it is constructed out of the Lorentz invariant product in $R^{2,d}$. Depending on the causal relationship between X and X', we have the following behaviour for $P(X,X')$:

- If X and X' are joined by a time-like geodesic $|P(X,X')| < 1$ and the geodesic distance is given by
$$d(X,X') = L\cos^{-1}(P) . \tag{E.38}$$
- If X and X' are space-like separated, $P(X,X') > 1$ and
$$d(X,X') = L\cosh^{-1}(P) . \tag{E.39}$$
- If X and X' are light-like separated, $P(X,X') = 1$ and $d(X,X') = 0$.

Notice that this is opposite of the de Sitter case. Furthermore, notice that in the anti-de Sitter space it is possible to reach the conformal boundary in a finite time. I.e. there exists a time-like geodesic connecting any point X with the conformal boundary. This is why AdS is often thought of as a finite 'box'.

Exercise E.2. *The geodesics in the anti-de Sitter space can be obtained by minimizing the distance in the embedding space subject to the constraint* (E.24). *Employ an appropriate Lagrange multiplier to solve this minimisation problem and explicitly find the geodesics in the de Sitter space.*

Euclidean AdS

In some situations it is more convenient to perform computations in Euclidean signature and after Euclidean AdS spacetime is the hyperboloid

$$-X_0^2 + \sum_{i=2}^{d+1} X_i^2 = -R^2 , \qquad X^0 > 0 , \tag{E.40}$$

embedded in $\mathbb{R}^{d+1,1}$. From this definition it is clear that Euclidean AdS is invariant under $SO(d+1,1)$. Let us be more explicit it this case and write out the symmetry generators as

$$J_{AB} = -i\left(X_A \frac{\partial}{\partial X^B} - X_B \frac{\partial}{\partial X^A}\right) . \tag{E.41}$$

Analogous to AdS , we define the Poincaré coordinates by

$$\begin{aligned} X_0 &= R\frac{1+x^2+z^2}{2z} \\ X_\mu &= R\frac{x^\mu}{z} \\ X_{d+1} &= R\frac{1-x^2-z^2}{2z} \end{aligned} \tag{E.42}$$

where $x^\mu \in \mathbb{R}^d$ and $z > 0$. In these coordinates the metric reads

$$ds^2 = R^2\frac{dz^2 + \delta_{\mu\nu}dx^\mu dx^\nu}{z^2} . \tag{E.43}$$

This shows that EAdS is conformal to $\mathbb{R}^+ \times \mathbb{R}^d$ whose boundary at $z = 0$ is just $\mathbb{R}^d$. These coordinates make explicit the subgroup $SO(1,1) \times ISO(d)$ of the full isometry group of EAdS. These correspond to dilatation and Poincaré symmetries inside the $d-$dimensional conformal group. In particular, the dilatation generator is

$$D = -i\,J_{0,d+1} = -X_0\frac{\partial}{\partial X^{d+1}} + X_{d+1}\frac{\partial}{\partial X^0} = -z\frac{\partial}{\partial z} - x^\mu\frac{\partial}{\partial x^\mu} . \tag{E.44}$$

Global coordinates in Euclidean AdS can simply be obtained from the global coordinates in AdS by analytically continuing $\tau \to i\tau$ such that the metric is given by

$$ds^2 = R^2\left[\cosh^2\rho\,\mathrm{d}\tau^2 + \mathrm{d}\rho^2 + \sinh^2\rho\,\mathrm{d}\Omega^2_{d-1}\right]. \tag{E.45}$$

To understand the global structure of this spacetime it is convenient to change the radial coordinate via $\tanh\rho = \sin r$ so that $r \in [0, \frac{\pi}{2}[$. Then, the metric becomes

$$ds^2 = \frac{R^2}{\cos^2 r}\left[d\tau^2 + dr^2 + \sin^2 r\,d\Omega^2_{d-1}\right] , \tag{E.46}$$

which is conformal to a solid cylinder whose boundary at $r = \frac{\pi}{2}$ is $\mathbb{R} \times S^{d-1}$. In these coordinates, the dilatation generator $D = -i\,J_{0,d+1} = -\frac{\partial}{\partial\tau}$ is the Hamiltonian conjugate to global time.

Exercise E.3. *Explicitly write out the symmetry generators for (Lorentzian) (A)dS spacetime, analogous to the discussion in this last subsection.*

E.2 Warped product manifolds and FLRW spaces

Apart from the maximally symmetric spacetimes, our second most loved example is given by warped product manifolds of the form $\mathbb{R} \times \mathcal{M}$ with metric

$$ds^2 = dt^2 - a(t)^2 ds^2_{\mathcal{M}}, \tag{E.47}$$

When $\mathcal{M}$ is a maximally symmetric Euclidean manifold, i.e. $\mathcal{M} = M_{k,d}$ these are the FLRW manifolds introduced in the main text. When $a(t)$ is a constant function these represent static spacetimes. In the case $k = +1$ this spacetime is often called the static Einstein universe.

On the other hand, when $a(t)$ is a non-trivial function of time, these manifold provide an excellent toy model for cosmology. The spatial section of the universe contracts or expands according to the scale factor $a(t)$. It is often useful to define the conformal time coordinate,

$$\eta = \int_{-\infty}^{t} \frac{dt'}{a(t')}, \tag{E.48}$$

in terms of which the FLRW metric becomes

$$ds^2 = a^2(\eta)\left(d\eta^2 - ds^2_{M_{k,d}}\right). \tag{E.49}$$

Note that the de Sitter space can be thought of as a FLRW space.

In maximally symmetric spacetimes, such as Minkowski or (A)dS, there can be no beginning or end of time. There cannot be any history because every time is equivalent. The simplest way to introduce some time dependence is to consider FLRW spacetimes. For this reason they are natural toy models in cosmology. The coordinate t introduced above, corresponds to the proper time of an observer at rest with respect to the co-moving spatial coordinates. The spatial manifold in an FLRW space is maximally symmetric hence such spacetimes describe a homogeneous and isotropic universe, i.e. a universe that looks the same at every point in space. This metric is simple enough that it allows for a very explicit, often exact, analysis. A most remarkable fact that should blow your mind is that this most simple spacetime for $k = 0$ is in fact a very good description of our own universe on distances much larger than average distance between galaxies, about a few Megaparsec (Mpc). Of course there are small deviations from perfect homogeneity in our universe but on large enough scales this gives an excellent description.

In contrast to the maximally symmetric spacetimes, in an FLRW universe time translation and Lorentz boost fail to be isometries because of the time dependence of the scale factor $a(t)$. A useful way to capture this dependence is by defining the Hubble parameter

$$H(t) = \frac{\dot{a}}{a}. \tag{E.50}$$

The absence of time translations has profound implications for constructing QFTs in these backgrounds, as energy is not preserved.

Next, let us quickly review the Einstein equations in such backgrounds. In order to have any hope to solve them, we need to impose some particularly symmetric stress tensor. The most general stress tensor consistent with the symmetries of FLRW spaces is given by

$$T_{\mu\nu} = \mathrm{diag}\,(\rho, -p, \cdots, -p)\,, \tag{E.51}$$

where the energy density ρ and pressure p are functions of time only. We can interpret this as the energy-momentum tensor of a homogeneous perfect fluid in its rest frame,

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - g_{\mu\nu}p\,, \tag{E.52}$$

where u_μ is the normalised fluid velocity, $|u|^2 = 1$, which in rest frame would be $u_\mu = \delta_{\mu t}$. Einsteins equation imply that the energy momentum tensor is covariantly conserved,

$$\nabla_\nu T^{\mu\nu} = \partial_\nu T^{\mu\nu} + \Gamma^\mu_{\alpha\nu}T^{\alpha\nu} + \Gamma^\nu_{\alpha\nu}T^{\mu\nu} = 0\,. \tag{E.53}$$

Plugging in the FLRW solution this reduces to the so-called continuity equation,

$$\dot\rho + d\,H(\rho + p) = 0\,. \tag{E.54}$$

This equation tells us that the energy density changes only if the universe expands or contracts, i.e. if $H \neq 0$. The Einstein equations however will not tell us what kind of matter permeates the universe. For that we need to specify an equation of state giving a relation between the pressure, density and possibly other thermodynamic variables. Most systems of interest in cosmology can be described to a good approximation by the very simple equation of state,

$$p = w\rho\,, \tag{E.55}$$

with a single parameter w. For this equation of state we can immediately solve the continuity equation giving us

$$\rho(t) = \rho_0 a(t)^{-d(1+w)}\,. \tag{E.56}$$

- Non-relativistic matter, a.k.a. dust, has a velocity much smaller than the speed of light. For this type of matter, the pressure is negligible compared to the energy density, $p \ll \rho$, or $0 < w \ll 1$. Therefore in an expanding universe, dust dilutes as $\rho \propto a^{-d}$.

- Relativistic matter, a.k.a. radiation, on the other hand has pressure and energy density of the same order. A statistical mechanics analysis furthermore predicts that $p = \rho/d$ and so $w = 1/d$. This is precisely the proportionality constant to make the matter conformal. Hence we find $\rho \propto a^{-(d+1)}$.

- Finally, a cosmological constant, or vacuum energy has $T_{\mu\nu} = -\Lambda g_{\mu\nu}$ and hence $p = -\rho = -\Lambda$, or $w = -1$. Since in this case we have $p + \rho = 0$, the continuity equation teaches us that the cosmological constant does not dilute, $\rho \propto a^0$.

