University of Cape Town

Department of Mathematics and Applied Mathematics

Dissertation presented for the degree of
Master of Science

Entanglement entropy, the Ryu-Takayanagi prescription, and conformal maps

Author:
Alastair Grant-Stuart

Supervisors:
Prof. Jeff Murugan
Dr. Jonathan Shock

March 2017

Abstract

We define and explore the concepts underpinning the Ryu-Takayanagi prescription for entanglement entropy in a holographic theory. We begin by constructing entanglement entropy in finite-dimensional quantum systems, and defining the boundary at infinity of a bulk spacetime. This is sufficient for a naïve application of the Ryu-Takayanagi prescription to some simple examples; nonetheless, we review the general theory of minimal submanifolds in Riemannian ambient manifolds in order to better characterise the objects involved in the prescription. Finally, we explore the symmetries of the the boundary theory to which the prescription applies, and thereby extend the aforementioned examples. Throughout, emphasis is placed on making explicit the mathematical structures that are taken for granted in the research literature.

Acknowledgements

I would like to thank Prof. Jeff Murugan and Dr. Jonathan Shock for introducing me to this field of research, and for their support, guidance and abundant patience. Additionally, I would like to thank S. Shajidul Haque for his assistance and discussions at the outset of this work.

For funding this project, I am grateful to the National Institute for Theoretical Physics.

For caring, I thank Emma Gibson.

Plagiarism declaration

I know the meaning of plagiarism and declare that all of the work in the dissertation, save for that which is properly acknowledged, is my own.

Contents

Chapter 1

Introduction

A holographic duality is a conjectured equivalence between a string theory in one spacetime, and a quantum field theory (QFT) on the boundary at infinity of that spacetime. The spacetime on which the string theory lives is often called the 'bulk', while its boundary at infinity is often simply called the 'boundary'. Such an equivalence between bulk and boundary theories bears the label of 'holography' since it implies that the lower-dimensional boundary QFT contains all of the same physical information as exists in the higher-dimensional bulk string theory.

There is a general class of models that relate a string theory on a spacetime of the form $\mathrm{AdS}_{d+2} \times \mathcal{M}$, where AdS_n denotes an n-dimensional anti-de Sitter spacetime and $\mathcal{M}$ is a compact Riemannian manifold, to a conformal field theory (CFT) on the $(d+1)$-dimensional boundary at infinity of the AdS_{d+2} component of the spacetime. This class of models and the assertion that they demonstrate holography is collectively termed the AdS/CFT conjecture.

The archetypal AdS/CFT model is the conjectured equivalence between a Type IIB closed string theory in an $\mathrm{AdS}_5 \times \mathbb{S}^5$ bulk spacetime, and a superconformal $\mathrm{SU}(N)$ Yang-Mills gauge theory on the boundary (Gubser, Klebanov, and Polyakov 1998; Maldacena 1999; Witten 1998). Here $\mathbb{S}^m$ denotes an m-dimensional sphere. This example is of special interest since it relates a string theory, and therefore a theory of quantum gravity, to a field theory that possesses many similarities to quantum chromodynamics. With such a duality, holographic techniques can provide access to theoretical predictions in the Yang-Mills theory when it is strongly coupled, and therefore intractable to the usual perturbative techniques of QFT.

In the low-energy limit, the string theory in the bulk can be replaced by a low-energy effective field theory. Since the string theories of interest are supersymmetric, so too are their low-energy effective field theories. Therefore, since the effective field theories possess gravitational degrees of freedom amongst others, they are often called supergravity theories.

1.1 Holographic entanglement entropy and the Ryu-Takayanagi conjecture

Recently, there has been much interest in the properties of quantum entanglement in such holographic theories. This has been spurred greatly by the proposal of Ryu and Takayanagi (2006a,b) of a prescription to calculate entanglement entropies of spatial regions in a $(d+1)$-dimensional boundary CFT using minimal surfaces in a spatial section of the associated bulk AdS_{d+2}.

The Ryu-Takayanagi prescription is as follows. Given a region A in a constant-time slice of the boundary CFT, label the boundary of A as ∂A. Let γ_A be the surface in the corresponding constant-time slice of the bulk spacetime, with the minimal area such that the boundary of γ_A is identical to the boundary of A. Then the holographic entanglement entropy $S(A)$ of region A is given by

$$S(A) = \frac{\mathrm{Area}(\gamma_A)}{4G_N^{(d+2)}}, \tag{1.1}$$

where $G_N^{(d+2)}$ is the $(d+2)$-dimensional Newton gravitational constant.

In fact, the holographic entanglement entropy $S(A)$ is divergent, since the area of a surface that extends towards the boundary at infinity of a manifold is necessarily infinite. This is resolved by imposing an artificial cut-off on how far towards the boundary the surface γ_A extends; this cut-off plays the same role as a UV cut-off in the conformal field theory on the boundary.

Since the original proposal of the Ryu-Takayanagi formula for holographic entanglement entropy, there have been many developments. Progress has been made towards a robust derivation of the Ryu-Takayanagi formula (Casini, M Huerta, and Myers 2011). A variety of explicit calculations, both analytic and numerical, of holographic entanglement entropy have been done for particular examples of regions A, revealing information about the dependence of entanglement entropy $S(A)$ on the particular shape of A (Allais and Mezei 2015; Carmi 2015; Fonda, Giomi, et al. 2015; Fonda, Seminara, and Tonni 2015; Krtouš and Zelnikov 2014). The holographic quantity $S(A)$ has been proven to obey many characteristic properties of quantum entanglement entropy, such as strong subadditivity, monogamy of mutual information and others (Hayden, Headrick, and Maloney 2013; Headrick 2014; Headrick and Takayanagi 2007). The Ryu-Takayanagi formula assumes restriction to a constant-time, and therefore Riemannian, submanifold of the bulk spacetime; a covariant generalisation of the Ryu-Takayanagi formula that requires no such restriction has been proposed and studied (Headrick, Hubeny, et al. 2014; Hubeny, Rangamani, and Takayanagi 2007). The behaviour of the holographic entanglement entropy under renormalisation group flow has been studied (Klebanov, Nishioka, et al. 2012; Liu and Mezei 2013). Additionally, a reformulation of the Ryu-Takayanagi prescription in terms of maximised flows,

or vector fields with pointwise bounded norm, rather than minimal surfaces has been proposed, giving an interpretation of holographic entanglement entropy in terms of bit threads (Freedman and Headrick 2016)

1.2 Aims and outline

In this work, we aim to lay out explicitly the mathematical structures and definitions required to make the statement of the Ryu-Takayanagi formula (1.1) as precise as possible, as a first step towards a concrete understanding of it. As always, in order to make sense of the concepts encountered, we present some simple examples along the way. We assume only introductory knowledge of quantum mechanics, quantum field theory and differential geometry on the part of the reader. We do this not only in the hope that this work may be as helpful as possible to any hypothetical future student looking to learn about holographic entanglement entropy, but also because it makes explicit many subtleties in the picture that are only tacitly assumed, or even neglected, in the research literature.

Before looking any further at holography and the Ryu-Takayanagi prescription, we first rigorously define entanglement in finite-dimensional quantum mechanics in Chapter 2. We also construct the notion of entropy as an information-theoretic property of a probability distribution; we do so in order to highlight its generality beyond its usual context of classical statistical physics. Combining entanglement and entropy, we obtain a definition of the entanglement entropy of a subsystem in a quantum mechanical system.

Another necessary prior construction is needed. The AdS/CFT conjecture in general, and the Ryu-Takayanagi prescription in particular, rely heavily on the geometric notion of the 'boundary at infinity' of a spacetime. However, this is not a boundary in the traditional, topological sense. In Chapter 3, we present the appropriate mathematics of conformal maps on Riemannian and pseudo-Riemannian manifolds necessary to understand the boundary at infinity. We also apply these tools to find the boundaries at infinity of some common manifolds, including the anti-de Sitter spacetime.

Understanding only the geometry of AdS and its boundary at infinity is sufficient to apply the Ryu-Takayanagi prescription in some simple examples, as we demonstrate in Chapter 4. These examples admit easy descriptions in particular coordinate systems, so that it is possible to use naïve techniques of variational calculus to find the area-minimising submanifold γ_A required to compute the holographic entanglement entropy.

However, such techniques are limited both in their application to particular examples, since they are heavily dependent on the minimal surfaces admitting 'nice' coordinate descriptions, and in their 'black box' nature that does not engage with or expose the underlying structure of area-minimising submanifolds. In Chapter 5, we present a more general mathematical descrip-

tion of area-minimising submanifolds . We content ourselves with defining the relevant geometric quantities necessary to characterise an area-minimising submanifold, and connecting them in their abstract form to their more computationally applicable reformulations. We emphasise coordinate-free constructions throughout the chapter. However, a full mathematical exposition of the theory of area-minimising submanifolds immersed in Riemannian or pseudo-Riemannian manifolds is beyond the scope of this work.

As important as understanding the mathematical definitions underlying holography and the Ryu-Takayanagi prescription, is understanding the physical symmetries of the theory. The quantum field theories on the boundary in the AdS/CFT possess conformal symmetry. In Chapter 6, after conformal symmetry transformations are defined in the general context, the case of a particular constant-time slice of the AdS_4 boundary at infinity is presented in detail. The family of conformal symmetry transformations on this manifold is elucidated in full via its connection to complex analysis; these conformal symmetries are then applied to infer holographic entanglement entropies of new regions from the calculations done in Chapter 4 on the basis of the Ryu-Takayanagi conjecture.

Chapter 2

Entanglement, information theory and entropy

Before we can explore the geometric notions involved in holographic entanglement entropy and the Ryu-Takayanagi prescription, we must understand the concept of entanglement entropy in a quantum system. To this end, we here provide a reasonably self-contained description of entanglement in finite-dimensional quantum systems, and define entropy from first principles in an information-theoretic context, in preparation for its application to describing entanglement. Subsequently, we introduce the density operator formalism of quantum mechanics, and using it combine the notions entanglement and entropy to define entanglement entropy, and note some of its simplest and most immediate implications.

2.1 Entanglement

In quantum mechanics, the state of a system can be represented by a (normalised) vector in a Hilbert space $\mathcal{H}$ over field $\mathbb{K}$, where the field is almost invariably the complex numbers, $\mathbb{K} = \mathbb{C}$. Suppose we have two separate subsystems A and B making up the total system; the states of these subsystems lie in Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. There are a number of ways to combine $\mathcal{H}_A$ and $\mathcal{H}_B$ mathematically to obtain a new Hilbert space $\mathcal{H}$ to describe the total system; two that have cogent physical interpretations are the tensor product, and the direct sum.

2.1.1 Tensor products and direct sums of Hilbert spaces

We denote the tensor product of $\mathcal{H}_A$ and $\mathcal{H}_B$ as $\mathcal{H}_A \otimes \mathcal{H}_B$, and the direct sum as $\mathcal{H}_A \oplus \mathcal{H}_B$.

As sets, both $\mathcal{H}_A \otimes \mathcal{H}_B$ and $\mathcal{H}_A \oplus \mathcal{H}_B$ consist of ordered 2-tuples of vectors in $\mathcal{H}_A$ and $\mathcal{H}_B$, although typical notation differs:

$$\mathcal{H}_A \otimes \mathcal{H}_B := \{a \otimes b | a \in \mathcal{H}_A, b \in \mathcal{H}_B\},$$
$$\mathcal{H}_A \oplus \mathcal{H}_B := \{(a, b) | a \in \mathcal{H}_A, b \in \mathcal{H}_B\}.$$

The two differ in their structure as Hilbert spaces over $\mathbb{C}$, i.e. in the definitions of addition and scalar multiplication of vectors, and in the definition of the inner product. We present only the key practical differences relevant for physics in Table 2.1 here. See, for instance, Reed and Simon (1981) for a more complete mathematical treatment.

At a heuristic level, a combination of two physical systems A and B will be described using a direct sum $\mathcal{H}_A \oplus \mathcal{H}_B$ if a state of the combined system may be in A 'or' B (Baez and J Huerta 2010). In particular, a state that is in A only would be $(a, 0)$ for some $a \in \mathcal{H}_A$. By the definition of the inner product $\langle ., . \rangle_{\mathcal{H}_A \oplus \mathcal{H}_B}$, this is automatically orthogonal to a state $(0, b)$ in B only.