Solving the Einstein equations for an FLRW metric results in the Friedmann equations, which can be written as

$$H^2 = \frac{16\pi G_N}{d(d-1)}\rho - \frac{k}{a^2}\,, \qquad \dot{H} = -\frac{8\pi G_N}{d-1}(\rho + p) + \frac{k}{a^2}\,. \tag{E.57}$$

Here p and ρ can be thought of as the effective energy density and pressure build from the combination of all the matter present in the universe, together with the cosmological constant,

$$p = \sum_m p_m - \frac{\Lambda}{8\pi G_N}\,, \qquad \rho = \sum_m \rho_m + \frac{\Lambda}{8\pi G_N}\,. \tag{E.58}$$

The first Friedmann equation can be used to estimate the age of the universe, while the second encodes the acceleration of the universe. Since most cosmological matter respects the null energy condition, which in this case reads $\rho + p > 0$, we find that H typically decreases during the expansion of the universe.

E.3 Black holes

Another important set of spacetimes that play a key role in this course are black holes. contrary to the above examples, such spacetimes are singular and hence not complete.

In this course we restrict ourselves to the simplest of black holes, the Schwarzschild black hole, as it will suffice to illustrate the relevant phenomena. This is the unique non-rotating neutral, asymptotically flat black hole. More generally, black holes can have mass and/or charge. For such more general solutions we refer the reader to the course General Relativity II.

The Schwarzschild black hole

The Schwarzschild black hole, with metric

$$\mathrm{d}s^2 = \left(1 - \frac{2M}{r}\right)\mathrm{d}t^2 - \left(1 - \frac{2M}{r}\right)^{-1}\mathrm{d}r^2 - \mathrm{d}\Omega_2^2\,, \tag{E.59}$$

is the unique non-rotating asymptotically flat black hole. This metric has a singularity at $r = 2M$, the location of the event horizon, and therefore only describes the exterior of the black hole. To see that the singularity at the event horizon is not a physical singularity it is useful to introduce the tortoise or Regge-Wheeler coordinate r^*, which is defined such that massless free falling observers follow the path $t = r^* + \text{constant}$. For such an observer one has,

$$\begin{aligned} \mathrm{d}s^2 = 0 \quad &\rightarrow \quad \mathrm{d}t = \frac{1}{1 - \frac{2M}{r}}\mathrm{d}r \equiv \mathrm{d}r^* \\ &\rightarrow \quad r^* = r + 2M\log\left(\frac{r}{2M} - 1\right)\,, \end{aligned} \tag{E.60}$$

where we put the arbitrary additive constant to zero. In these coordinates the metric takes the form,

$$\mathrm{d}s^2 = \left(1 - \frac{2M}{r}\right)\left(\mathrm{d}t^2 - \mathrm{d}r^{*2}\right) - r^2\mathrm{d}\Omega_2^2\,. \tag{E.61}$$

From this line element we see that the two-dimensional metric is conformally equivalent to Minkowski space. Next we introduce the retarded and advanced Eddington-Finkelstein coordinates

$$u = t + r^*, \qquad v = t - r^*, \tag{E.62}$$

which places the horizon at $(u, v) \to (\infty, -\infty)$. In these coordinates the metric becomes

$$ds^2 = \left(1 - \frac{2M}{r}\right) du dv - r^2 ds^2_{S^2}, \tag{E.63}$$

where r can be expressed as a complicated function of $u - v$. In these coordinates the metric is still singular at $r = 2M$ but we can introduce one more coordinate transformation

$$U = -4M \exp\left(-\frac{u}{4M}\right), \qquad V = 4M \exp\left(\frac{v}{4M}\right), \tag{E.64}$$

called Kruskal-Szekeres coordinates. The final metric is then given by

$$ds^2 = \frac{2M}{r} e^{-\frac{r}{2M}} dU dV - r^2 d\Omega_2^2. \tag{E.65}$$

which makes it clear that there is no singularity at $r = 2M$. In these coordinates the event horizon is at $U = 0$ or $V = 0$ and the original Schwarzschild metric only covers the patch $U < 0$ and $V > 0$. However, there is no obstruction to extend U, V to the full real line. The fully extended metric covers both the inside and outside of the Schwarzschild black hole and is the maximal extension of this spacetime. Finally, an explicit map between U, V and r, t is given by

$$UV = e^{\frac{r}{2M}} \left(1 - \frac{r}{2M}\right), \qquad \frac{U}{V} = e^{\frac{t}{2M}}. \tag{E.66}$$

We see in fact that $r = 2M$ is a null hypersurface ruled by outgoing null geodesics. The presence of the event horizon means that not all light rays escape to infinity. For $r > 2M$, light rays with $\dot{r} > 0$ can and do escape. However, for $r < 2M$, all causal geodesics have future end point at the true singularity at $r = 0$. More precisely, we can define the event horizon is as

Definition E.1. The event horizon is the boundary of the past of $\mathscr{I}^+$.

this can easily be seen from the metric using the retarded Eddington-Finkelstein coordinate

$$ds^2 = \left(1 - \frac{2M}{r}\right) du^2 + 2 du dr - r^2 d\Omega_2^2, \tag{E.67}$$

as the boundary is at $u \to \infty$, which corresponds to $r \to 2M$. There is a corresponding time-reversed picture using coordinates (v, r). However, now we have that $r = 2M$ is the past horizon being the boundary $v = -\infty$ of the future of $\mathscr{I}^-$. The full causal structure of the Schwarzschild black hole can be summarised in its Penrose diagram, see Figure E.1.

Similar diagrams can be drawn for Reissner-Nordstrom, Kerr and the Kerr-Newman, see GR II, although the latter have the novelty of having Cauchy horizons, hypersurfaces beyond which neither fields nor space-time itself are determined by Cauchy data essentially as a consequence of naked singularities, singularities in the past of observers. However, these cannot be seen from infinity. These

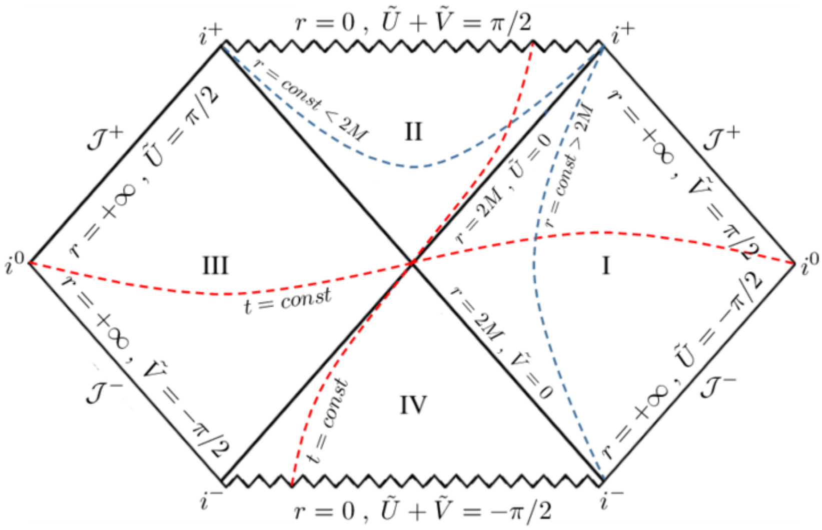


Figure E.1: Penrose diagram for the Kruskal extension of the Schwarzschild spacetime. The singularity at $r = 0$ (which is a genuine curvature singularity) is a black hole to the future of every observer that crosses the future event horizon (or a white hole in the past).

black hole solutions are unique subject to various assumptions (like the existence of a stationary Killing vector that looks like a time translation at large distances). Similarly, all of these spacetimes have generalisations with non-vanishing cosmological constant.

This final state is tightly constrained as in four dimensions we have powerful uniqueness theorems. Birkhoff's theorem says that any spherically symmetric vacuum solution is static, which then implies that it must be Schwarzschild. For Einstein-Maxwell system this extends to show that the only spherically symmetric solution is Reissner-Nordstrom. But suppose we know only that the metric exterior to a star is static. We further have:

Theorem E.1 (Israel). *If $(\mathcal{M}, g)$ is an asymptotically-flat, static, vacuum space-time that is non-singular on and outside an event horizon, then (M,g) is Schwarzschild.*

More remarkably we have

Theorem E.2 (Carter-Robinson). *If $(\mathcal{M}, g)$ is an asymptotically-flat stationary and axi-symmetric vacuum spacetime that is non-singular on and outside an event horizon, then (M,g) is a member of the two-parameter Kerr family. The parameters are the mass M an the angular momentum J.*

The assumption of axi-symmetry has since been shown to be unnecessary by Hawking and Wald , i.e., for black holes, stationarity ⇒ axisymmetry.

A digression on null congruences and hypersurfaces

An important role is played by null hypersurfaces, i.e., hypersurfaces $u(x) = \text{constant}$ such that $\nabla_a u$ is a null vector. It is a standard elementary exercise that $l_a = \nabla_a u$ is tangent to a family of null geodesics.

More generally, a null congruence is a foliation of a region of space-time by null geodesics. It can be defined by a null vector field l^a whose integral curves are the null geodesics through each point. If it is tangent to a congruence of affinely parametrised null geodesics, then

$$\nabla_l l^b := l^a \nabla_a l^b = 0 \tag{E.68}$$

The geometry can be studied by introducing a 2-dimensional *screen space* consisting of tangent vectors perpendicular to the direction of null geodesic. This space can be spanned by a pair of orthonormal vectors X^a, Y^a, $X \cdot X = Y \cdot Y = 1$ and $X \cdot l = Y \cdot l = 0$. Given l, these are defined up to standard 2d rotations. The general vector in this screen space can be parametrized as

$$V^a = xX^a + yY^a \,. \tag{E.69}$$

It is conventional to introduce the complex coordinate $\sqrt{2}\zeta = x + iy$ on screen space and associated complex null vectors $m^a, \bar{m}^a$ so that

$$m^a = \frac{1}{\sqrt{2}}(X^a - iY^a), \qquad V^a = \zeta \bar{m}^a + \bar{\zeta} m^a \,. \tag{E.70}$$

We can choose X^a and Y^a to be parallel propagated along the null geodesics, $\nabla_l X^a = \nabla_l Y^a = \nabla_l m^a = 0$.

We wish to measure how images on the screen are distorted as they are propagated along the light rays of the congruence. If V^a connects nearby geodesics of the congruence, then it is Lie derived along l^a, i.e.,

$$[l, V]^a = \nabla_l V^a - \nabla_V l^a = 0 \,. \tag{E.71}$$

It is easy to see that if $l \cdot V = 0$ initially, then it remains zero. This gives

$$\nabla_l \zeta = -\rho\zeta - \sigma\bar{\zeta} \,, \tag{E.72}$$

for some complex parameters ρ, σ. These can be interpreted as follows:

1. The imaginary part of ρ is the twist and generates rotations of the ζ plane. It vanishes iff the congruence is hypersurface forming, $l_{[a}\nabla_b l_{c]} = 0$ which implies that there is a rescaling of l_a so that $l_a = \nabla_a u$ for some function u.

2. The real part of ρ gives the *expansion*, $\nabla_a l^a = -2\rho$ and the area element of the orthogonal transverse plane evolves by

   $$A = -im_a dx^a \wedge \bar{m}_b dx^b,$$

   satisfies

   $$\mathcal{L}_l A = -2\rho A \tag{E.73}$$

3. The complex scalar σ is the shear in the sense that a circle in the ζ plane evolves into an ellipse.

4. Equation (E.71) implies the geodesic deviation equation

   $$\nabla_l \nabla_l V^a = l^b l^c V^d R_{bdc}{}^a \tag{E.74}$$

and this combines with (E.72) to give the *Sachs equations*

$$\nabla_l \rho = \rho^2 + \sigma\bar{\sigma} + \Phi_{00} \tag{E.75}$$

$$\nabla_l \sigma = (\rho + \bar{\rho})\sigma + \Psi_0 \tag{E.76}$$

Here $\Psi_0 = C_{abcd} l^a m^b l^c m^d$, $\Phi_{00} = \Phi_{ab} l^a l^b = -\frac{1}{2} R_{ab} l^a l^b$ and is positive when the dominant energy condition is satisfied. An important consequence for horizons and singularity theorems is that the whole RHS of (E.75) is manifestly positive definite.

5. If a null hypersurface has vanishing shear, then it has the intrinsic geometry of a light cone or null hyperplane in Minkowski space up to scaling (i.e. the metric restricts to a multiple of $d\zeta d\bar{\zeta}$ on $\mathbb{R}^3$ or $S^2 \times \mathbb{R}$ where $l^a \partial_a = \partial_v$ for a third coordinate v).

Horizons and black hole thermodynamics

For an asymptotically flat space-time, we define

Definition E.2. The *event horizon* $\mathcal{H}$ is the boundary of the past $J^-(\mathscr{I}^+)$ of $\mathscr{I}^+$, that is, it is the boundary of the region from which it is possible to escape to infinity along a causal curve.

Much is known about event horizons under reasonable assumptions appropriate to isolated systems that settle down:

- $\mathcal{H}$ is a null hypersurface being the boundary of a past set (it clearly cannot be time-like as causal paths could then cross both ways, and if it were space-like there would be regions to its past that could not exit to $\mathscr{I}^+$).
- $\mathcal{H}$ is ruled (or foliated) by complete null geodesics.
- If $\mathscr{I}$ has topology $S^2 \times \mathbb{R}$, as appropriate for the exterior of an isolated system, then so does $\mathcal{H}$, with the $\mathbb{R}$ factor being the null geodesics.
- The cross-sectional area is bounded above.

This is a rather excessively global definition that requires knowledge of the whole space-time. One can also define with just local knowledge:

Definition E.3. a *closed trapped surface* is a two-surface of topology S^2 such that the outward pointing null geodesics have nonpositive expansion (i.e., the area will drop or be constant in any outward going null direction or $\rho \geq 0$ where ρ is the spin coefficient in the definition of the Sachs equation).

Penrose's original singularity theorem deduces the existence of a singularity (in the form of geodesic incompleteness) from the existence of such a closed trapped surface. It is easy to see from the signs in the Sachs equations and following the outward going null geodesic normals off hte surface that a closed trapped surface leads to:

Definition E.4. an *apparent horizon* is a null hypersurface of topology $S^2 \times \mathbb{R}$ such that the expansion of the outward going null rays is nonpositive (i.e., the area is non-increasing to the future).

The first of the Sachs equation for a null geodesic congruence generated by l gives

$$\nabla_l \rho = \rho^2 + \sigma\bar{\sigma} + \Phi_{00} \geq \rho^2 . \tag{E.77}$$

Thus if $\rho \geq 0$ then it cannot decrease. (Recall that if A is the area element, $\mathcal{L}_l A = -2\rho A$.)

However, Penrose's theorem doesnt deduce the location of the singularity! In particular it is not clear that an apparent horizon is hidden inside an event horizon and the following is open:

The cosmic censorship hypothesis: All singularities that arise from evolving from an initial data hypersurface are hidden behind an event horizon and so cannot be seen from infinity.

Generally speaking we assume that a black hole settles down to being stationary or static. Then, the event horizon must settle down to a null hypersurface with finite cross sectional area (otherwise geodesics will be escaping to infinity). Once a black hole horizon settles down, its area is constant. Assuming that the black hole is becoming stationary or static, it follows that it is (under suitable analyticity assumptions) a *Killing horizon*:

Definition E.5. A Killing horizon is a null hypersurface on which a Killing vector k_a becomes null, so that the surface is defined by $k_a k^a = 0$ and $k^a \neq 0$. Thus k^a is tangent to the null geodesic generators of the horizon.

The fact that k^a is Killing means that we must have $\rho = \sigma = 0$.

It follows from the black hole uniqueness theorems that, even if we started from some collapse scenario, the final black hole, if essentially static or stationary, is Kerr Newman or Schwarzschild. These all have a similar structure to Schwarzschild in that they can be continued analytically back to a point where the standard future event horizon intersects a past one at a 2-surface C.

Definition E.6. Such a Killing Horizon is said to be a bifurcate Killing horizon if there exists cross-section C of topology S^2 on which k^a vanishes—it is bifurcate because then in a neighbourhood of there is a transverse horizon such that $k^a = U\partial_U - V\partial_V$ as for the crossover in Schwarzschild.

On a Killing horizon we can define

Definition E.7. The *surface gravity* κ is defined by

$$\nabla_a k_b k^b = -2\kappa k_a , \quad \text{or } k^a \nabla_a k^b = \kappa k^b , \tag{E.78}$$

where in the static case, k_a is understood to be normalized to have $k_a k^a = 1$ at large distances. For Schwarzschild $\kappa = 1/4m$.

Black hole thermodynamics starts with the Bekenstein bound on the entropy S: in a region of radius R and mass-energy E the entropy is constrained by

$$S < \frac{2\pi k R E}{\hbar c} \tag{E.79}$$

where we have not set the usual fundamental constants $k, \hbar, c$ to unity. It was arrived at by consideration of throwing objects with entropy into black holes and trying to avoid violations of the second law arising from the black hole eating entropy. In this view, the black does have entropy

$$S_{BH} = \frac{kA}{4G} \tag{E.80}$$

and this is taken to be the maximal entropy state, i.e., the Bekenstein bound is saturated by the black hole entropy.

Classically, one does not think of black holes as having microstates that could give rise to an entropy in view of the black hole uniqueness theorems. These seem to imply that the black hole state is unique, whereas an enropy suggests the existence of many equally likely microstates compatible with given macroscopic observables. The black hole entropy is usually understood as having its origin in quantum gravity.

This chain of reasoning subsequently led to the *Holographic principle*, that the maximum number N of states in a spatial region of radius R satisfies

$$N < \exp S_{BH}(R) \tag{E.81}$$

This comes from the definition of entropy of a system as $S := -\sum_i p_i \log p_i$ where p_i is the probability of the ith state. If the system is equidistributed, $p_i = 1/N$, where N is the number of states (the dimension of the Hilbert space of the system) we obtain $S = \log N$. This is counter-intuitive without general relativity because one thinks of the number of states in a region as being the exponential of the volume rather than the area. However, gravitational collapse reduces this if there is too much matter (too many particles) and indeed the vast bulk of the entropy is undertood to be gravitational.

The second law of thermodynamics: if the entropy is equated with the area of the event horizon in black hole thermodynamics, the 2nd law states that it can only increase. We have the area theorem

Theorem E.3. *The area of an event horizon is non-decreasing.*

Proof: This is a simple consequence of the Sachs (or Raychauduri) equations

$$\dot{\rho} = \rho^2 + |\sigma|^2 + \Phi_{00}\,. \tag{E.82}$$

This shows that in particular $\dot{\rho} \geq \rho^2$ when the dominant energy condition is satisfied. Thus, if $\rho = \rho_0 > 0$ at some affine parameter value $t = 0$ on the generator, it is bounded below by the solution

$$\rho_0(t) = \frac{\rho_0}{1 - \rho_0 t}\,, \tag{E.83}$$

the solution to $\dot{\rho} = \rho^2$ with the same initial condition. Thus $\rho \to \infty$ in finite time. This introduces a *cusp* after which the null geodesic must then leave the horizon (see picture), contradicting its being a generator of $\mathcal{H}$. □

The first law of black hole thermodynamics: for a variation of a closed system with rotation and charge can be stated as

$$dE = TdS + \Omega dJ + \Phi_H dQ \tag{E.84}$$

Here E is the total energy, T the temerature, Ω the angular velocity, J the angular momentum, Q the charge and ϕ the electrostatic potential. In the context of black holes, the total energy is the mass, we identify the temperature with the surface gravity by

$$T = \kappa/2\pi \tag{E.85}$$

and S with the area.