On the other hand, the tensor product space $\mathcal{H}_A \otimes \mathcal{H}_B$ is used when the state is necessarily in A 'and' B simultaneously. After all, if we attempt to construct a state in $\mathcal{H}_A \otimes \mathcal{H}_B$ with no overlap with any states involving $\mathcal{H}_B$, for instance, we can get only the zero vector: for any $a \in \mathcal{H}_A$, $a \otimes 0 = 0$.

The Fock spaces of quantum field theory are effective examples to illustrate the differences between tensor product and direct sum spaces. Say the state of a given single particle is an element of Hilbert space $\mathcal{H}$. Then a state of two such particles lives in the tensor product space $S_{\pm}(\mathcal{H} \otimes \mathcal{H})$, where $S_{\pm}$ projects out the symmetric $(+)$ or anti-symmetric $(-)$ subspaces, for bosons or fermions respectively. A two-particle state must be described in 'both' copies of $\mathcal{H}$ simultaneously, so that the tensor product construction is necessary.

This may be extended to a general n-particle state, which lives in

$$S_{\pm}(\mathcal{H}^{\otimes n}) = S_{\pm}(\underbrace{\mathcal{H} \otimes \mathcal{H} \otimes \ldots \otimes \mathcal{H}}_{n \text{ copies}}).$$

The Fock space is then the (metric space completion of the) direct sum of all n-particle spaces for $n \in \mathbb{N}$:

$$F_{\pm}(\mathcal{H}) := \overline{\bigoplus_{n=1}^{\infty} S_{\pm} H^{\otimes n}}$$

A state in the Fock space may describe specifically m particles for some $m \in \mathbb{N}$, and thus live in only $S_{\pm}(\mathcal{H}^{\otimes m})$. It will have zero overlap with any n-particle state for any $n \neq m$. Thus the use of a direct sum construction.

	$\mathcal{H}_A \otimes \mathcal{H}_B$	$\mathcal{H}_A \oplus \mathcal{H}_B$
Addition	bilinear: $a_1 \otimes b + a_2 \otimes b = (a_1 + a_2) \otimes b$, $a \otimes b_1 + a \otimes b_2 = a \otimes (b_1 + b_2)$	componentwise: $(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$
Scalar multiplication	$\lambda(a \otimes b) = (\lambda a) \otimes b = a \otimes (\lambda b)$	$\lambda(a, b) = (\lambda a, \lambda b)$
Inner product	$\langle a_1 \otimes b_1, a_2 \otimes b_2 \rangle_{\mathcal{H}_A \otimes \mathcal{H}_B}$ $= \langle a_1, a_2 \rangle_{\mathcal{H}_A} \cdot \langle b_1, b_2 \rangle_{\mathcal{H}_B}$	$\langle (a_1, b_1), (a_2, b_2) \rangle_{\mathcal{H}_A \oplus \mathcal{H}_B}$ $= \langle a_1, a_2 \rangle_{\mathcal{H}_A} + \langle b_1, b_2 \rangle_{\mathcal{H}_B}$

Table 2.1: Comparison of operations in a tensor product and direct sum of Hilbert spaces. Here, $a, a_1, a_2 \in \mathcal{H}_A$ and $b, b_1, b_2 \in \mathcal{H}_B$ and $\lambda \in \mathbb{C}$.

2.1.2 Entangled states

Entanglement occurs in tensor product spaces $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, where a state must necessarily live simultaneously in $\mathcal{H}_A$ and $\mathcal{H}_B$.

For convenience, we will work with finite-dimensional spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. Much of what we present can be extended rigorously to infinite-dimensional but separable $\mathcal{H}_A$ and $\mathcal{H}_B$ (i.e. Hilbert spaces for which countable bases exist). The generalisation to non-separable spaces is less rigorous. For the remainder of this chapter, we use the bra-ket notation.

Say we have orthonormal bases $\{|a\rangle_A\}_{a \in \{1,\dots,N_A\}}$ of $\mathcal{H}_A$ and $\{|b\rangle_B\}_{b \in \{1,\dots,N_B\}}$ of $\mathcal{H}_B$, where $N_A = \dim \mathcal{H}_A$ and $N_B = \dim \mathcal{H}_B$. A general normalised vector $|\psi\rangle \in \mathcal{H}$ can therefore be written as

$$|\psi\rangle = \sum_{a=1}^{N_A} \sum_{b=1}^{N_B} C_{ab} \, |a\rangle_A \otimes |b\rangle_B \,, \tag{2.1}$$

for some coefficients $C_{ab} \in \mathbb{C}$. That $|\psi\rangle$ is normalised means that $\langle \psi | \psi \rangle = 1$; since the basis vectors $|a\rangle_A$ and $|b\rangle_B$ are also normalised, we must have

$$\sum_{a=1}^{N_A} \sum_{b=1}^{N_B} |C_{ab}|^2 = 1. \tag{2.2}$$

Definition 2.1 (Entangled states)**.** The state $|\psi\rangle$ defined above is *not entangled* if there exist $d_a^A, d_b^B \in \mathbb{C}$ such that $d_a^A d_b^B = C_{ab}$ for all $a \in \{1, \dots, N_A\}$ and $b \in \{1, \dots, N_B\}$. The state $|\psi\rangle$ is *entangled* if no such d_a^A, d_b^B exist.

A state that is not entangled may be written as

$$|\psi\rangle = \sum_{a=1}^{N_A} \sum_{b=1}^{N_B} d_a^A d_b^B \, |a\rangle_A \otimes |b\rangle_B = |\psi\rangle_A \otimes |\psi\rangle_B \,,$$

where

$$|\psi\rangle_A = \sum_{a=1}^{N_A} d_a^A \, |a\rangle_A \qquad \text{and} \qquad |\psi\rangle_B = \sum_{b=1}^{N_B} d_b^B \, |b\rangle_B \,.$$

An entangled state can never be written as a single term $|\psi\rangle_A \otimes |\psi\rangle_B$, but only as a linear combination of such terms.

Although this consequence of Definition 2.1 is clear enough, it is not yet clear how to practically determine whether the decomposition of C_{ab} into d_a^A and d_b^B exists. Therefore, we present the foundation of an algorithmic approach to the problem of determining whether or not $|\psi\rangle$ is entangled, commonly found in textbooks such as Nielsen and Chuang (2010).

Lemma 2.1 (Schmidt decomposition)**.** *For any normalised state vector $|\psi\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_B$, there exist Schmidt coefficients $\sigma_i \in [0,1]$, and Schmidt vectors $|\alpha_i\rangle_A \in \mathcal{H}_A$, $|\beta_i\rangle_B \in \mathcal{H}_B$ for $i \in \{1,\ldots,N\}$, $N = \min(N_A, N_B)$ where:*

(i) $|\psi\rangle = \sum_{i=1}^{N} \sigma_i |\alpha_i\rangle_A \otimes |\beta_i\rangle_B$.

(ii) $\{|\alpha_i\rangle_A\}_{i\in\{1,\ldots,N\}}$ *is orthonormal in* $\mathcal{H}_A$.

(iii) $\{|\beta_i\rangle_B\}_{i\in\{1,\ldots,N\}}$ *is orthonormal in* $\mathcal{H}_B$.

(iv) $\sum_{i=1}^{N} \sigma_i^2 = 1$.

Proof. Expand $|\psi\rangle$ in orthonormal bases of $\mathcal{H}_A$ and $\mathcal{H}_B$ as per (2.1), so that normalisation of $|\psi\rangle$ gives condition (2.2).

The coefficients C_{ab} of the expansion form a complex $N_A \times N_B$ matrix. We may always write this matrix in terms of its singular value decomposition:

$$C_{ab} = \sum_{c=1}^{N_A} \sum_{d=1}^{N_B} U_{ac} \Sigma_{cd} V_{db}^{\dagger},$$

or $C = U\Sigma V^{\dagger}$, where U is a $N_A \times N_A$ unitary matrix, V is a $N_B \times N_B$ unitary matrix and Σ is a $N_A \times N_B$ diagonal matrix whose entries are real and non-negative. Let $\sigma_i = \Sigma_{ii}$ be the diagonal elements of Σ; the index i takes values from 1 to $N = \min(N_A, N_B)$.

Using this,

$$\begin{aligned} |\psi\rangle &= \sum_{a,b,c,d} \left(U_{ac} \Sigma_{cd} V_{db}^{\dagger} \right) |a\rangle_A \otimes |b\rangle_B \\ &= \sum_{i=1}^{N} \sigma_i \left(\sum_{a=1}^{N_A} U_{ai} |a\rangle_A \right) \otimes \left(\sum_{b=1}^{N_B} V_{ib}^{\dagger} |b\rangle_B \right) \end{aligned}$$

By letting

$$|\alpha_i\rangle_A = \sum_{a=1}^{N_A} U_{ai} |a\rangle_A \qquad \text{and} \qquad |\beta_i\rangle_B = \sum_{b=1}^{N_B} V_{ib}^{\dagger} |b\rangle_B ,$$

we obtain Item (i).

By the unitarity of U and orthonormality of the vectors $|a\rangle_A$,

$$\langle \alpha_i | \alpha_j \rangle = \sum_{a_1,a_2} \langle a_1 | U^*_{a_1 i} U_{a_2 j} | a_2 \rangle = \sum_{a_1,a_2} U^\dagger_{i a_1} U_{a_2 j} \langle a_1 | a_2 \rangle = \sum_{a_1} U^\dagger_{i a_1} U_{a_1 j} = \delta_{ij},$$

where δ_{ij} is the Kronecker delta symbol. This demonstrates Item (ii); Item (iii) holds similarly.

Recall that (2.2) holds since $|\psi\rangle$ is taken to be normalised; this condition can be rewritten as $\mathrm{tr} CC^\dagger = 1$. Using the singular value decomposition,

$$\mathrm{tr} CC^\dagger = \mathrm{tr}\left(U\Sigma V^\dagger\right)\left(V\Sigma^\dagger U^\dagger\right) = \mathrm{tr}\left(\Sigma\Sigma^\dagger\right) = \sum_{i=1}^{N} \sigma_i^2,$$

using the cyclicity of the trace, the unitarity of U and V, and that Σ is diagonal with real elements σ_i. Therefore, (2.2) gives Item (iv). □

Once a vector $|\psi\rangle$ has been written in the Schmidt decomposition, we can immediately see whether the state is entangled or not.

Corollary 2.2. *A state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is not entangled if and only if it has exactly one non-zero Schmidt coefficient. In this case, that Schmidt coefficient has value 1. For any other set of Schmidt coefficients, the state is entangled.*

Since the Schmidt decomposition is essentially a singular value decomposition, which can be performed algorithmically, this gives a definitive way to determine whether or not any given state in a tensor product of two (finite-dimensional) Hilbert spaces is entangled.

2.2 Entropy

We now introduce the fundamental concept of entropy. Although the use of entropy originates in thermodynamics, and is made more substantial in statistical mechanics, perhaps the most informative view of entropy arises from pure probability and information theory.

Given a random variable X which takes values in a finite space of possible outcomes $\Omega = \{x_1, \ldots, x_n\}$, we can construct a distribution $p_X(x_i) = \mathbb{P}(X = x_i)$ of the probabilities of each outcome. As a probability (mass) distribution p_X satisfies $p_X(x) \in [0,1]$ for all $x \in \Omega$ and $\sum_{x\in\Omega} p_X(x) = 1$.

The probability distribution p_X can be seen as a model of our lack of information about the true value that X will take upon sampling or observing it. To illustrate: the trivial example is that of a probability distribution which is zero for all but one possible outcome x_j. Then we have $p_X(x_i) = \delta_{ij}$. In this case, it is known with certainty that the value X will take is precisely x_j. Therefore, the distribution p_X actually describes no lack of information at all.

The next-less trivial example is that of a uniform probability distribution, $p_X(x) = \frac{1}{n}$ for all $x \in \Omega$. In this case, the distribution indicates that all we know about X is the space of possible values it may take – we have no more knowledge about which x_i is more or less likely than any other. We can thus regard the uniform distribution as describing a maximum possible 'lack of information' about X, for a given set of known possible outcomes.

The entropy $S(p_X)$ is a property of a distribution p_X that quantifies this lack-of-information described by p_X, or equivalently the amount of information to gained by actually measuring X.

Definition 2.2. Given a random variable X with finite set of possible values Ω and probability distribution $p_X(x) := \mathbb{P}(X = x)$ for $x \in \Omega$, the *entropy* S of the distribution p_X is the real-valued quantity satisfying the following conditions (Jaynes 1957):

(i) S is a continuous function of the probabilities $p_X(x)$ for $x \in \Omega$ only, $S = S(\{p_X(x) : x \in \Omega\})$. For brevity, we denote this dependence as $S(p_X)$.

(ii) If p_X is uniform with $|\Omega| = n$, so that $p_X(x) = 1/n$ for all $x \in \Omega$, then $T(n) := S(p_X) = S(\{1/n, \ldots, 1/n\})$ is a monotonically increasing function of n

(iii) Let $\mathcal{I}$ denote a partition of Ω, i.e. let $\mathcal{I}$ be a family of disjoint non-empty sets $I \in \mathcal{I}$ so that $\Omega = \cup_{I \in \mathcal{I}} I$. Let $p_{\mathcal{I}}(I) := \mathbb{P}(X \in I) = \sum_{x \in I} p_X(x)$, and $p_{X|I}(x) = \mathbb{P}(X = x | X \in I)$. Then

$$S(p_X) = S(p_{\mathcal{I}}) + \sum_{I \in \mathcal{I}} p_{\mathcal{I}}(I) S(p_{X|I}) \tag{2.3}$$

Let us unpack this definition. Item (i) is reasonably clear: S depends on the probability distribution of interest and nothing else, and that dependence is continuous in those probabilities.

Item (ii) requires that, if we regard specifically uniform distributions, then the value of S increases with the number of possible outcomes. Under the interpretation of $S(p_X)$ as a description of the amount of lack-of-information represented by the probability distribution p_X, this means that – assuming equal probabilities – the more possible outcomes there are, the less we know.

Item (iii) is perhaps the least obvious. It allows one to relate the entropy of an underlying distribution to that of a 'coarse-grained' description of the outcomes. Consider dividing the possible outcomes into sets (as per the partition $\mathcal{I}$). Then an observation of outcome $X = x$ can be rephrased as first the observation that X lies in a particular set $I \in \mathcal{I}$ and subsequently that X takes the particular value x in I. The information to be gained from the

observation (and thus the entropy) should not depend on how the outcomes are so coarse-grained.

The 'information gain' in the course-grained case can be broken down into two parts. First is the information that $X \in I$ for some particular I; this is quantified by $S(p_{\mathcal{I}})$. Then is the information that, given that X is in a particular I, X takes a specific value $x \in I$; this is quantified by the entropy of the conditional distribution $S(p_{X|I})$. However, all possible I's in the coarse-graining must be accounted for; this is done by summing the entropies $S(p_{X|I})$ weighted with the probabilities $p_{\mathcal{I}}(I)$ that the outcome lies in each possible I.

Note that other axioms can also be taken to define entropy (Cover and Thomas 2006; Shannon 1948); however, all lead to the following proposition.

Proposition 2.3. *The entropy of a finite probability distribution p_X is*

$$S(p_X) = -k \sum_{x \in X} p_X(x) \log p_X(x),$$

where k may be any positive number.

The constant k here sets the units in which entropy is measured; conventional choices for k vary between different fields of study. In information theory and communication, for instance, usually $k = 1/\log 2$ so that $S(p_X) = -\sum p_X(x) \log_2 p_X(x)$; then entropy is measured in bits. In physics, k is chosen as the Boltzmann constant k_B; then the expression above gives the Gibbs entropy, measured in units of energy over temperature, e.g. $J \cdot K^{-1}$.

To prove Proposition 2.3, we will first need a technical lemma. We use $\mathbb{N}$ and $\mathbb{N}_0$ to denote the positive and non-negative integers respectively.

Lemma 2.4. *If $f : \mathbb{N} \to \mathbb{R}$ is a monotonically increasing function satisfying $f(mn) = f(m) + f(n)$, then $f(n) = k \log n$ for some $k > 0$.*

Proof. First note that $f(n^p) = pf(n)$ for any $n \in \mathbb{N}$ and $p \in \mathbb{N}_0$. This can be shown inductively: let $m = 1$ in $f(mn) = f(m) + f(n)$. Then $f(1 \cdot n) = f(1) + f(n)$, or $f(1) = 0$. This shows that $f(n^p) = pf(n)$ for $p = 0$. Next, assume $f(n^q) = qf(n)$ for some $q \in \mathbb{N}_0$. Then

$$f(n^{q+1}) = f(n \cdot n^q) = f(n) + f(n^q) = f(n) + qf(n) = (q+1)f(n).$$

Since $f(1) = 0$ and f is monotonically increasing, $f(n) > 0$ for any integer $n > 1$.

Now, let $m, n \in \mathbb{N}$ with $n, m > 1$, and take any $p \in \mathbb{N}$. Then there exists some $q \in \mathbb{N}_0$ such that

$$n^q \leq m^p < n^{q+1}.$$

Since f is monotonically increasing, this means that

$$f(n^q) \leq f(m^p) \leq f(n^{q+1}).$$

Using the properties demonstrated above, this can be rearranged into

$$0 \leq \frac{f(m)}{f(n)} - \frac{q}{p} \leq \frac{1}{p}.$$

The logarithm is also monotonically increasing, satisfies $\log n^p = p \log n$ and $\log n > 0$ for $n > 1$, so similarly

$$0 \leq \frac{\log m}{\log n} - \frac{q}{p} \leq \frac{1}{p}.$$

Combining these, we have

$$-\frac{1}{p} \leq \left(\frac{f(m)}{f(n)} - \frac{q}{p}\right) - \left(\frac{\log m}{\log n} - \frac{q}{p}\right) \leq \frac{1}{p},$$

or

$$\left|\frac{f(m)}{f(n)} - \frac{\log m}{\log n}\right| \leq \frac{1}{p}.$$

Since this holds for any $p \in \mathbb{N}$, we must have

$$\frac{f(m)}{f(n)} = \frac{\log m}{\log n}, \qquad \text{i.e.} \qquad \frac{f(m)}{\log m} = \frac{f(n)}{\log n}.$$

This holds for any integers n and m greater than 1, so $f(n) = k \log n$ for some constant k whenever $n > 1$. When $n = 1$, $f(1) = 0 = k \log 1$ as well. For f to be monotonically increasing, necessarily $k > 0$.

□

With this, we can return to entropy:

Proof of Proposition 2.3. Since $S(p_X)$ depends continuously on the probabilities $p_X(x)$, it will be sufficient to prove the result for rational probabilities $p_X(x)$ – any real-valued probability can approached as a limit of a sequence of rational probabilities.

Therefore, we consider probabilities

$$p_X(x) = \frac{n_x}{\sum\limits_{x' \in X} n_{x'}},$$

where $n_x \in \mathbb{N}$ for all $x \in X$.

Using Item (iii) of Definition 2.2, we can write any $S(p_X)$ for such rational probabilities in terms of entropies $T(n)$ of finite uniform distributions. Consider a variable Y distributed uniformly over $\sum_{x \in X} n_x$ possible outcomes y, so $p_Y(y) = 1/\left(\sum_{x \in X} n_x\right)$. Let $\mathcal{I}$ be a partition of these outcomes into sets I_x for $x \in X$, with n_x outcomes in I_x. Then the probability that Y is in I_x is $p_{\mathcal{I}}(I_x) = n_x/\left(\sum_{x \in X} n_x\right) = p_X(x)$. Bayes' theorem gives that

the conditional probability of Y taking value y, given that Y is in I_x, is $p_{Y|I_x}(y) := \mathbb{P}(Y = y|Y \in I_x) = p_Y(y)/p_{\mathcal{I}}(I_x)$ for any $y \in I_x$, so here $p_{Y|I_x}(y) = 1/n_x$.

Using this, (2.3) gives

$$T\left(\sum_{x' \in X} n_{x'}\right) = S(p_{\mathcal{I}}) + \sum_{I_x \in \mathcal{I}} p_{\mathcal{I}}(I_x) T(n_x),$$

or since $p_{\mathcal{I}}(I_x) = p_X(x)$,

$$T\left(\sum_{x' \in X} n_{x'}\right) = S(p_X) + \sum_{x \in X} p_X(x) T(n_x) \tag{2.4}$$

Now consider the case where $n_x = n$ for all x, so that p_X is uniform. With $|X| = m$ so $p_X(x) = 1/m$, (2.4) becomes

$$T(mn) = T(m) + T(n) \sum_{x \in X} p_X(x) = T(m) + T(n).$$

This holds for any $m, n \in \mathbb{N}$. Thus from Lemma 2.4, $T(n) = k \log n$ for some $k > 0$.

Inserting the form of T into (2.4), we have

$$\begin{aligned} S(p_X) &= k \left\{ \log\left(\sum_{x' \in X} n_{x'}\right) - \sum_{x \in X} p_X(x) \log n_x \right\} \\ &= k \sum_{x \in X} p_X(x) \left\{ \log\left(\sum_{x' \in X} n_{x'}\right) - \log n_x \right\} \\ &= -k \sum_{x \in X} p_X(x) \log\left(\frac{n_x}{\sum n_{x'}}\right) \\ &= -k \sum_{x \in X} p_X(x) \log p_X(x). \end{aligned}$$

□

2.3 Entanglement entropy

We are now equipped with the separate notions of entanglement and entropy. We wish to combine them to form a notion of the entropy of an entangled state in a quantum system, called the entanglement entropy.

To do so, we must first associate a probability distribution to a given state. In quantum mechanics, a normalised state vector $|\psi\rangle$ can be expanded in the eigenbasis $|e_i\rangle$ of some observable $\hat{\mathcal{O}}$:

$$|\psi\rangle = \sum_i a_i |e_i\rangle, \qquad \text{where} \qquad \hat{\mathcal{O}} |e_i\rangle = \lambda_i |e_i\rangle.$$

Then upon taking a single measurement of observable $\hat{\mathcal{O}}$, the probability of measuring the eigenvalue λ_i is given by $|a_i|^2$.

In such a way, we are accustomed to associating probability distributions with quantum mechanical states, and may try to describe the entropy of such a distribution. However, such a distribution is a property of not only the state $|\psi\rangle$, but also the basis in which it is expanded, and therefore also the observable $\hat{\mathcal{O}}$.

We are looking explicitly for an entropy that is purely a property of the state of the system, not of any choice of basis. Therefore, the distribution above is *not* the one to use to define it.

To properly introduce a notion of entanglement entropy into a quantum system, we need to introduce a new formalism that will enable a sensible definition of entropy (in general, not specifically of entanglement) in the quantum-mechanical framework.

2.3.1 Density operator formalism

The usual formulation of quantum mechanics has states of the system represented as normalised state vectors (or wavefunctions) in a Hilbert space $\mathcal{H}$. With a state vector $|\psi\rangle \in \mathcal{H}$ one can:

- propagate the state information through time using a Hamiltonian operator $\hat{H}$:
$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\psi\rangle = \hat{H} |\psi\rangle .$$
- calculate the expected value of an observable $\hat{\mathcal{O}}$:
$$\langle \hat{\mathcal{O}} \rangle = \langle \psi | \hat{\mathcal{O}} | \psi \rangle .$$
- calculate the probability of the state being in a subspace Ω of $\mathcal{H}$:
$$\mathbb{P}(\Omega) = \langle \psi | \hat{P}_\Omega | \psi \rangle ,$$
where $\hat{P}_\Omega : \mathcal{H} \to \mathcal{H}$ is the projection operator ($\hat{P}_\Omega^2 = \hat{P}_\Omega$ and $\hat{P}_\Omega^\dagger = \hat{P}_\Omega$) whose image is Ω.