For Reissner Nordstrom, $\Phi = \Phi_H$ is the electric potential at the horizon and Q the total charge, and, in the case of the Kerr solution, Ω is the angular velocity, and J the angular momentum.

There are a number of strategies for proving these formulae. The most basic is to simply establish sufficient relations between the various quantities $(M, A, \Omega_H, J, \phi_H, Q)$ as can be read off from the black hole metric, and then to differentiate it. The simplest example is for Schwarschild where the area is that associated with the Schwarzschild radius $r = 2M$, $A = 16\pi M^2$, upon which differentiation yields

$$dM = \frac{dA}{32\pi M} = \frac{\kappa}{2\pi} dA, \tag{E.86}$$

giving the most basic version.

If we wish to introduce charge, we must consider the Reissner-Nordstrom solution

$$ds^2 = \frac{\Delta(r)}{r^2} dt^2 - \frac{r^2}{\Delta(r)} dr^2 - r^2 ds^2_{S^2}, \qquad \Delta(r) = r^2 - 2Mr + Q^2. \tag{E.87}$$

This satisfies the Einstein equations with electromagnetic potential

$$A = \frac{Q}{r} dt\,. \tag{E.88}$$

The Killing horizons are where $\Delta = 0$ giving

$$r_\pm = M \pm \sqrt{M^2 - Q^2}, \tag{E.89}$$

assuming $Q < M$. The outer one is the event horizon and a short calculation shows that differentiating the obvious relation $A = \pi r_+^2$ now gives

$$dM = \frac{\sqrt{M^2 - Q^2}}{2\pi r_+^2} dA + \frac{Q}{r_+} dQ\,. \tag{E.90}$$

The coefficient of dQ is indeed the value of the potential at the horizon. It is a more complicated task to see that the surface gravity does indeed appropriately give the coefficient of dA (see the exercises). Even more nontrivially, this works as stated above for the Kerr-Newman solution where there is also rotation.

The zeroth law of Black hole thermodynamics: In the analogy with thermodynamics, κ plays the role of temperature via $T = \kappa/2\pi$. The zeroth law is that the temperature is constant in equilibrium. It is easy to see that the surface gravity κ is constant up the generators of $\mathcal{H}$, because k^a is Killing. We will see in the problems that for a bifurcate Killing horizon κ is actually constant over the horizon. Hence it is constant everywhere. The next result follows in greater generality but we will not prove it here.

There is also a third law, that the entropy of an object at absolute zero is zero. This fails for black holes for a number of reasons, but a vaguer version, that once cannot approach absolute zero temperature with a finite number of processes does seem reasonable, as $T \to 0$ corresponds to $M \to \infty$.

The glaring omission in all this is of course that the temperature of a black hole classically would seem to have to be zero. This will be seen to be resolved by Hawking radiation.

Appendix F

Hypergeometric functions

In this appendix we review the definition and various properties of hypergeometric functions. For further reference see for example [AAR99].

The hypergeometric function is a solution of Euler's hypergeometric differential equation,

$$z(1-z)\partial_z^2 F + [\gamma - (\alpha+\beta+1)z]\,\partial_z F - \alpha\beta F = 0\,. \tag{F.1}$$

which has three regular singular points at $z = 0, 1$ and ∞. Any second order linear equation with three regular singular points can be converted to this equation through a change of variables.

For $|z| < 1$, the hypergeometric function can be defined through the following series expansion

$${}_2F_1(\alpha,\beta;\gamma|z) = \sum_{n=0}^{\infty} \frac{(\alpha)_n(\beta)_n}{(\gamma)_n}\frac{z^n}{n!}\,, \tag{F.2}$$

where we introduced the (rising) Pochhammer symbol

$$(q)_n = \frac{\Gamma(q+n)}{\Gamma(q)}\,. \tag{F.3}$$

when either α or β is a non-positive integer this series terminates in which case the hypergeometric function reduces to a polynomial. For complex $|z| \geq 1$ it can be analytically continued along any path that avoids the branch points at $z = 1$ and $z = \infty$.

Depending on the sign of $\mathrm{Re}(\gamma - \alpha - \beta)$ we find the following behaviour near $z = 1$,

$${}_2F_1(\alpha,\beta;\gamma|z) \overset{z\to 1}{\simeq} \begin{cases} \frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)}\,, & \text{for} \quad \mathrm{Re}(\gamma-\alpha-\beta) > 0\,, \\ \frac{\Gamma(\gamma)\Gamma(\alpha+\beta-\gamma)}{\Gamma(\alpha)\Gamma(\beta)}(1-z)^{\gamma-\alpha-\beta}\,, & \text{for} \quad \mathrm{Re}(\gamma-\alpha-\beta) < 0\,. \end{cases} \tag{F.4}$$

In addition,we have the following identities for the analytic continuation of the hypergeometric functions,

$$\begin{aligned} {}_2F_1(\alpha,\beta;\gamma|z) =& \frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)}{}_2F_1(\alpha,\beta;1+\alpha+\beta-\gamma|1-z) \\ &+ \frac{\Gamma(\gamma)\Gamma(\alpha+\beta-\gamma)}{\Gamma(\alpha)\Gamma(\beta)}(1-z)^{\gamma-\alpha-\beta}{}_2F_1(\gamma-\alpha,\gamma-\beta;1_\gamma-\alpha-\beta|1-z)\,. \end{aligned} \tag{F.5}$$

and

$${}_2F_1(\alpha,\beta;\gamma|z) = (1-z)^{\gamma-\alpha-\beta}{}_2F_1(\gamma-\alpha,\gamma-\beta;\gamma|z)\,. \tag{F.6}$$

Appendix G

A mini introduction to CFT

This section briefly describes the basic concepts necessary to formulate a non-perturbative definition of CFT.

G.1 Conformal Transformations

For simplicity, in most formulas, we will consider Euclidean signature. We start by discussing conformal transformations of $\mathbb{R}^d$ in Cartesian coordinates,

$$ds^2 = \delta_{\mu\nu} dx^\mu dx^\nu \ . \tag{G.1}$$

A conformal transformation is a coordinate transformation that preserves the form of the metric tensor up to a scale factor,

$$\delta_{\mu\nu} \frac{d\tilde{x}^\mu}{dx^\alpha} \frac{d\tilde{x}^\nu}{dx^\beta} = \Omega^2(x) \delta_{\alpha\beta} \ . \tag{G.2}$$

In other words, a conformal transformation is a local dilatation.

Exercise G.1. *Show that, for $d > 2$, the most general infinitesimal conformal transformation is given by $\tilde{x}^\mu = x^\mu + \epsilon^\mu(x)$ with*

$$\epsilon^\mu(x) = a^\mu + \lambda x^\mu + m^{\mu\nu} x_\nu + x^2 b^\mu - 2x^\alpha b_\alpha x^\mu \ . \tag{G.3}$$

In spacetime dimension $d > 2$, conformal transformations form the group SO$(d+1,1)$. The generators P_μ and $M_{\mu\nu}$ correspond to translation and rotations and they are present in any relativistic invariant QFT. In addition, we have the generators of dilatations D and special conformal transformations K_μ. It is convenient to think of the special conformal transformations as the composition of an inversion followed by a translation followed by another inversion. Inversion is the conformal transformation[1]

$$x^\mu \to \frac{x^\mu}{x^2} \ . \tag{G.4}$$

Exercise G.2. *Verify that inversion is a conformal transformation.*

The form of the generators of the conformal algebra acting on functions can be obtained from

$$\phi\left(x^\mu + \epsilon^\mu(x)\right) = \left[1 + i\, a^\mu P_\mu - \lambda D + \frac{i}{2} m^{\mu\nu} M_{\mu\nu} + i\, b^\mu K_\mu\right] \phi\left(x^\mu\right) \ , \tag{G.5}$$

[1] Inversion is outside the component of the conformal group connected to the identity. Thus, it is possible to have CFTs that are not invariant under inversion. In fact, CFTs that break parity also break inversion.

which leads to [2]

$$P_\mu = -i\partial_\mu \,, \qquad\qquad D = -x^\mu \partial_\mu \,, \tag{G.6}$$
$$M_{\mu\nu} = -i\left(x_\mu \partial_\nu - x_\nu \partial_\mu\right) \,, \qquad\qquad K_\mu = 2i x_\mu x^\nu \partial_\nu - i\, x^2 \partial_\mu \,. \tag{G.7}$$

Exercise G.3. *Show that the generators obey the following commutation relations*

$$\begin{gathered}\left[D, P_\mu\right] = P_\mu \,, \qquad \left[D, K_\mu\right] = -K_\mu \,, \qquad \left[K_\mu, P_\nu\right] = 2\delta_{\mu\nu} D - 2i\, M_{\mu\nu} \,, \\ \left[M_{\mu\nu}, P_\alpha\right] = i\left(\delta_{\mu\alpha} P_\nu - \delta_{\nu\alpha} P_\mu\right) \,, \qquad \left[M_{\mu\nu}, K_\alpha\right] = i\left(\delta_{\mu\alpha} K_\nu - \delta_{\nu\alpha} K_\mu\right) \,, \\ \left[M_{\alpha\beta}, M_{\mu\nu}\right] = i\left(\delta_{\alpha\mu} M_{\beta\nu} + \delta_{\beta\nu} M_{\alpha\mu} - \delta_{\beta\mu} M_{\alpha\nu} - \delta_{\alpha\nu} M_{\beta\mu}\right) \,. \end{gathered} \tag{G.8}$$