Definition 2.3 (Density operator representing a quantum state)**.** Given a state represented by normalised state vector $|\psi\rangle \in \mathcal{H}$, the *density operator* representing the state is a linear operator $\hat{\rho}$: $\mathcal{H} \to \mathcal{H}$ defined by $\hat{\rho} = |\psi\rangle \langle\psi|$.

The density operator $\hat{\rho}$ is another way of representing the state information of the system. As with the state vector formulation above, with $\hat{\rho}$ one can:

- propagate the state information through time using a Hamiltonian operator $\hat{H}$:
$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \hat{\rho} = [\hat{H}, \hat{\rho}] = \hat{H}\hat{\rho} - \hat{\rho}\hat{H}.$$

- calculate the expected value of an observable $\hat{\mathcal{O}}$:

$$\langle \hat{\mathcal{O}} \rangle = \mathrm{tr}\left(\hat{\rho}\hat{\mathcal{O}}\right).$$

- calculate the probability of the state being in a subspace Ω of $\mathcal{H}$:

$$\mathbb{P}(\Omega) = \mathrm{tr}\left(\hat{\rho}\hat{P}_\Omega\right),$$

  where $\hat{P}_\Omega$ is the projection operator whose image is Ω.

Using the definition of $\hat{\rho}$ it easy to show that each of these listed properties is equivalent to its state vector counterpart.

These state vector and density operator formalisms differ in a key manner. Say we have two different states, represented by state vectors $|\psi_1\rangle, |\psi_2\rangle$ and equivalently by density operators $\hat{\rho}_1 = |\psi_1\rangle\langle\psi_1|$, $\hat{\rho}_2 = |\psi_2\rangle\langle\psi_2|$. Define the expectation values of observable $\hat{\mathcal{O}}$ in state i by $\langle\hat{\mathcal{O}}\rangle_i = \langle\psi_i|\hat{\mathcal{O}}|\psi_i\rangle = \mathrm{tr}\hat{\rho}_i\hat{\mathcal{O}}$.

Consider taking linear combinations of the state vectors: for $c_1, c_2 \in \mathbb{C}$, let

$$|\psi\rangle = c_1|\psi_1\rangle + c_2|\psi_2\rangle.$$

(Let us assume that c_1 and c_2 satisfy conditions for $|\psi\rangle$ to be a normalised state vector.) Then this new state gives expectation values

$$\langle\hat{\mathcal{O}}\rangle = \langle\psi|\hat{\mathcal{O}}|\psi\rangle = |c_1|^2\langle\hat{\mathcal{O}}\rangle_1 + |c_2|^2\langle\hat{\mathcal{O}}\rangle_2 + c_1^*c_2\langle\psi_1|\hat{\mathcal{O}}|\psi_2\rangle + c_1c_2^*\langle\psi_2|\hat{\mathcal{O}}|\psi_1\rangle.$$

As we are familiar with, the superposition of state vectors $|\psi_1\rangle$ and $|\psi_2\rangle$ gives a new quantum state, with new observable properties not present in either $|\psi_1\rangle$ or $|\psi_2\rangle$ alone – these arise as the cross-terms $\langle\psi_1|\hat{\mathcal{O}}|\psi_2\rangle$ and $\langle\psi_2|\hat{\mathcal{O}}|\psi_1\rangle$ above.

Now, consider taking a linear superposition of the corresponding density operators: with $C_1, C_2 \in \mathbb{C}$, let

$$\hat{\rho} = C_1\hat{\rho}_1 + C_2\hat{\rho}_2.$$

(Again, let us assume that C_1 and C_2 are such that $\hat{\rho}$ is itself a density operator; we will describe the necessary conditions below). Then this new state gives expectation values

$$\langle\hat{\mathcal{O}}\rangle = \mathrm{tr}\left(\hat{\rho}\hat{\mathcal{O}}\right) = C_1\langle\hat{\mathcal{O}}\rangle_1 + C_2\langle\hat{\mathcal{O}}\rangle_2.$$

Here, we do not find the cross terms we saw in the analogous state vector case. We must conclude that a linear combination of density operators does not represent a new quantum state constructed as a superposition of the old ones.

Rather, by choosing the coefficients C_i of the linear combination to have the properties of a classical probability distribution (i.e. $C_i \in [0, 1]$ and

$\sum_i C_i = 1$), we find that the linear combination of density operators delivers the behaviour expected of a *classical* probability distribution of quantum states:

$$\mathrm{tr}\hat{\rho}\hat{\mathcal{O}} = \sum_i C_i \langle\hat{\mathcal{O}}\rangle_i = \overline{\langle\hat{\mathcal{O}}\rangle},$$

where we have included the over-line above $\langle\hat{\mathcal{O}}\rangle$ to explicitly denote that the quantity takes the form of a classical mean of the quantum expectation values.

The density operator formalism therefore gives us a natural way in which to include classical statistical information in the quantum system.

From the construction of density operators out of state vectors $|\psi_i\rangle$ and probabilities C_i, we can generalise to a list of properties that define a general density operator $\hat{\rho}$.

Definition 2.4 (Density operator)**.** A *density operator* $\hat{\rho} : \mathcal{H} \to \mathcal{H}$ is a linear operator satisfying

(i) Hermiticity: $\hat{\rho}^\dagger = \hat{\rho}$.

(ii) Normalisation: $\mathrm{tr}\hat{\rho} = 1$.

(iii) Non-negative definiteness: for all $|\phi\rangle \in \mathcal{H}$, $\langle\phi|\hat{\rho}|\phi\rangle \geq 0$.

From Item (i), we know that any density operator $\hat{\rho}$ can be diagonalised,

$$\hat{\rho} = \sum_i p_i |e_i\rangle \langle e_i| ,$$

for orthonormal vectors $|e_i\rangle$ and real eigenvalues p_i. Items (ii) and (iii) then ensure that $\sum_i p_i = 1$ and $p_i > 0$ for all i respectively, so that the set of eigenvalues p_i can be interpreted as a classical probability distribution.

Each $|e_i\rangle \langle e_i|$ is itself also a density operator, of the simpler type seen in Definition 2.3 representing a single quantum state. We can formalise the distinction between such simpler operators and the more general ones allowed by Definition 2.4.

Definition 2.5 (Pure and mixed states)**.** A density operator represents a *pure* state if it has only one non-zero eigenvalue. Otherwise, it represents a *mixed* state.

By the normalisation condition of the density operator, it is clear that the single non-zero eigenvalue of a pure state density operator must be exactly 1. A pure state is thus one whose density operator satisfies Definition 2.3; it is purely quantum mechanical in nature. A mixed state, on the other hand, can be interpreted as a classical 'mixture' of multiple quantum states (or more precisely a state which takes takes one of several quantum values, with classical probabilities equal to the density operator's eigenvalues).

Since any density operator has eigenvalues p_i that may be treated as probabilities in a classical mixture of the quantum states, we can now sensibly define an entropy on this probability distribution:

$$S(\hat{\rho}) = -k \sum_i p_i \log p_i.$$

Note that this may be written equivalently in terms of the density operator directly as follows.

Definition 2.6 (von Neumann entropy)**.** Given a density operator $\hat{\rho}$ describing a (pure or mixed) state, the *von Neumann entropy* associated with that state is

$$S_{\text{vN}}(\hat{\rho}) = -k_B \text{tr}\left(\hat{\rho} \log \hat{\rho}\right) = -k_B \sum_{i \in I} p_i \log p_i,$$

where k_B is the Boltzmann constant and $\{p_i\}_{i \in I}$ is the set of non-zero eigenvalues of p_i.

An important point here is that the von Neumann entropy is purely a function of the state of the system, as represented in $\hat{\rho}$, and so is independent of basis.

Theorem 2.5. *Density operator $\hat{\rho}$ represents a pure state if and only if $S_{vN}(\hat{\rho}) = 0$.*

Proof. If $\hat{\rho}$ is a pure state $\hat{\rho} = |e\rangle \langle e|$ then it has exactly one non-zero eigenvalue $p = 1$. So $S_{\text{vN}}(\hat{\rho}) = -k_B 1 \log 1 = 0$.

Conversely, say $S_{\text{vN}}(\hat{\rho}) = 0$. Then $\sum_{i \in I} p_i \log p_i = 0$, where p_i are the non-zero eigenvalues of $\hat{\rho}$. As a result of Definition 2.4, we saw that $p_i > 0$ for all $i \in I$ and that $\sum_{i \in I} p_i = 1$. From the latter, we see that $p_i \leq 1$ also. Therefore $p_i \log p_i \leq 0$ for each $i \in I$. In order for $S_{\text{vN}}(\hat{\rho}) = 0$, we must therefore have $p_i \log p_i = 0$ and so $p_i = 1$ for each $i \in I$. Since $\sum_{i \in I} p_i = 1$, there must be only one non-zero eigenvalue $p_i = 1$. Therefore, $\hat{\rho}$ is pure. □

2.3.2 Reduced density operators and entanglement entropy

Let us return our attention to a tensor product Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, and a density operator $\hat{\rho} \colon \mathcal{H} \to \mathcal{H}$. We will denote the dimension of $\mathcal{H}_A$ by N_A, and likewise the dimension of $\mathcal{H}_B$ by N_B; also, let $N = \min(N_A, N_B)$. When $\hat{\rho}$ represents a pure state, we can write $\hat{\rho} = |e\rangle \langle e|$ for some normalised $|e\rangle \in \mathcal{H}$.

Definition 2.7. Given tensor product Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ and density operator $\hat{\rho} \colon \mathcal{H} \to \mathcal{H}$, the *reduced density operator* $\hat{\rho}_A \colon \mathcal{H}_A \to \mathcal{H}_A$ is given by

$$\hat{\rho}_A = \text{tr}_B \hat{\rho},$$

where tr_B denotes the partial trace over the $\mathcal{H}_B$ portions of the operator.

Likewise, the reduced density operator $\hat{\rho}_B \colon \mathcal{H}_B \to \mathcal{H}_B$ is given by

$$\hat{\rho}_B = \mathrm{tr}_A \hat{\rho}.$$

Given an observer to whom $\mathcal{H}_A$ is accessible but $\mathcal{H}_B$ is not, the reduced density operator $\hat{\rho}_A$ contains all the state information to which the observer has access.

Lemma 2.6. *If the pure state density operator $\hat{\rho} = |e\rangle \langle e|$ represents quantum state $|e\rangle$ with Schmidt decomposition*

$$|e\rangle = \sum_{i=1}^{N} \sigma_i \, |\alpha_i\rangle_A \otimes |\beta_i\rangle_B \,,$$

then the reduced density operators $\hat{\rho}_A$ and $\hat{\rho}_B$ can be written as

$$\hat{\rho}_A = \sum_{i=1}^{N} \sigma_i^2 \left(|\alpha_i\rangle \langle \alpha_i| \right)_A \qquad \textit{and} \qquad \hat{\rho}_B = \sum_{i=1}^{N} \sigma_i^2 \left(|\beta_i\rangle \langle \beta_i| \right)_B .$$

Proof. By direct calculation,

$$\begin{aligned} \hat{\rho} &= |e\rangle \langle e| \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_i \sigma_j^* \left(|\alpha_i\rangle_A \otimes |\beta_i\rangle_B \right) \left({}_A\langle \alpha_j| \otimes {}_B\langle \beta_j| \right) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_i \sigma_j \left(|\alpha_i\rangle \langle \alpha_j| \right)_A \otimes \left(|\beta_i\rangle \langle \beta_j| \right)_B . \end{aligned}$$

Therefore

$$\hat{\rho}_A = \mathrm{tr}_B \hat{\rho} = \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_i \sigma_j \left(|\alpha_i\rangle \langle \alpha_j| \right)_A \mathrm{tr}_B \left(|\beta_i\rangle \langle \beta_j| \right),$$

but $\mathrm{tr}_B \left(|\beta_i\rangle \langle \beta_j| \right) = \langle \beta_j | \beta_i \rangle_B = \delta_{ij}$, so

$$\hat{\rho}_A = \sum_{i=1}^{N} \sigma_i^2 \left(|\alpha_i\rangle \langle \alpha_i| \right)_A .$$

A similar calculation gives the result for $\hat{\rho}_B$. □

The important consequence of the lemma above is that the Schmidt coefficients σ_i of a pure state in $\mathcal{H}_A \otimes \mathcal{H}_B$ determine the eigenvalues of the reduced density operators $\hat{\rho}_A$ and $\hat{\rho}_B$. This provides us with a connection between notions of entanglement, as described by the Schmidt decomposition and Corollary 2.2, and whether or not the reduced states $\hat{\rho}_A$ and $\hat{\rho}_B$ are pure or mixed.

Theorem 2.7. *For pure state $\hat{\rho} = |e\rangle\langle e|$, the reduced density operator $\hat{\rho}_A$ represents a pure state in $\mathcal{H}_A$ if and only if $\hat{\rho}$ represents a non-entangled state in $\mathcal{H}$.*

Proof. From Lemma 2.6, we have that

$$\hat{\rho}_A = \sum_{i=1}^{N} \sigma_i^2 \left(|\alpha_i\rangle\langle\alpha_i|\right)_A .$$

Then $\hat{\rho}_A$ is pure if and only if $\sigma_i^2 = \delta_{ik}$ for some $k \in \{1, \ldots, N\}$. This is true if and only if $|e\rangle = |\alpha_k\rangle_A \otimes |\beta_k\rangle_B$, i.e. $\hat{\rho}$ represents a non-entangled state. □

This is an important piece of physical information. Even if we assume perfect knowledge of the quantum state of the total system in $\mathcal{H}_A \otimes \mathcal{H}_B$ so that $\hat{\rho}$ is pure rather than mixed, an observer with access to only part of the system (i.e. only $\mathcal{H}_A$ via $\hat{\rho}_A$, or only $\mathcal{H}_B$ via $\hat{\rho}_B$) will effectively see a classical mixture of different quantum states whenever the full state is entangled between A and B.

A simple example may serve to illustrate this. Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be identical two-dimensional spaces, with orthonormal basis $\{|0\rangle, |1\rangle\}$ for both. Consider the pure state of the combined system $\mathcal{H}_A \otimes \mathcal{H}_B$ given by the vector (called a Bell state)

$$|\phi\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle_A \otimes |0\rangle_B - |1\rangle_A \otimes |1\rangle_B \Big).$$

In the chosen basis, $|\phi\rangle$ is already in its Schmidt-decomposed form, with Schmidt coefficients $\sigma_1 = \sigma_2 = 1/\sqrt{2}$ so that the state is entangled. The density operator describing this same pure entangled state is

$$\begin{aligned}\hat{\rho} = |\phi\rangle\langle\phi| = \frac{1}{2}\Big[&\left(|0\rangle\langle 0|\right)_A \otimes \left(|0\rangle\langle 0|\right)_B - \left(|0\rangle\langle 1|\right)_A \otimes \left(|0\rangle\langle 1|\right)_B \\ &- \left(|1\rangle\langle 0|\right)_A \otimes \left(|1\rangle\langle 0|\right)_B + \left(|1\rangle\langle 1|\right)_A \otimes \left(|1\rangle\langle 1|\right)_B \Big].\end{aligned}$$

Then using that $\mathrm{tr}\left(|i\rangle\langle j|\right) = \delta_{ij}$ to trace over $\mathcal{H}_B$, we find the reduced density operator,

$$\hat{\rho}_A = \frac{1}{2}\Big[|0\rangle\langle 0| + |1\rangle\langle 1| \Big]_A = \frac{1}{2}\mathbb{1}_A.$$

Evidently, $\hat{\rho}_A$ has multiple non-zero eigenvalues, and so represents a mixed state in the component $\mathcal{H}_A$ as per Definition 2.5. This is in agreement Theorem 2.7, since the original state $|\phi\rangle$ in the full space $\mathcal{H}_A \otimes \mathcal{H}_B$ was entangled.

On the other hand, we can consider a non-entangled state,

$$|\theta\rangle = |0\rangle_A \otimes |1\rangle_B .$$

This has corresponding density operator $\hat{\rho} = (|0\rangle\langle 0|)_A \otimes (|1\rangle\langle 1|)_B$ and therefore reduced density operator

$$\hat{\rho}_A = \mathrm{tr}_B \hat{\rho} = (|0\rangle\langle 0|)_A .$$

Since $\hat{\rho}_A$ has only one non-zero eigenvalue, it represents a pure state on $\mathcal{H}_A$, again in agreement with Theorem 2.7.

Definition 2.8 (Entanglement entropy)**.** With density operator $\hat{\rho}$ as above, the *entanglement entropy* $S(A)$ of the state between subsystem A and its complement B is given by the von Neumann entropy of the reduced density operator $\hat{\rho}_A$:

$$S(A) = S_{\mathrm{vN}}(\hat{\rho}_A).$$

From this definition and the preceding discussion, we can immediately see the following properties of entanglement entropy for a pure state in particular.

Corollary 2.8. *Say $\hat{\rho}$ represents a pure state in $\mathcal{H}_A \otimes \mathcal{H}_B$. Then*

$$S(A) = S(B).$$

Proof. From Lemma 2.6, the reduced density matrices $\hat{\rho}_A$ and $\hat{\rho}_B$ of the pure state $\hat{\rho}$ share exactly the same eigenvalues. Therefore, their von Neumann entropies are identical. □

Corollary 2.9. *Let $\hat{\rho}$ represent a pure state in $\mathcal{H}_A \otimes \mathcal{H}_B$. Then the entanglement entropies vanish, $S(A) = S(B) = 0$, if and only if the state $\hat{\rho}$ is not entangled between A and B.*

Proof. This follows directly from Theorems 2.5 and 2.7. □

This establishes the most basic groundwork of an interpretation of entanglement entropy as a measure of how entangled a state is between two subsystems of a physical system.

There are many more properties of both classical (Shannon) entropy of information and entanglement entropy, of increasing importance and sophistication (Nielsen and Chuang 2010). For instance, the entanglement between various parts of a many-component tensor product space $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C \otimes \ldots$ is often of interest. In general, this is challenging since a full generalisation of the Schmidt decomposition to this multipartite system only exists under certain conditions (Bengtsson and Życzkowski 2017).

Instead, the properties of the bipartite entanglement and entanglement entropy described above can be applied by dividing the many tensor components of the Hilbert space into two groups. So, splitting it into $\mathcal{H}_A$ and $(\mathcal{H}_B \otimes \mathcal{H}_B \otimes \ldots)$, we can define reduced density operators as $\hat{\rho}_A = \mathrm{tr}_{B,C,\ldots}(\hat{\rho})$ by tracing out all components in the tensor product other than A. Likewise, we can split the Hilbert space into $(\mathcal{H}_A \otimes \mathcal{H}_B)$ and $(\mathcal{H}_C \otimes \mathcal{H}_D \otimes \ldots)$;

then $\hat{\rho}_{AB} = \mathrm{tr}_{C,D,\ldots}(\hat{\rho})$, and we denote the associated entanglement entropy $S(AB) = S_{\mathrm{vN}}(\hat{\rho}_{AB})$. In this context, two well-known properties of entanglement entropy are subadditivity,

$$S(AB) \leq S(A) + S(B),$$

and the Araki-Lieb (or triangle) inequality,

$$S(AB) \geq |S(A) - S(B)|\,.$$

An especially important result, relatively easy to prove for Shannon entropy but more challenging for von Neumann entropy, is strong subadditivity:

$$S(ABC) + S(B) \leq S(AB) + S(BC).$$

We will not prove these important results; the basic construction of entanglement entropy we have given in this chapter will suffice to underpin the notions of holographic entanglement entropy that we explore in the remainder of this work.

2.4 Entanglement entropy in quantum field theories

In the preceding discussion of entanglement and entanglement entropy, the Hilbert spaces of interest have been finite dimensional. However, the Hilbert space of states in a quantum field theory is infinite-dimensional. While axiomatic approaches to QFT do pose that the Hilbert spaces of physical interest are at least separable (Streater and Wightman 1964, pp. 85-87) and so have countable basis sets, we can nonetheless expect complications when discussing entanglement entropy in QFT.

At a naïve level, one can think of a QFT as a continuum limit of a quantum mechanical system defined on a lattice. By 'continuum limit', we mean taking the lattice spacing to be infinitesimally small, so that the lattice looks a little bit like a smooth space. In this picture, the total Hilbert space of the theory is a tensor product of the state spaces at each lattice point. Therefore, it is sensible to consider the entanglement of a spatial region A (i.e. the state spaces of all lattice points within A) with the rest of the lattice.

Since a true QFT corresponds to the continuum limit with infinitely many lattice points in A, it is unsurprising that any entanglement entropy one calculates is divergent and in need of regularisation. This is familiar in the context of QFT, and can be accomplished using, for instance, a UV cut-off scale. Equivalently, one can impose an infinitesimal but non-zero lattice-spacing ϵ between points on the lattice, and calculate an entanglement entropy S with this spacing as a parameter. This results in an expression for S that diverges as $\epsilon \to 0$ in the continuum limit.

There are proposals (Liu and Mezei 2013) on how to obtain a finite expression representing a renormalised entanglement entropy from this divergent S. Alternatively, such divergent quantities S written in terms of a cut-off ϵ can be used to calculate other physically meaningful quantities such as the mutual information between two regions A and B,

$$I(A:B) := S(A) + S(B) - S(AB),$$

wherein the divergent parts of the entanglement entropies are expected to cancel out as $\epsilon \to 0$.

However, actually calculating an entanglement entropy directly in a field theory is challenging. In brief, rather than directly computing a reduced density operator $\hat{\rho}_A$ and its von Neumann entropy, one calculates the Rényi entropies

$$S^{(n)}(A) := \frac{1}{1-n} \log \mathrm{tr} \hat{\rho}_A^n,$$

for integers $n > 1$. Here, the quantity $\mathrm{tr}\hat{\rho}_A^n$ can be calculated by path integral methods via the 'replica trick', which is detailed in Calabrese and Cardy (2009). From these Rényi entropies, one infers the von Neumann entanglement entropy as $S(A) = \lim_{n\to 1} S^{(n)}(A)$. This limiting process is often termed an analytic continuation, though it is not mathematically rigorous.

Compared to this procedure, the Ryu-Takayanagi prescription (1.1) offers an alternative that is simple and elegantly geometric. Of course, in order to establish the validity of (1.1), it must be checked against the aforementioned procedure. However, our aim is not to prove, but simply to define and use, the Ryu-Takayanagi holographic entanglement entropy. To do so, we must first understand some of the geometry underlying the holographic principle.

Chapter 3

Conformal completions and the boundary at infinity

The Ryu-Takayanagi prescription (1.1) for holographic entanglement entropy relates the entanglement entropy in a conformal field theory on the 'boundary at infinity' of a bulk spacetime, to a minimal surface within that bulk spacetime. Our task in this chapter is to understand the geometric notion of a 'boundary at infinity' of a spacetime manifold. In order to do so, we will first need to introduce the concepts of conformal maps between finite-dimensional Riemannian or pseudo-Riemannian manifolds. This is because the 'boundary at infinity' of a spacetime manifold is in actuality the conformal boundary, and must be understood by conformally embedding the spacetime of interest into another manifold.

The introduction of conformal maps here also lays the groundwork for Chapter 6, where we explore some of the geometric consequences of the conformal symmetry of a conformal field theory.

3.1 Conformal maps

In preparation for an exposition of conformal maps and their role in understanding the 'boundary at infinity' of a spacetime, let us recall the notions of the pullback f^* as well as the pushforward (or tangent map) f_* associated with a smooth map $f : \mathcal{M} \to \mathcal{N}$ between smooth manifolds $\mathcal{M}$ and $\mathcal{N}$. The pushforward at p, $f_* : T_p\mathcal{M} \to T_{f(p)}\mathcal{N}$, is a map associated with f that takes vectors tangent to $\mathcal{M}$ at p to vectors tangent to $\mathcal{N}$ at $f(p)$. The pushforward is defined such that for any $X_p \in T_p\mathcal{M}$,

$$[f_* X_p](F) := X_p(F \circ f) \qquad \text{for all } F \in C^\infty(\mathcal{N}). \tag{3.1}$$

When unambiguous, we will use the notation $X'_{f(p)} := f_* X_p$ so that under the pushforward, $X_p \mapsto X'_{f(p)}$.