G.2 Local Operators

Local operators are divided into two types: primary and descendant. Descendant operators are operators that can be written as (linear combinations of) derivatives of other local operators. Primary operators can not be written as derivatives of other local operators. Primary operators at the origin are annihilated by the generators of special conformal transformations. Moreover, they are eigenvectors of the dilatation generator and form irreducible representations of the rotation group SO(d),

$$\left[K_\mu, \mathcal{O}(0)\right] = 0 \,, \quad \left[D, \mathcal{O}(0)\right] = \Delta\, \mathcal{O}(0) \,, \quad \left[M_{\mu\nu}, \mathcal{O}_A(0)\right] = \left[M_{\mu\nu}\right]_A^B \mathcal{O}_B(0) \,. \tag{G.9}$$

Correlation functions of scalar primary operators obey

$$\langle \mathcal{O}_1(\tilde{x}_1) \dots \mathcal{O}_n(\tilde{x}_n) \rangle = \left|\frac{\partial \tilde{x}}{\partial x}\right|_{x_1}^{-\frac{\Delta_1}{d}} \dots \left|\frac{\partial \tilde{x}}{\partial x}\right|_{x_n}^{-\frac{\Delta_n}{d}} \langle \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle \tag{G.10}$$

for all conformal transformations $x \to \tilde{x}$. As explained above, it is sufficient to impose Poincaré invariance and this transformation rule under inversion,

$$\left\langle \mathcal{O}_1\left(\frac{x_1}{x_1^2}\right) \dots \mathcal{O}_n\left(\frac{x_n}{x_n^2}\right) \right\rangle = \left(x_1^2\right)^{\Delta_1} \dots \left(x_n^2\right)^{\Delta_n} \langle \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle \,.$$

This implies that vacuum one-point functions $\langle \mathcal{O}(x) \rangle$ vanish except for the identity operator (which is the unique operator with $\Delta = 0$). It also fixes the form of the two and three point functions,

$$\left\langle \mathcal{O}_i(x) \mathcal{O}_j(y) \right\rangle = \frac{\delta_{ij}}{(x-y)^{2\Delta_i}} \,, \tag{G.11}$$
$$\langle \mathcal{O}_1(x_1) \mathcal{O}_2(x_2) \mathcal{O}_3(x_3) \rangle = \frac{C_{123}}{|x_{12}|^{\Delta_1+\Delta_2-\Delta_3} \, |x_{13}|^{\Delta_1+\Delta_3-\Delta_2} \, |x_{23}|^{\Delta_2+\Delta_3-\Delta_1}} \,,$$

where we have normalized the operators to have unit two point function.

The four-point function is not fixed by conformal symmetry because with four points one can construct

[2] We define the dilatation generator D in a non-standard fashion so that it has real eigenvalues in unitary CFTs.

two independent conformal invariant cross-ratios

$$u = z\bar{z} = \frac{x_{12}^2 x_{34}^2}{x_{13}^2 x_{24}^2}\,, \qquad v = (1-z)(1-\bar{z}) = \frac{x_{14}^2 x_{23}^2}{x_{13}^2 x_{24}^2}\,. \tag{G.12}$$

The general form of the four point function is

$$\big\langle \mathcal{O}(x_1)\dots\mathcal{O}(x_4)\big\rangle = \frac{\mathcal{A}(u,v)}{\left(x_{13}^2 x_{24}^2\right)^{\Delta}}\,. \tag{G.13}$$

G.3 **Ward identities**

To define the stress-energy tensor it is convenient to consider the theory in a general background metric $g_{\mu\nu}$. Formally, we can write

$$\langle \mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\rangle_g = \frac{1}{Z[g]}\int [d\phi] e^{-S[\phi,g]}\mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\,, \tag{G.14}$$

where $Z[g] = \int [d\phi] e^{-S[\phi,g]}$ is the partition function for the background metric $g_{\mu\nu}$. Recalling the classical definition

$$T^{\mu\nu}(x) = -\frac{2}{\sqrt{g}}\frac{\delta S}{\delta g_{\mu\nu}(x)}\,, \tag{G.15}$$

it is natural to define the quantum stress-energy tensor operator via the equation

$$\frac{Z[g+\delta g]}{Z[g]} = 1 + \frac{1}{2}\int dx\sqrt{g}\delta g_{\mu\nu}(x)\,\langle T^{\mu\nu}(x)\rangle_g + O(\delta g^2)\,, \tag{G.16}$$

and

$$\begin{aligned}
&\langle \mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\rangle_{g+\delta g} - \langle \mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\rangle_g \\
=&\frac{1}{2}\int dx\sqrt{g}\delta g_{\mu\nu}(x)\big[\langle T^{\mu\nu}(x)\mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\rangle_g \\
&\qquad - \langle T^{\mu\nu}(x)\rangle_g\,\langle \mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\rangle_g\big] + O(\delta g^2)\,.
\end{aligned} \tag{G.17}$$

Under an infinitesimal coordinate transformation $\tilde{x}^\mu = x^\mu + \epsilon^\mu(x)$, the metric tensor changes $\tilde{g}_{\mu\nu} = g_{\mu\nu} - \nabla_\mu\epsilon_\nu - \nabla_\nu\epsilon_\mu$ but the physics should remain invariant. In particular, the partition function $Z[g] = Z[\tilde{g}]$ and the correlation functions [3]

$$\langle \mathcal{O}_1(\tilde{x}_1)\dots\mathcal{O}_n(\tilde{x}_n)\rangle_{\tilde{g}} = \langle \mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\rangle_g\,, \tag{G.18}$$

[3] If the operators are not scalars (e.g. if they are vector operators) then one also needs to take into account the rotation of their indices.

do not change. This leads to the conservation equation $\left\langle \nabla_\mu T^{\mu\nu}(x) \right\rangle_g$ and

$$\sum_{i=1}^{n} \epsilon^\mu(x_i) \frac{\partial}{\partial x_i^\mu} \langle \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle_g \qquad \text{(G.19)}$$
$$= -\int dx \sqrt{g} \epsilon_\nu(x) \left\langle \nabla_\mu T^{\mu\nu}(x) \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \right\rangle_g$$

for all $\epsilon^\mu(x)$ that decays sufficiently fast at infinity. Thus $\nabla_\mu T^{\mu\nu} = 0$ up to contact terms.

Correlation functions of primary operators transform homogeneously under Weyl transformations of the metric [4]

$$\langle \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle_{\Omega^2 g} = \frac{\langle \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle_g}{[\Omega(x_1)]^{\Delta_1} \dots [\Omega(x_n)]^{\Delta_n}} \,. \qquad \text{(G.20)}$$

Exercise G.4. *Show that this transformation rule under local rescalings of the metric (together with coordinate invariance) implies (G.10) under conformal transformations.*

Consider now an infinitesimal Weyl transformation $\Omega = 1 + \omega$, which corresponds to a metric variation $\delta g_{\mu\nu} = 2\omega g_{\mu\nu}$. From (G.17) and (G.20) we conclude that

$$\sum_{i=1}^{n} \Delta_i \, \omega(x_i) \langle \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle_g$$
$$= -\int dx \sqrt{g} \, \omega(x) g_{\mu\nu} \big[\langle T^{\mu\nu}(x) \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle_g \qquad \text{(G.21)}$$
$$- \langle T^{\mu\nu}(x) \rangle_g \langle \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle_g \big] \,.$$

Consider the following codimension 1 integral over the boundary of a region B, [5]

$$I = \int_{\partial B} dS_\mu \epsilon_\nu(x) \big[\langle T^{\mu\nu}(x) \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle_g \qquad \text{(G.22)}$$
$$- \langle T^{\mu\nu}(x) \rangle_g \langle \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle_g \big] \,.$$

One can think of this as the total flux of the current $\epsilon_\nu T^{\mu\nu}$, where $\epsilon_\nu(x)$ is an infinitesimal conformal transformation. Gauss law tells us that this flux should be equal to the integral of the divergence of the current

$$\nabla_\mu (\epsilon_\nu T^{\mu\nu}) = \epsilon_\nu \nabla_\mu T^{\mu\nu} + \nabla_\mu \epsilon_\nu T^{\mu\nu} = \epsilon_\nu \nabla_\mu T^{\mu\nu} + \frac{1}{d} \nabla_\alpha \epsilon^\alpha g_{\mu\nu} T^{\mu\nu} \,, \qquad \text{(G.23)}$$

where we used the symmetry of the stress-energy tensor $T^{\mu\nu} = T^{\nu\mu}$ and the definition of an infinitesimal conformal transformation $\nabla_\mu \epsilon_\nu + \nabla_\nu \epsilon_\mu = \frac{2}{d} \nabla_\alpha \epsilon^\alpha g_{\mu\nu}$. Using Gauss law and (G.19) and (G.21) we conclude that

$$I = -\sum_{x_i \in B} \left[\epsilon^\mu(x_i) \frac{\partial}{\partial x_i^\mu} + \frac{\Delta_i}{d} \nabla_\alpha \epsilon^\alpha(x_i) \right] \langle \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle_g \,. \qquad \text{(G.24)}$$

[4] In general, the partition fungion is not invariant in even dimensions. This is the Weyl anomaly $Z[\Omega^2 g] = Z[g] e^{-S_{Weyl}[\Omega, g]}$.

[5] In the notation of the *Conformal Bootstrap* chapter [?] this is the topological operator $Q_\epsilon[\partial B]$ inserted in the correlation function $\langle \mathcal{O}_1(x_1) \dots \mathcal{O}_n(x_n) \rangle_g$.

The equality of (G.22) and (G.24) for any infinitesimal conformal transformation (G.3) is the most useful form of the conformal Ward identities.