In coordinates $\mathbf{x}$ on $\mathcal{M}$, and with $\mathbf{y} = f(\mathbf{x})$ on $\mathcal{N}$, a tangent vector X_p at p has coordinate representation

$$X_p = \alpha^i \left.\frac{\partial}{\partial x^i}\right|_p ,$$

for some real components α^i. Then, from the coordinate-free definition (3.1) and the chain rule, we see that the pushforward is represented in the coordinate basis by the Jacobian matrix $J_f = \frac{\partial f}{\partial x}$ of $\mathbf{y} = f(\mathbf{x})$:

$$X'_{f(p)} = \alpha^i \frac{\partial f^j}{\partial x^i}(p) \left.\frac{\partial}{\partial y^j}\right|_{f(p)} = \left[J_f(p)^j{}_i \; \alpha^i\right] \left.\frac{\partial}{\partial y^j}\right|_{f(p)} . \tag{3.2}$$

The term 'pushforward' denotes the fact that tangent vectors are mapped 'forwards' or in the same direction as f, i.e. from tangent spaces of $\mathcal{M}$ to those of $\mathcal{N}$.

We can push any particular tangent vector forward, $X_p \mapsto X'_{f(p)}$. However, if we take a whole vector field X on $\mathcal{M}$ and apply pushforwards at each point, $X \mapsto X'$, then the X' so constructed will not necessarily be a well-defined vector field on $\mathcal{N}$. For instance, if f is not surjective then there will be regions of $\mathcal{N}$ with no associated tangent vector. If f happens to be a diffeomorphism, then this X' will be a vector field. In particular, if vector field X on $\mathcal{M}$ is given by

$$X = \alpha^i(\mathbf{x}) \frac{\partial}{\partial x^i},$$

then using the inverse map $\mathbf{x} = f^{-1}(\mathbf{y})$, we can write the pushforward X' as

$$X' = \left[J_f(f^{-1}(\mathbf{y}))^j{}_i \; \alpha^i(f^{-1}(\mathbf{y}))\right] \frac{\partial}{\partial y^j}.$$

The pullback f^* is defined on cotangent spaces and maps covectors of $\mathcal{N}$ to those of $\mathcal{M}$. It therefore maps 'backwards' or in the opposite direction to f. Specifically, $f^* : T^*_{f(p)}\mathcal{N} \to T^*_p\mathcal{M}$ is defined such that for covector $\phi_{f(p)} \in T^*_{f(p)}\mathcal{N}$,

$$\left[f^*\phi_{f(p)}\right](X_p) := \phi_{f(p)}\left(f_* X_p\right) \qquad \text{for all } X_p \in T_p\mathcal{M}. \tag{3.3}$$

Analogous to the notation used for vectors, we write $\phi'_p := f^*\phi_{f(p)}$, so that the pullback maps $\phi_{f(p)} \mapsto \phi'_p$.

We can represent a covector field ϕ on $\mathcal{N}$ in our coordinates as

$$\phi = \beta_j(\mathbf{y}) \mathrm{d}y^j.$$

Then, from the coordinate free definition (3.3) and the coordinate representation of the pushforward (3.2), we have ϕ' given in coordinates as

$$\phi' = \beta_j(f(\mathbf{x})) \frac{\partial f^j}{\partial x^i} \mathrm{d}x^i.$$

Unlike the pushforward, one can always pull whole covector fields ϕ on $\mathcal{N}$ back to give new covector fields, $\phi \mapsto \phi'$ on $\mathcal{M}$.

The definitions of both the pushforward and pullback can be naturally extended to tensors: the pushforward can map arbitrary contravariant tensors on $\mathcal{M}$ to those on $\mathcal{N}$, while the pullback can map arbitrary covariant tensors in the opposite direction.

In particular, for Riemannian or pseudo-Riemannian manifolds $(\mathcal{M}, g)$ and $(\mathcal{N}, h)$, the metrics g and h are covariant 2-tensors. So, given a smooth map $f : \mathcal{M} \to \mathcal{N}$, we can act on h with the pullback f^*. For $X_p, Y_p \in T_p\mathcal{M}$,

$$f^*h\big(X_p, Y_p\big) = h\big(f_*X_p, f_*Y_p\big) = h\big(X'_{f(p)}, Y'_{f(p)}\big). \tag{3.4}$$

In coordinates, the metric on $\mathcal{N}$ can be represented as $h = h_{ij}(\mathbf{y})\mathrm{d}y^i\mathrm{d}y^j$, where the juxtaposition of 1-forms represents the symmetric tensor product:

$$\mathrm{d}y^i\mathrm{d}y^j := \frac{1}{2}\left(\mathrm{d}y^i \otimes \mathrm{d}y^j + \mathrm{d}y^j \otimes \mathrm{d}y^i\right).$$

Then, in coordinates, the pullback of the metric is given by

$$f^*\left[h_{ij}(\mathbf{y})\mathrm{d}y^i\mathrm{d}y^j\right] = h_{mn}(f(\mathbf{x}))\frac{\partial f^m}{\partial x^i}\frac{\partial f^n}{\partial x^j}\mathrm{d}x^i\mathrm{d}x^j = h'_{ij}(\mathbf{x})\mathrm{d}x^i\mathrm{d}x^j. \tag{3.5}$$

Therefore, the components h_{ij} of the coordinate representation of the metric tensor h transform as

$$h'_{ij}(\mathbf{x}) := h_{mn}(f(\mathbf{x}))\frac{\partial f^m}{\partial x^i}\frac{\partial f^n}{\partial x^j}.$$

This is applicable not only to smooth maps $f : \mathcal{M} \to \mathcal{N}$ between manifolds $\mathcal{M}$ and $\mathcal{N}$, but also for transformations $f : \mathcal{M} \to \mathcal{M}$ of a manifold $\mathcal{M}$ back to itself (including changes of coordinates).

With the language of pullbacks and pushforwards (both with and without coordinate representation) at our disposal, we can explore classes of maps between manifolds that we will find useful. For the rest of the discussion, let $(\mathcal{M}, g)$ and $(\mathcal{N}, h)$ be Riemannian or pseudo-Riemannian manifolds. The first and most obvious sort of map that may be of interest is one which preserves the metric:

Definition 3.1 (Local isometry)**.** A smooth map $f : (\mathcal{M}, g) \to (\mathcal{N}, h)$ is called a *local isometry* if

$$f^*h = g.$$

The metric tensor g is used to define the notion of distance (Lee 1997), i.e. a metric in the sense of metric spaces, on $(\mathcal{M}, g)$. For $a, b \in \mathcal{M}$, the distance between a and b can be defined as

$$d(a, b) = \inf_\gamma \int_a^b \sqrt{g(\dot\gamma(t), \dot\gamma(t))}\mathrm{d}t,$$

where the infimum is taken over a space of suitably well-behaved curves γ connecting a and b.

Provided that a local isometry is suitably well-behaved at the global level, it may be able to preserve the notion of distance on a manifold. In particular, a local isometry that is also a diffeomorphism will preserve distances between points. Such 'global isometries' are simply called isometries.

However, this class of maps is too restrictive for our purposes; we will be interested in transformations of spacetime manifolds that do not preserve distance, but which do preserve as much other structure as possible.

Definition 3.2 (Conformal map)**.** A smooth map $f : (\mathcal{M}, g) \to (\mathcal{N}, h)$ is called *conformal* if there is some $\Lambda \in C^\infty(\mathcal{M})$, with $\Lambda(p) > 0$ for all $p \in \mathcal{M}$, such that

$$f^*h = \Lambda^2 g.$$

Λ is called the *conformal factor* of f.

Local isometries can be regarded as a special case of conformal maps with constant conformal factor $\Lambda = 1$.

The conformal property of a map has a very specific geometric meaning. To elucidate this meaning, we can rephrase the definition of conformality:

Proposition 3.1. *A smooth map $f : (\mathcal{M}, g) \to (\mathcal{N}, h)$ is conformal if and only if for all $p \in \mathcal{M}$,*

$$f_*|_p = \Lambda(p) R_p,$$

where $\Lambda(p)$ is the conformal factor of f and $R_p : T_p\mathcal{M} \to T_{f(p)}\mathcal{N}$ is a linear isometry, i.e. a linear map such that $h(R_pX_p, R_pY_p) = g(X_p, Y_p)$ for all $X_p, Y_p \in T_p\mathcal{M}$.

Proof. Let p be any point in $\mathcal{M}$, and consider any $X_p, Y_p \in T_p\mathcal{M}$.

Say f is a smooth map such that $f_*|_p = \Lambda(p)R_p$ where R_p is a linear isometry. Then

$$\begin{aligned} f^*h(X_p, Y_p) &= h\left(\Lambda(p)R_pX_p, \Lambda(p)R_pY_p\right) = \Lambda^2(p)\; h(R_pX_p, R_pY_p) \\ &= \Lambda^2(p)\; g(X_p, Y_p), \end{aligned}$$

so f is conformal.

Conversely, say f is a smooth conformal map. From (3.4), we have that

$$h\left(f_*X_p, f_*Y_p\right) = \Lambda^2(p) g\left(X_p, Y_p\right).$$

Since a conformal factor must satisfy $\Lambda(p) \neq 0$, and tensors (in particular, h) are multilinear over $C^\infty(\mathcal{M})$, this can be rearranged as

$$h\left(\frac{f_*}{\Lambda}X_p, \frac{f_*}{\Lambda}Y_p\right) = g\left(X_p, Y_p\right),$$

so that $\left(f_*/\Lambda\right)|_p$ is a linear isometry. □

Note that the linear isometries R_p above are not the local isometries of Definition 3.1, despite the similar terminology. Linear isometries are maps between inner product spaces that preserve both vector space structure and inner products. Of course, the dependence of R_p on the point p is as smooth as the dependence of f_* on p.

Whereas our original definition of conformality, Definition 3.2, applied globally everywhere on the domain of the map, this reformulation is inherently local. In particular, it allows us to define conformality at a specific point $p \in \mathcal{M}$.

Definition 3.3. A map $f : (\mathcal{M}, g) \to (\mathcal{N}, h)$ that is smooth at point $p \in \mathcal{M}$, is *conformal at* p if

$$f_*|_p = \Lambda(p) R_p,$$

where $\Lambda(p)$ is a positive number, and $R_p : T_p\mathcal{M} \to T_{f(p)}\mathcal{M}$ is a linear isometry.

Then a map is conformal in our original global sense if it is smooth and conformal at every point $p \in \mathcal{M}$.

Since linear isometries are injective, and since $\Lambda(p) \neq 0$ always, Proposition 3.1 already shows that f_* is necessarily injective at all points on $\mathcal{M}$ if f is conformal. However, the proposition contains much more information than this.

Geometrically, since linear isometries preserve inner products, they are exactly those maps which preserve both lengths of, and angles between, vectors. For instance, linear isometries mapping the Euclidean space to itself are just rotations and reflections. Therefore, Proposition 3.1 gives us an elegant interpretation: the pushforward of a conformal map at a point always preserves angles between vectors at that point, and scales all such vectors (independent of their direction) by the conformal factor at that point. In this sense, we say that conformal maps are those which, locally, act as an isotropic scaling which preserves angles. This must be a local notion since we are actually talking about the tangent map rather than the original map;

Let us demonstrate these properties in action. To avoid complications, we specialise to Riemannian rather than pseudo-Riemannian manifolds, so that the metric tensors we consider are positive definite. Since a metric tensor defines an inner product on each tangent space, it also defines a norm:

$$|X_p|_g := \sqrt{g\left(X_p, X_p\right)}.$$

Therefore, from (3.4) and Definition 3.2,

$$\left|X'_{f(p)}\right|_h = \sqrt{f^*h\left(X_p, X_p\right)} = |X_p|_{f^*h} = \Lambda(p)\left|X_p\right|_g, \tag{3.6}$$

which is precisely the aforementioned local isotropic scaling.

As well as that, take $X_p, Y_p \in T_p\mathcal{M}$. Then we can determine the angle θ between X_p and Y_p:

$$\cos\theta = \frac{g(X_p, Y_p)}{|X_p|_g |Y_p|_g}.$$

This is a direct generalisation of the familiar formula $\cos\theta = X \cdot Y/(|X||Y|)$ for vectors X, Y in Euclidean space.

We can likewise determine the angle θ' between $X'_{f(p)}$ and $Y'_{f(p)}$:

$$\cos\theta' = \frac{h\big(X'_{f(p)}, Y'_{f(p)}\big)}{\big|X'_{f(p)}\big|_h \big|Y'_{f(p)}\big|_h} = \frac{f^*h\big(X_p, Y_p\big)}{\big|X_p\big|_{f^*h} \big|Y_p\big|_{f^*h}}.$$

Then recalling both (3.6) and the definition of conformal maps, we find $\cos\theta' = \cos\theta$. Evidently, the conformal property is precisely the property which gives at least $|\theta| = |\theta'|$, i.e. preserves the magnitude of angles.

We may also ask whether there is a relationship between the signs of θ and θ'. However, it is not always possible to sensibly and uniformly assign a sign to angles, as we know from the familiar case of angles between vectors in $\mathbb{R}^3$.

In the case where the target manifold $\mathcal{N}$ has the same dimension as the domain $\mathcal{M}$, there is a more general property of f that describes the same information as this relative sign of angles. Let J_f be the Jacobian of a smooth map f. Since the $\mathcal{M}$ and $\mathcal{N}$ have the same dimension, so do their tangent spaces; therefore, J_f is a square matrix and we can calculate its determinant. If $\det J_f > 0$ at a point, we say f is *orientation-preserving* at that point; if $\det J_f < 0$ then f is *orientation-reversing*. Comparing this to Proposition 3.1 and recalling that the Jacobian J_f is nothing but a coordinate representation of the pushforward f_*, we see that if a conformal map f with $f_* = \Lambda R$ is orientation-preserving, it must have $\det R > 0$. This means that R is orientation-preserving in the sense of linear maps. Likewise, if f is orientation-reversing, so must R be.

It is clear that $\det J_f \neq 0$ everywhere, since linear isometry R must have non-zero determinant, and $\Lambda > 0$ strictly. Since the determinant is a continuous map, this means that $\det J_f$ has definite sign on each connected component of $\mathcal{M}$. In particular, if $\mathcal{M}$ is a connected manifold then the conformal map f is either everywhere orientation-preserving or everywhere orientation-reversing.

It is sometimes convenient to distinguish between those conformal maps which are orientation-preserving and those which are orientation-reversing. In such cases, we call only those which preserve orientation conformal, while those which reverse orientation are called anti-conformal.

In cases such as $\mathbb{R}^2$ where angles can be given a meaningful sign, f will give θ' the same sign as θ when f is orientation-preserving, and the opposite sign if it is orientation-reversing. Since $\mathbb{R}^2$ is connected, this means that a

conformal map f on $\mathbb{R}^2$ either preserves the relative sign of angles everywhere, or swaps them everywhere.

3.2 Conformal completion of space and spacetime manifolds

Holography aims to relate two physical theories: one in the bulk of a spacetime manifold, and one on the 'boundary at infinity' of that manifold. With an understanding of conformal maps from the previous section, we explore the notion of a 'boundary at infinity'.

In order to do this, we need to develop an adequate means of talking about 'infinity' on a pseudo-Riemannian spacetime manifold $(\mathcal{M}, g)$. The strategy we use is to find a map $f : \mathcal{M} \to \mathcal{N}$ that embeds $\mathcal{M}$ in another manifold, $f(\mathcal{M}) \subseteq \mathcal{N}$, such that 'points' corresponding to infinity on $\mathcal{M}$ are mapped to finite points on the boundary (in the topological sense) of $f(\mathcal{M})$. In other words, $\mathcal{M}$'s points at infinity are $\partial\,[f(\mathcal{M})] \subseteq \mathcal{N}$; one can regard the completed spacetime, including its points at infinity, as the closed region $\overline{f(\mathcal{M})} = f(\mathcal{M}) \cup \partial\,[f(\mathcal{M})]$.

The chosen map – and therefore chosen $\overline{f(\mathcal{M})}$ – ought to preserve as much of the structure of $\mathcal{M}$ as possible. How much structure can be preserved, while still changing infinite distances into finite ones?

We certainly cannot preserve the metric entirely, i.e. we cannot choose f to be an isometry. Distances on a manifold are determined by the metric tensor. So, if the metric was preserved, then an infinite distance on $\mathcal{M}$ would also be infinite on $f(\mathcal{M}) \subset \mathcal{N}$.

However, by insisting that our map be conformal, we can at least preserve the notion of angle or its generalisation to the pseudo-Riemannian manifolds such as are used to model spacetime. In particular, a conformal map f will send null vectors to null vectors, timelike vectors to timelike vectors and spacelike vectors to spacelike vectors – so a conformal map f will preserve causal structure of the spacetime.

The process of conformally embedding a spacetime manifold $\mathcal{M}$ into another manifold $\mathcal{N}$ so that the infinity of the first is brought to a finite distance in the second can be called 'conformal completion' of $\mathcal{M}$. The 'boundary at infinity' of $\mathcal{M}$ is thus more precisely named the conformal boundary of $\mathcal{M}$. To denote such a boundary, and emphasise that it is not simply the usual topological boundary of a manifold, we will denote it as $\partial^c\mathcal{M}$.

Conformal completion is related to, but not the same as, the topological notion of compactification. For instance, the conformal completion of $\mathbb{E}^d$ we present below is an example of a one-point compactification (Munkres 2000). For this reason, the physics literature sometimes uses the term 'conformal compactification' for the resulting $\overline{f(\mathcal{M})}$, but this need not necessarily be

compact. Below we will see a case in which the timelike dimension of a spacetime manifold does not get compactified, for instance.

To begin with, in Section 3.2.1 we will demonstrate a conformal completion of d-dimensional Euclidean space $\mathbb{E}^d$. This will provide a suitably simple first example before we look at Minkowski and AdS spacetimes in Sections 3.2.2 and 3.2.3 respectively. Beyond its simplicity as a first example, we will find the compactification of $\mathbb{E}^d$ useful in understanding constant-time sections of (conformally completed) Minkowski spacetime, and also of the conformal boundary of AdS.

In the Euclidean case, we simply propose an appropriate map and subsequently show it to be a conformal embedding. For Minkowski and AdS spacetimes we will use the method, familiar from Penrose diagrams, of explicitly building an appropriate conformal embedding.

3.2.1 Conformal completion of Euclidean space $\mathbb{E}^d$

The Euclidean space $\mathbb{E}^d = (\mathbb{R}^d, g_{\mathbb{E}^d})$ can be described by rectangular coordinates $\mathbf{x} = \left(x^1, \ldots, x^d\right) \in \mathbb{R}^d$, with the metric

$$g_{\mathbb{E}^d} = \sum_{i=1}^{d} (\mathrm{d}x^i)^2.$$

We take the d-dimensional unit sphere $\mathbb{S}^d$ to be defined as a surface in $\mathbb{E}^{d+1}$:

$$\mathbb{S}^d := \left\{ \mathbf{y} \in \mathbb{E}^{d+1} \ : \ \mathbf{y} \cdot \mathbf{y} = \sum_{i=1}^{d+1} \left(y^i\right)^2 = 1 \right\}.$$

If we denote the natural inclusion $\iota : \mathbb{S}^d \hookrightarrow \mathbb{E}^{d+1}$, then a metric on $\mathbb{S}^d$ is induced: $g_{\mathbb{S}^d} = \iota^* g_{\mathbb{E}^{d+1}}$.

With rectangular coordinates for $\mathbb{E}^{d+1}$, we use coordinates $\mathbf{y} \in \mathbb{S}^d \subseteq \mathbb{E}^{d+1}$ for the sphere, so necessarily $\mathbf{y}^2 := \mathbf{y} \cdot \mathbf{y} = 1$. Let $\mathbf{N} = (0, \ldots, 0, 1)$ define the North pole of $\mathbb{S}^d$.

We will likewise treat $\mathbb{E}^d$ as a surface in $\mathbb{E}^{d+1}$. Using rectangular coordinates $\mathbf{x}$ of $\mathbb{E}^{d+1}$, the $x^{d+1} = 0$ plane of $\mathbb{E}^{d+1}$ is identical (isometric) to $\mathbb{E}^d$. Hence a coordinate $\mathbf{x} \in \mathbb{E}^{d+1}$ describing $\mathbb{E}^d$ satisfies $\mathbf{N} \cdot \mathbf{x} = x^{d+1} = 0$.

We describe the well-known stereographic projection $\phi : \mathbb{E}^d \to \mathbb{S}^d$, which we will subsequently show to be the desired conformal embedding.

Let $\mathbf{y} = \phi(\mathbf{x})$ be the point on $\mathbb{S}^d$ that lies on the straight line passing through $\mathbf{x}$ and $\mathbf{N}$, as shown in Fig. 3.1. It is clear geometrically that all points $\mathbf{y} \in \mathbb{S}^d$ other than $\mathbf{N}$ have some corresponding line, and thus some such corresponding $\mathbf{x} \in \mathbb{E}^d$.

However, there is no finite $\mathbf{x}$ on the $x^{d+1} = 0$ plane for which the constructed line will only touch the sphere at $\mathbf{y} = \mathbf{N}$; there are always

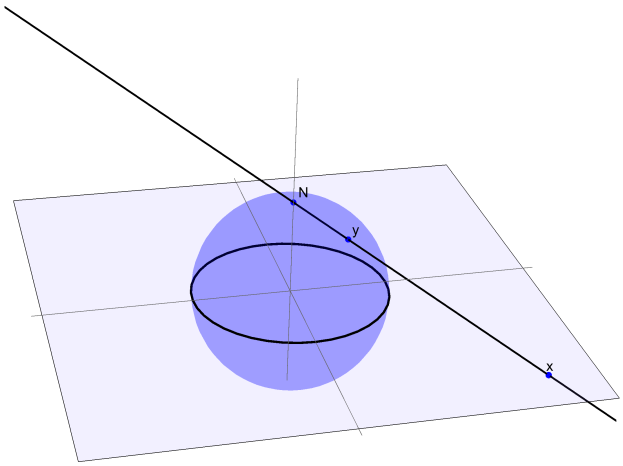


Figure 3.1: Construction of the stereographic projection from $\mathbb{E}^2$ to $\mathbb{S}^2$, embedded in $\mathbb{E}^3$. $\mathbf{N} = (0, 0, 1)$ is the North pole of the sphere $\mathbb{S}^2$, while $\mathbf{x}$ and $\mathbf{y}$ are corresponding points on $\mathbb{E}^2$ and $\mathbb{S}^2$ respectively.

two intersections of the line with the sphere. From this, we conclude that $\phi\left(\mathbb{E}^d\right) = \mathbb{S}^d \setminus \{\mathbf{N}\}$.

In the limit $\mathbf{x} \to \infty$, though, the constructed line becomes parallel with the $x^{d+1} = 0$ plane, and so $\mathbf{y} \to \mathbf{N}$. Thus (at least at the level of such limits) ϕ would seem to map the infinity of $\mathbb{E}^d$ to finite point $\mathbf{N}$ on $\mathbb{S}^d$; so, our choice is a good candidate for the conformal completion of $\mathbb{E}^d$.