Exercise G.5. *Conformal symmetry fixes the three-point function of a spin 2 primary operator and two scalars up to an overall constant,* [6]

$$\langle \mathcal{O}(x_1)\mathcal{O}(x_2)T^{\mu\nu}(x_3)\rangle = C_{12T}\frac{H^{\mu\nu}(x_1,x_2,x_3)}{|x_{12}|^{2\Delta-d+2}|x_{13}|^{d-2}|x_{23}|^{d-2}}\,, \tag{G.25}$$

where

$$H^{\mu\nu} = V^\mu V^\nu - \frac{1}{d}V_\alpha V^\alpha \delta^{\mu\nu}\,, \qquad V^\mu = \frac{x_{13}^\mu}{x_{13}^2} - \frac{x_{23}^\mu}{x_{23}^2}\,. \tag{G.26}$$

Write the conformal Ward identity (G.22)=(G.24) for the three point function $\langle T^{\mu\nu}(x)\mathcal{O}(0)\mathcal{O}(y)\rangle$ for the case of an infinitesimal dilation $\epsilon^\mu(x) = \lambda x^\mu$ and with the surface ∂B being a sphere centred at the origin and with radius smaller than $|y|$. Use this form of the conformal Ward identity in the limit of an infinitesimally small sphere ∂B and formula (G.25) for the three point function to derive

$$C_{\mathcal{O}\mathcal{O}T} = -\frac{d\Delta}{d-1}\frac{1}{S_d}\,, \tag{G.27}$$

where $S_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ is the volume of a $(d-1)$-dimensional unit sphere.

G.4 State-Operator Map

Consider $\mathbb{R}^d$ in spherical coordinates. Writing the radial coordinate as $r = e^\tau$ we find

$$ds^2 = dr^2 + r^2 d\Omega_{d-1}^2 = e^{2\tau}\left(d\tau^2 + d\Omega_{d-1}^2\right)\,. \tag{G.28}$$

Thus, the cylinder $\mathbb{R}\times S^{d-1}$ can be obtained as a Weyl transformation of euclidean space $\mathbb{R}^d$.

Exercise G.6. *Compute the two-point function of a scalar primary operator on the cylinder using the Weyl transformation property (G.20).*

A local operator inserted at the origin of $\mathbb{R}^d$ prepares a state at $\tau = -\infty$ on the cylinder. On the other hand, a state on a constant time slice of the cylinder can be propagated backwards in time until it corresponds to a boundary condition on a arbitrarily small sphere around the origin of $\mathbb{R}^d$, which defines a local operator. Furthermore, time translations on the cylinder correspond to dilatations on $\mathbb{R}^d$. This teaches us that the spectrum of the dilatation generator on $\mathbb{R}^d$ is the same as the energy spectrum for the theory on $\mathbb{R}\times S^{d-1}$. [7]

[6] You can try to derive this formula using the embedding space formalism of section G.7.

[7] More precisely, there can be a constant shift equal to the Casimir energy of the vacuum on S^{d-1}, which is related with the Weyl anomaly. In $d=2$, this gives the usual energy spectrum $\left(\Delta - \frac{c}{12}\right)\frac{1}{L}$ where c is the central charge and L is the radius of S^1.

G.5 Operator Product Expansion

The Operator Product Expansion (OPE) between two scalar primary operators takes the following form

$$\mathcal{O}_i(x)\mathcal{O}_j(0) = \sum_k C_{ijk}|x|^{\Delta_k-\Delta_i-\Delta_j}\left[\mathcal{O}_k(0) + \underbrace{\beta\, x^\mu \partial_\mu \mathcal{O}_k(0) + \dots}_{\text{descendants}}\right] \tag{G.29}$$

where β denotes a number determined by conformal symmetry. For simplicity we show only the contribution of a scalar operator $\mathcal{O}_k$. In general, in the OPE of two scalars there are primary operators of all spins.

Exercise G.7. *Compute β by using this OPE inside a three-point function.*

The OPE has a finite radius of convergence inside correlation functions. This follows from the state operator map with an appropriate choice of origin for radial quantization.

G.6 Conformal Bootstrap

Using the OPE successively one can reduce any n—point function to a sum of one-point functions, which all vanish except for the identity operator. Thus, knowing the operator content of the theory, *i.e.* the scaling dimensions Δ and $SO(d)$ irreps $\mathcal{R}$ of all primary operators, and the OPE coefficients C_{ijk},[8] one can determine all correlation functions of local operators. This set of data is called CFT data because it essentially defines the theory. [9] The CFT data is not arbitrary, it must satisfy several constraints:

- **OPE associativity** - Different ways of using the OPE to compute a correlation function must give the same result. This leads to the conformal bootstrap equations described below.
- **Existence of stress-energy tensor** - The stress-energy tensor $T_{\mu\nu}$ is a conserved primary operator (with $\Delta = d$) whose correlation functions obey the conformal Ward identities.
- **Unitarity** - In our Euclidean context this corresponds to reflection positivity and it implies lower bounds on the scaling dimensions. It also implies that one can choose a basis of real operators where all OPE coefficients are real. In the context of statistical physics, there are interesting non-unitary CFTs.

It is sufficient to impose OPE associativity for all four-point functions of the theory. For a four-point function of scalar operators, the bootstrap equation reads

$$\sum_k C_{12k}C_{k34}G^{(12)(34)}_{\Delta_k,l_k}(x_1,\dots,x_4) = \sum_q C_{13q}C_{q24}G^{(13)(24)}_{\Delta_q,l_q}(x_1,\dots,x_4)\,,$$

[8]For primary operators $\mathcal{O}_1$, $\mathcal{O}_2$, $\mathcal{O}_3$ transforming in non-trivial irreps of $SO(d)$ there are several OPE coefficients C_{123}. The number of OPE coefficients C_{123} is given by the number of symmetric traceless tensor representations that appear in the tensor product of the 3 irreps of $SO(d)$ associated to $\mathcal{O}_1$, $\mathcal{O}_2$ and $\mathcal{O}_3$.

[9]However, there are observables besides the vacuum correlation functions of local operators. It is also interesting to study non-local operators (line operators, surface operators, boundary conditions, etc) and correlation functions in spaces with non-trivial topology (for example, correlators at finite temperature).

where $G_{\Delta,l}$ are conformal blocks, which encode the contribution from a primary operator of dimension Δ and spin l and all its descendants.

G.7 Embedding Space Formalism

The conformal group $SO(d+1,1)$ acts naturally on the space of light rays through the origin of $\mathbb{R}^{d+1,1}$,

$$-\left(P^0\right)^2+\left(P^1\right)^2+\cdots+\left(P^{d+1}\right)^2=0\ . \tag{G.30}$$

A section of this light-cone is a $d-$dimensional manifold where the CFT lives. For example, it is easy to see that the Poincaré section $P^0+P^{d+1}=1$ is just $\mathbb{R}^d$. To see this parametrize this section using

$$P^0(x)=\frac{1+x^2}{2}\ , \qquad P^\mu(x)=x^\mu\ , \qquad P^{d+1}(x)=\frac{1-x^2}{2}\ , \tag{G.31}$$

with $\mu=1,\ldots,d$ and $x^\mu\in\mathbb{R}^d$ and compute the induced metric. In fact, any conformally flat manifold can be obtained as a section of the light-cone in the embedding space $\mathbb{R}^{d+1,1}$. Using the parametrization $P^A=\Omega(x)P^A(x)$ with $x^\mu\in\mathbb{R}^d$, one can easily show that the induced metric is simply given by $ds^2=\Omega^2(x)\delta_{\mu\nu}dx^\mu dx^\nu$. With this is mind, it is natural to extend a primary operator from the physical section to the full light-cone with the following homogeneity property

$$\mathcal{O}(\lambda P)=\lambda^{-\Delta}\mathcal{O}(P)\ , \qquad \lambda\in\mathbb{R}\ . \tag{G.32}$$

This implements the Weyl transformation property (G.20). One can then compute correlation functions directly in the embedding space, where the constraints of conformal symmetry are just homogeneity and $SO(d+1,1)$ Lorentz invariance. Physical correlators are simply obtained by restricting to the section of the light-cone associated with the physical space of interest. This idea goes back to Dirac [**?**] and has been further develop by many authors [**?**, **?**, **?**, **?**, **?**, **?**, **?**].

Exercise G.8. *Rederive the form of two and three point functions of scalar primary operators in $\mathbb{R}^d$ using the embedding space formalism.*

Vector primary operators can also be extended to the embedding space. In this case, we impose

$$P^A\mathcal{O}_A(P)=0\ , \qquad \mathcal{O}_A(\lambda P)=\lambda^{-\Delta}\mathcal{O}_A(P)\ , \qquad \lambda\in\mathbb{R}\ , \tag{G.33}$$

and the physical operator is obtained by projecting the indices to the section,

$$\mathcal{O}_\mu(x)=\frac{\partial P^A}{\partial x^\mu}\mathcal{O}_A(P)\bigg|_{P^A=P^A(x)}\ . \tag{G.34}$$

Notice that this implies a redundancy: $\mathcal{O}_A(P)\to\mathcal{O}_A(P)+P_A\Lambda(P)$ gives rise to the same physical operator $\mathcal{O}(x)$, for any scalar function $\Lambda(P)$ such that $\Lambda(\lambda P)=\lambda^{-\Delta-1}\Lambda(P)$. This redundancy together with the constraint $P^A\mathcal{O}_A(P)=0$ remove 2 degrees of freedom of the $(d+2)$-dimensional vector $\mathcal{O}_A$.

Exercise G.9. *Show that the two-point function of vector primary operators is given by*

$$\left\langle \mathcal{O}^A(P_1)\mathcal{O}^B(P_2)\right\rangle = const\, \frac{\eta^{AB}\,(P_1\cdot P_2) - P_2^A P_1^B}{(-2P_1\cdot P_2)^{\Delta+1}}\ , \tag{G.35}$$

up to redundant terms.

Exercise G.10. *Consider the parametrization* $P^A = \left(P^0, P^\mu, P^{d+1}\right) = (\cosh\tau, \Omega^\mu, -\sinh\tau)$ *of the global section* $\left(P^0\right)^2 - \left(P^{d+1}\right)^2 = 1$, *where* Ω^μ *(*$\mu = 1,\dots,d$*) parametrizes a unit* $(d-1)$*–dimensional sphere,* $\Omega\cdot\Omega = 1$. *Show that this section has the geometry of a cylinder exactly like the one used for the state-operator map.*

Conformal correlation functions extended to the light-cone of $\mathbb{R}^{1,d+1}$ are annihilated by the generators of $SO(1,d+1)$

$$\sum_{i=1}^{n} J_{AB}^{(i)}\,\langle \mathcal{O}_1(P_1)\dots\mathcal{O}_n(P_n)\rangle = 0\ , \tag{G.36}$$

where $J_{AB}^{(i)}$ is the generator

$$J_{AB} = -i\left(P_A\frac{\partial}{\partial P^B} - P_B\frac{\partial}{\partial P^A}\right)\ , \tag{G.37}$$

acting on the point P_i. For a given choice of light cone section, some generators will preserve the section and some will not. The first are Killing vectors (isometry generators) and the second are conformal Killing vectors. The commutation relations give the usual Lorentz algebra

$$[J_{AB}, J_{CD}] = i\,(\eta_{AC}J_{BD} + \eta_{BD}J_{AC} - \eta_{BC}J_{AD} - \eta_{AD}J_{BC})\ . \tag{G.38}$$

Exercise G.11. *Check that the conformal algebra (G.8) follows from (G.38) and*

$$\begin{aligned} D &= -i\,J_{0,d+1}\ , & P_\mu &= J_{\mu 0} - J_{\mu,d+1}\ ,\\ M_{\mu\nu} &= J_{\mu\nu}\ , & K_\mu &= J_{\mu 0} + J_{\mu,d+1}. \end{aligned} \tag{G.39}$$

Exercise G.12. *Show that equation (G.36) for* $J_{AB} = J_{0,d+1}$ *implies time translation invariance on the cylinder*

$$\sum_{i=1}^{n}\frac{\partial}{\partial\tau_i}\,\langle \mathcal{O}_1(\tau_1,\Omega_1)\dots\mathcal{O}_n(\tau_n,\Omega_n)\rangle = 0\ , \tag{G.40}$$

and dilatation invariance on $\mathbb{R}^d$

$$\sum_{i=1}^{n}\left(\Delta_i + x_i^\mu\frac{\partial}{\partial x_i^\mu}\right)\langle \mathcal{O}_1(x_1)\dots\mathcal{O}_n(x_n)\rangle = 0\ . \tag{G.41}$$

In this case, you will need to use the differential form of the homogeneity property $P^A\frac{\partial}{\partial P^A}\mathcal{O}_i(P) = -\Delta_i\mathcal{O}_i(P)$. *It is instructive to do this exercise for the other generators as well.*

Figure G.1: Vacuum diagrams in the double line notation. Interaction vertices are marked with a small blue dot. The left diagram is planar while the diagram on the right has the topology of a torus (genus 1 surface).

G.8 Large N Factorization

Consider a $U(N)$ gauge theory with fields valued in the adjoint representation. Schematically, we can write the action as

$$S = \frac{N}{\lambda} \int dx \, \mathrm{Tr}\left[(D\Phi)^2 + c_3 \Phi^3 + c_4 \Phi^4 + \ldots \right] \tag{G.42}$$

where we introduced the 't Hooft coupling $\lambda = g_{YM}^2 N$ and c_i are other coupling constants independent of N. Following 't Hooft [**?**], we consider the limit of large N with λ kept fixed. The propagator of an adjoint field obeys

$$\left\langle \Phi^i_j \Phi^k_l \right\rangle \propto \frac{\lambda}{N} \delta^i_l \delta^k_j \tag{G.43}$$

where we used the fact that the adjoint representation can be represented as the direct product of the fundamental and the anti-fundamental representation. This suggests that one can represent a propagator by a double line, where each line denotes the flow of a fundamental index. Start by considering the vacuum diagrams in this language. A diagram with V vertices, E propagators (or edges) and F lines (or faces) scales as

$$\left(\frac{N}{\lambda}\right)^V \left(\frac{\lambda}{N}\right)^E N^F = \left(\frac{N}{\lambda}\right)^\chi \lambda^F \,, \tag{G.44}$$

where $\chi = V + F - E = 2 - 2g$ is the minimal Euler character of the two dimensional surface where the double line diagram can be embedded and g is the number of handles of this surface. Therefore, the large N limit is dominated by diagrams that can be drawn on a sphere ($g = 0$). These diagrams are called planar diagrams. For a given topology, there is an infinite number of diagrams that contribute with increasing powers of the coupling λ, corresponding to tesselating the surface with more and more faces. Figure G.1 shows two examples of vacuum diagrams in the double line notation. This topological expansion has the structure of string perturbation theory with λ/N playing the role of the string coupling. As we shall see this is precisely realized in maximally supersymmetric Yang-Mills theory (SYM).

Let us now consider single-trace local operators of the form $\mathcal{O} = c_J \mathrm{Tr}\left(\Phi^J\right)$, where c_J is a normalization constant independent of N. Adapting the argument above, it is easy to conclude that the connected correlators are given by a large N expansion of the form

$$\langle \mathcal{O}_1 \dots \mathcal{O}_n \rangle_c = \sum_{g=0}^{\infty} N^{2-n-2g} f_g(\lambda) \ , \tag{G.45}$$

which is dominated by the planars diagrams ($g = 0$). Moreover, we see that the planar two-point function is independent of N while connected higher point functions are suppressed by powers of N. This is large N factorization. In particular it implies that the two-point function of a multi-trace operator $\tilde{\mathcal{O}}(x) =: \mathcal{O}_1(x) \dots \mathcal{O}_k(x)$: is dominated by the product of the two-point functions of its single-trace constituents

$$\left\langle \tilde{\mathcal{O}}(x) \tilde{\mathcal{O}}(y) \right\rangle \approx \prod_i \langle \mathcal{O}_i(x) \mathcal{O}_i(y) \rangle = \frac{1}{(x-y)^{2\sum_i \Delta_i}} \ , \tag{G.46}$$

where we assumed that the single-trace operators were scalar conformal primaries properly normalized. We conclude that the scaling dimension of the multi-trace operator $\tilde{\mathcal{O}}$ is given by $\sum_i \Delta_i + O(1/N^2)$. In other words, the space of local operators in a large N CFT has the structure of a Fock space with single-trace operators playing the role of single particle states of a weakly coupled theory. This is the form of large N factorization relevant for AdS/CFT. However, notice that conformal invariance was not important for the argument. It is well known that large N factorization also occurs in confining gauge theories. Physically, it means that colour singlets (like glueballs or mesons) interact weakly in large N gauge theories (see [**?**] for a clear summary).

The stress tensor has a natural normalization that follows from the action, $T_{\mu\nu} \sim \frac{N}{\lambda} \mathrm{Tr}\left(\partial_\mu \Phi \partial_\nu \Phi\right)$. This leads to the large N scaling

$$\left\langle T_{\mu_1 \nu_1}(x_1) \dots T_{\mu_n \nu_n}(x_n) \right\rangle_c \sim N^2 \ , \tag{G.47}$$

which will be important below. This normalization of $T_{\mu\nu}$ is also fixed by the Ward identities.

Bibliography

[AAR99] George E. Andrews, Richard Askey, and Ranjan Roy. *Special Functions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1999.

[AEMM19] Ahmed Almheiri, Netta Engelhardt, Donald Marolf, and Henry Maxfield. The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole. *JHEP*, 12:063, 2019.

[AGM+00] Ofer Aharony, Steven S. Gubser, Juan Martin Maldacena, Hirosi Ooguri, and Yaron Oz. Large N field theories, string theory and gravity. *Phys. Rept.*, 323:183–386, 2000.

[AIS78] S. J. Avis, C. J. Isham, and D. Storey. Quantum Field Theory in anti-De Sitter Space-Time. *Phys. Rev. D*, 18:3565, 1978.

[All85] Bruce Allen. Vacuum States in de Sitter Space. *Phys. Rev. D*, 32:3136, 1985.

[Ann12] Dionysios Anninos. De Sitter Musings. *Int. J. Mod. Phys. A*, 27:1230013, 2012.

[BBB+09] Christian Bär, Christian Becker, Romeo Brunetti, Klaus Fredenhagen, Nicolas Ginoux, Frank Pfäffle, and Alexander Strohmaier. *Quantum Field Theory on Curved Spacetimes*. Lecture Notes in Physics. Springer, 2009.

[BCH73] James M. Bardeen, B. Carter, and S. W. Hawking. The Four laws of black hole mechanics. *Commun. Math. Phys.*, 31:161–170, 1973.

[BD77] Claude W. Bernard and Anthony Duncan. Regularization and Renormalization of Quantum Field Theory in Curved Space-Time. *Annals Phys.*, 107:201, 1977.

[BD78] T. S. Bunch and P. C. W. Davies. Quantum Field Theory in de Sitter Space: Renormalization by Point Splitting. *Proc. Roy. Soc. Lond. A*, 360:117–134, 1978.

[BD84] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge Univ. Press, Cambridge, UK, 2 1984.

[Bek73] Jacob D. Bekenstein. Black holes and entropy. *Phys. Rev.*, D7:2333–2346, 1973.

[BF82] Peter Breitenlohner and Daniel Z. Freedman. Stability in Gauged Extended Supergravity. *Annals Phys.*, 144:249, 1982.

[BMS02] Raphael Bousso, Alexander Maloney, and Andrew Strominger. Conformal vacua and entropy in de Sitter space. *Phys. Rev. D*, 65:104039, 2002.

[BS05] Antonio N. Bernal and Miguel Sanchez. Smoothness of time functions and the metric splitting of globally hyperbolic space-times. *Commun. Math. Phys.*, 257:43–50, 2005.

[BT13] R. Bott and L.W. Tu. *Differential Forms in Algebraic Topology*. Graduate Texts in Mathematics. Springer New York, 2013.

[CGP14] Miguel S. Costa, Vasco Gonçalves, and João Penedones. Spinning AdS Propagators. *JHEP*, 09:064, 2014.

[CJ86] P. Candelas and B. P. Jensen. Feynman green function inside a schwarzschild black hole. *Phys. Rev. D*, 33:1596–1603, Mar 1986.

[CT68] N. A. Chernikov and E. A. Tagirov. Quantum theory of scalar fields in de Sitter space-time. *Ann. Inst. H. Poincare Phys. Theor. A*, 9:109, 1968.

[DFM+99a] Eric D'Hoker, Daniel Z. Freedman, Samir D. Mathur, Alec Matusis, and Leonardo Rastelli. Graviton and gauge boson propagators in AdS(d+1). *Nucl. Phys. B*, 562:330–352, 1999.

[DFM+99b] Eric D'Hoker, Daniel Z. Freedman, Samir D. Mathur, Alec Matusis, and Leonardo Rastelli. Graviton exchange and complete four point functions in the AdS / CFT correspondence. *Nucl. Phys. B*, 562:353–394, 1999.

[DO01] F. A. Dolan and H. Osborn. Conformal four point functions and the operator product expansion. *Nucl. Phys. B*, 599:459–496, 2001.

[Dod03] Scott Dodelson. *Modern Cosmology*. Academic Press, Amsterdam, 2003.

[dRNS15] Adrián del Río and José Navarro-Salas. Equivalence of adiabatic and dewitt-schwinger renormalization schemes. *Phys. Rev. D*, 91:064031, Mar 2015.

[EL10] Chas A. Egan and Charles H. Lineweaver. A Larger Estimate of the Entropy of the Universe. *Astrophys. J.*, 710:1825–1834, 2010.

[ESP12] Sheer El-Showk and Kyriakos Papadodimas. Emergent Spacetime and Holographic CFTs. *JHEP*, 10:106, 2012.

[FG85] C. Fefferman and C.R. Graham. Conformal invariants. *Societé Mathématique de France*, 1985.

[Fri22] Alexander Friedman. über die Krümmung des Raumes. *Z. Physik*, 10:377–386, 1922.

[Fri24] Alexander Friedman. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Z. Physik*, 21:326–332, 1924.

[FS11] Guido Festuccia and Nathan Seiberg. Rigid Supersymmetric Theories in Curved Superspace. *JHEP*, 06:114, 2011.

[FVP12] D.Z. Freedman and A. Van Proeyen. *Supergravity*. Cambridge University Press, 2012.

[Ger70] Robert P. Geroch. The domain of dependence. *J. Math. Phys.*, 11:437–439, 1970.

[GH77a] G. W. Gibbons and S. W. Hawking. Action Integrals and Partition Functions in Quantum Gravity. *Phys. Rev. D*, 15:2752–2756, 1977.

[GH77b] G. W. Gibbons and S. W. Hawking. Cosmological Event Horizons, Thermodynamics, and Particle Creation. *Phys. Rev. D*, 15:2738–2751, 1977.

[GH94] G. W. Gibbons and S. W. Hawking, editors. *Euclidean quantum gravity*. 1994.

[Har17] Thomas Hartman. Lecture notes on classical de sitter space. 2017.

[Haw76] S. W. Hawking. Breakdown of Predictability in Gravitational Collapse. *Phys. Rev. D*, 14:2460–2473, 1976.

[HH83] J. B. Hartle and S. W. Hawking. Wave Function of the Universe. *Phys. Rev. D*, 28:2960–2975, 1983.

[Kro23] Juan A. Valiente Kroon. *Conformal Methods in General Relativity*. Cambridge University Press, 2023.

[KS11] Zohar Komargodski and Adam Schwimmer. On Renormalization Group Flows in Four Dimensions. *JHEP*, 12:099, 2011.

[Kub57] R. Kubo. Statistical-mechanical theory of irreversible processes i. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12:570–586, 1957.

[Lab00] J. M. F. Labastida. Knot invariants and Chern-Simons theory. In *3rd European Congress of Mathematics: Shaping the 21st Century*, 7 2000.

[Lem31] Abbé Georges Lemaître. A Homogeneous Universe of Constant Mass and Increasing Radius accounting for the Radial Velocity of Extra-galactic Nebulae. *Monthly Notices of the Royal Astronomical Society*, 91(5):483–490, 03 1931.

[Lem33] Abbé Georges Lemaître. L'Univers en expansion. *Annales de la Société Scientifique de Bruxelles*, 53:51, January 1933.

[LM05] Jose Labastida and Marcos Mariño. *Topological Quantum Field Thoery and Four Manifolds*. Springer, 2005.

[LT98] Hong Liu and Arkady A. Tseytlin. D = 4 superYang-Mills, D = 5 gauged supergravity, and D = 4 conformal supergravity. *Nucl. Phys. B*, 533:88–108, 1998.

[LT99] Hong Liu and Arkady A. Tseytlin. On four point functions in the CFT / AdS correspondence. *Phys. Rev. D*, 59:086002, 1999.

[Mar01] M. Marino. An introduction to Donaldson-Witten theory. In *Summer School 2001 on Geometric and Topological Methods for Quantum Field Theory*, pages 265–311, 7 2001.

[Mar05] Marcos Mariño. Chern-Simons theory and knot invariants. In *Chern-Simons Theory, Matrix Models, and Topological Strings*. Oxford University Press, 09 2005.

[Moo12] G. Moore. A very long lecture on the physical approach to donaldson and seiberg-witten invariants of four-manifolds. 2012.

[Mot85] E. Mottola. Particle Creation in de Sitter Space. *Phys. Rev. D*, 31:754, 1985.

[MS59] Paul C. Martin and Julian Schwinger. Theory of many-particle systems. i. *Physical Review*, 115:1342–1373, 1959.

[Nab10] G.L. Naber. *Topology, Geometry and Gauge fields: Foundations*. Texts in Applied Mathematics. Springer New York, 2010.

[Nak90] Mikio Nakahara. *Geometry, topology and physics*. Graduate student series in physics. Hilger, Bristol, 1990.

[O'N83] Barrett O'Neill. *Semi-Riemannian Geometry With Applications to Relativity, 103, Volume 103 (Pure and Applied Mathematics)*. Academic Press, 1983.

[Paj20] Enrico Pajer. *Field Theory in Cosmology*. Part III, 2020.

[Pen20] Geoffrey Penington. Entanglement Wedge Reconstruction and the Information Paradox. *JHEP*, 09:002, 2020.

[Poi04] Eric Poisson. *A Relativist's Toolkit: The Mathematics of Black-Hole Mechanics*. Cambridge University Press, 2004.

[PR84] Roger Penrose and Wolfgang Rindler. *Spinors and Space-Time*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1984.

[PS95] M.E. Peskin and D.V. Schroeder. *An Introduction To Quantum Field Theory*. Frontiers in Physics. Avalon Publishing, 1995.

[PT09] Leonard E. Parker and D. Toms. *Quantum Field Theory in Curved Spacetime: Quantized Field and Gravity*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 8 2009.

[Rob35] H. P. Robertson. Kinematics and World-Structure. *Astrophysical Journal*, 82:284, November 1935.

[Rob36a] H. P. Robertson. Kinematics and World-Structure II. *Astrophysical Journal*, 83:187, April 1936.

[Rob36b] H. P. Robertson. Kinematics and World-Structure III. *Astrophysical Journal*, 83:257, May 1936.

[Sre07] M. Srednicki. *Quantum Field Theory*. Cambridge University Press, 2007.

[SS76] Christiane Schomblond and Philippe Spindel. Conditions d'unicité pour le propagateur $\delta^1(x, y)$ du champ scalaire dans l'univers de de Sitter. *Annales de l'institut Henri Poincaré. Section A, Physique Théorique*, 25(1):67–78, 1976.

[SSV01] Marcus Spradlin, Andrew Strominger, and Anastasia Volovich. Les Houches lectures on de Sitter space. In *Les Houches Summer School: Session 76: Euro Summer School on Unity of Fundamental Physics: Gravity, Gauge Theory and Strings*, pages 423–453, 10 2001.

[STY95] Misao Sasaki, Takahiro Tanaka, and Kazuhiro Yamamoto. Euclidean vacuum mode functions for a scalar field on open de Sitter space. *Phys. Rev. D*, 51:2979–2995, 1995.

[SV96] Andrew Strominger and Cumrun Vafa. Microscopic origin of the Bekenstein-Hawking entropy. *Phys. Lett.*, B379:99–104, 1996.

[Unr76] W. G. Unruh. Notes on black hole evaporation. *Phys. Rev. D*, 14:870, 1976.

[VP99] Antoine Van Proeyen. Tools for supersymmetry. *Ann. U. Craiova Phys.*, 9(I):1–48, 1999.

[Wal37] A. G. Walker. On milne's theory of world-structure. *Proceedings of the London Mathematical Society*, s2-42:90–127, 1937.

[Wal84] Robert M. Wald. *General Relativity*. The University od Chicago Press, 1984.

[Wei95] Steven Weinberg. *The Quantum Theory of Fields*, volume 1-3. Cambridge University Press, 1995.

[Zam86] A. B. Zamolodchikov. Irreversibility of the Flux of the Renormalization Group in a 2D Field Theory. *JETP Lett.*, 43:730–732, 1986. [Pisma Zh. Eksp. Teor. Fiz.43,565(1986)].