It remains to demonstrate that ϕ is a conformal map. It follows from the geometric construction of Fig. 3.1 that for some $t \in \mathbb{R}$,

$$t\,(\mathbf{y} - \mathbf{N}) = \mathbf{x} - \mathbf{N}. \tag{3.7}$$

The value of t can be determined from the $(d+1)$-th component of this equation:

$$t\left(y^{d+1} - 1\right) = -1 \qquad \Longrightarrow \qquad t = \frac{1}{1 - y^{d+1}}.$$

Inserting this back into (3.7), we find after some manipulation that

$$\mathbf{y} = \left(1 - y^{d+1}\right)\mathbf{x} + y^{d+1}\mathbf{N}. \tag{3.8}$$

We can substitute this into the condition that $\mathbf{y}^2 = 1$. Doing so, and using that $\mathbf{N} \cdot \mathbf{x} = 0$ and $\mathbf{N}^2 = 1$, gives a quadratic for y^{d+1}:

$$\left(1 - y^{d+1}\right)^2 \mathbf{x}^2 + \left(y^{d+1}\right)^2 = 1,$$

which can be written as

$$\left[y^{d+1}-1\right]\left[\left(1+\mathbf{x}^2\right)y^{d+1}+\left(1-\mathbf{x}^2\right)\right]=0.$$

The two solutions of this quadratic correspond to the two points at which the line we have constructed intersects the sphere; the $y^{d+1}=1$ solution describes the point $\mathbf{N}$. The other solution is $y^{d+1}=\left(\mathbf{x}^2-1\right)/\left(\mathbf{x}^2+1\right)$. Using this in (3.8), we find

$$\mathbf{y}=\phi(\mathbf{x})=\frac{2}{\mathbf{x}^2+1}\mathbf{x}+\frac{\mathbf{x}^2-1}{\mathbf{x}^2+1}\mathbf{N}.$$

In components, this means

$$\left(y^1,\ldots,y^d,y^{d+1}\right)=\left(\frac{2x^1}{\mathbf{x}^2+1},\ldots,\frac{2x^d}{\mathbf{x}^2+1},\frac{\mathbf{x}^2-1}{\mathbf{x}^2+1}\right). \tag{3.9}$$

To show that ϕ is conformal, we can now explicitly evaluate $\phi^*g_{\mathbb{S}^d}=\phi^*\iota^*g_{\mathbb{E}^{d+1}}$ using (3.9). Note that $\phi^*\iota^*=(\iota\circ\phi)^*=\phi^*$ since $\iota\circ\phi=\phi$; this is just a fancy way of writing that $\phi(\mathbf{x})^2=1$ always, which we know by construction. So, recalling the pullback of a metric tensor (3.5),

$$\phi^*g_{\mathbb{S}^d}=\phi^*\left(\sum_{i=1}^{d+1}\left(\mathrm{d}y^i\right)^2\right)=\sum_{i=1}^{d+1}\left(\sum_{j=1}^{d}\frac{\partial y^i}{\partial x^j}\mathrm{d}x^j\right)^2.$$

Using (3.9) to evaluate this expression, one finds that

$$\phi^*g_{\mathbb{S}^d}=\frac{4}{\left(\mathbf{x}^2+1\right)^2}\sum_{j=1}^{d}\left(\mathrm{d}x^j\right)^2=\frac{4}{\left(\mathbf{x}^2+1\right)^2}g_{\mathbb{E}^d},$$

demonstrating that $\phi:\mathbb{E}^d\to\mathbb{S}^d$ is conformal.

We can conclude that $\overline{\phi(\mathbb{E}^d)}=\overline{\mathbb{S}^d\setminus\{\mathbf{N}\}}=\mathbb{S}^d$ is a conformal completion of the Euclidean space. Indeed in this case it is compact, and so could be called a conformal compactification of the Euclidean space.

3.2.2 Conformal completion of Minkowski spacetime $\mathbb{R}^{1,d}$

Next, consider the conformal completion of the simplest pseudo-Riemannian spacetime manifold, Minkowski spacetime $\mathbb{R}^{1,d}=(\mathbb{R}^{d+1},\eta)$, where η is the flat metric of signature $(-,+,\ldots,+)$. In the usual rectangular coordinates $\mathbf{x}=(x^0,\ldots,x^d)\in\mathbb{R}^{d+1}$, this metric is

$$\eta=-\left(\mathrm{d}x^0\right)^2+\sum_{i=1}^{d}\left(\mathrm{d}x^i\right)^2.$$

Physically, the coordinate x^0 describes the time component of the spacetime.

We can alternatively use spherical coordinates (i.e. $(d-1)$-spherical coordinates in the language of Appendix A) for the spatial part, and for convenience rename $t := x^0$, so that our coordinates are $(t, r, \phi_{d-1}, \ldots, \phi_1)$ with spatial coordinates given by radius r and angular coordinates ϕ_i. Then the Minkowski metric becomes

$$\eta = -\mathrm{d}t^2 + \mathrm{d}r^2 + r^2 \mathrm{d}\Omega_{d-1}^2,$$

where $\mathrm{d}\Omega_{d-1}^2$ is the metric tensor on the $(d-1)$-sphere as defined by (A.4).

From geometric intuition, we would like to say that the 'infinity' of $\mathbb{R}^{1,d}$ is related to the limits $t \to \pm\infty$ and $r \to \infty$. To make this intuition precise, we 'compactify' these two coordinates, i.e. make transformations that replace them with coordinates which remain finite. By transforming the metric η into such an appropriate coordinate system, we will identify the conformal embedding we need.

Let $u = t + r$, $v = t - r$ so that $u, v \in (-\infty, \infty)$ and $u \geq v$ with $u = v$ if $r = 0$. Then as per transformation law (3.5) for the metric,

$$\eta = -\mathrm{d}u\, \mathrm{d}v + \frac{(u-v)^2}{4} \mathrm{d}\Omega_{d-1}^2.$$

Next, let $\widetilde{u} = \arctan u$, $\widetilde{v} = \arctan v$ so that $\widetilde{u}, \widetilde{v} \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and (since the arctangent function is monotonic) $\widetilde{u} \geq \widetilde{v}$ with $\widetilde{u} = \widetilde{v}$ if $r = 0$ once again. Then

$$\eta = \frac{1}{4} \sec^2 \widetilde{u} \sec^2 \widetilde{v} \left[-4\mathrm{d}\widetilde{u}\, \mathrm{d}\widetilde{v} + \sin^2(\widetilde{u} - \widetilde{v})\, \mathrm{d}\Omega_{d-1}^2 \right].$$

Finally, let $\widetilde{t} = \widetilde{u} + \widetilde{v}$, $\widetilde{r} = \widetilde{u} - \widetilde{v}$, so that the possible values of $\widetilde{t}$ and $\widetilde{r}$ are constrained by the equations $-\pi < \widetilde{t} + \widetilde{r} < \pi$, $-\pi < \widetilde{t} - \widetilde{r} < \pi$ and $\widetilde{r} \geq 0$, where the case $\widetilde{r} = 0$ occurs when $r = 0$. Geometrically, this region is a triangle, as illustrated in Fig. 3.2a. The metric in these coordinates is

$$\eta = \frac{1}{4} \sec^2\left(\tfrac{\widetilde{t}+\widetilde{r}}{2}\right) \sec^2\left(\tfrac{\widetilde{t}-\widetilde{r}}{2}\right) \left[-\mathrm{d}\widetilde{t}^2 + \mathrm{d}\widetilde{r}^2 + \sin^2\widetilde{r}\, \mathrm{d}\Omega_{d-1}^2 \right].$$

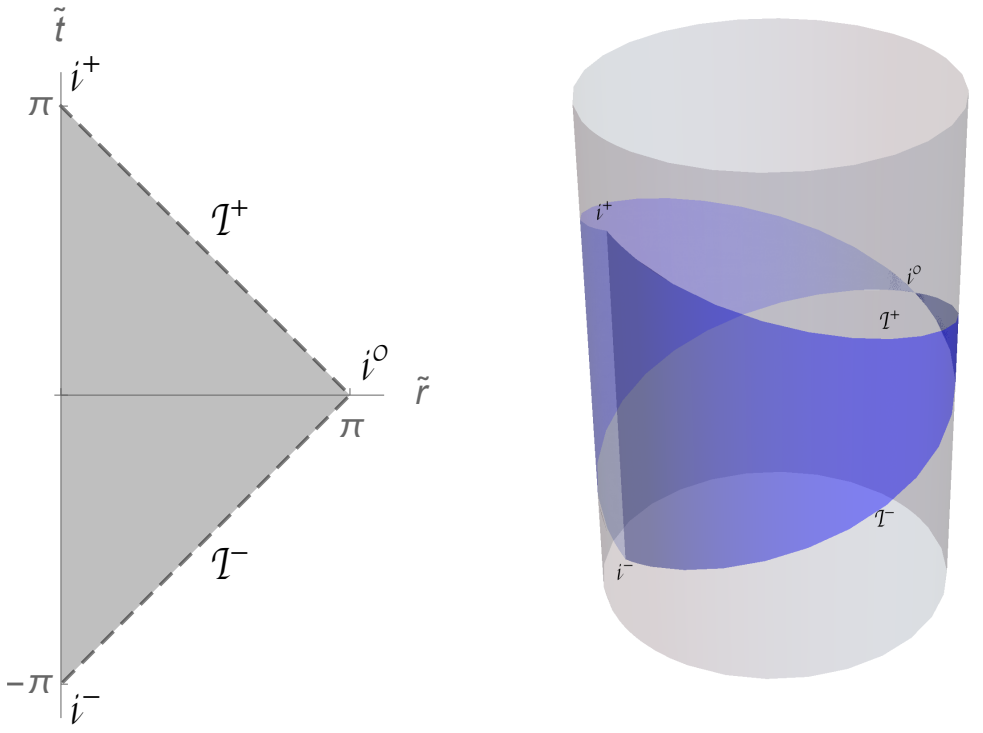
Now, consider a manifold $\mathbb{R} \times \mathbb{S}^d$ with coordinates $(\widehat{t}, \widehat{\phi}_d, \ldots, \widehat{\phi}_1)$ where $\widehat{t}$ is the usual coordinate on the $\mathbb{R}$ component, and $\widehat{\phi}_i$ are angular coordinates on the $\mathbb{S}^d$ component as per the spherical coordinates of Appendix A. In particular, $\widehat{\phi}_1 \in [0, 2\pi)$ and $\widehat{\phi}_i \in [0, \pi]$ for all $2 \leq i \leq d$. Equip this manifold with pseudo-Riemannian metric

$$g_{\mathbb{R}\times\mathbb{S}^d} = -\mathrm{d}\widehat{t}^{\,2} + \mathrm{d}\widehat{\Omega}_d^2.$$

This is the manifold of the so-called Einstein static universe (Aharony, Gubser, et al. 2000).

Since $0 \leq \widetilde{r} < \pi$ always, we can define an injective map $f : \mathbb{R}^{1,d} \to \mathbb{R} \times \mathbb{S}^d$ easily in these coordinates by

$$\left(\widetilde{t}, \widetilde{r}, \phi_{d-1}, \ldots, \phi_1\right) \overset{f}{\mapsto} \left(\widehat{t}, \widehat{\phi}_d, \widehat{\phi}_{d-1}, \ldots, \widehat{\phi}_1\right) = \left(\widetilde{t}, \widetilde{r}, \phi_{d-1}, \ldots, \phi_1\right).$$



(a) Penrose diagram of Minkowski space. i^0, i^+ and i^- are the spatial, future timelike and past timelike infinities, while $\mathcal{I}^+$ and $\mathcal{I}^-$ are the future and past null infinities.

(b) Conformal completion of Minkowski spacetime, by conformal embedding of the Penrose diagram into $\mathbb{R} \times \mathbb{S}^d$. The spatial infinity i^0 is mapped to a single point on $\mathbb{R} \times \mathbb{S}^d$.

Figure 3.2: The Penrose diagram of Minkowski spacetime $\mathbb{R}^{1,d}$, and its conformal embedding into an Einstein static universe $\mathbb{R} \times \mathbb{S}^d$.

Then the calculations above, along with the formula (A.4) for the metric of $\mathbb{S}^d$, show that

$$\begin{aligned} f^* g_{\mathbb{R}\times\mathbb{S}^d} &= f^* \left(-\mathrm{d}\widehat{t}^{\,2} + \mathrm{d}\widehat{\phi}_d^{\,2} + \sin^2 \widehat{\phi}_d \, \mathrm{d}\widehat{\Omega}_{d-1}^2 \right) \\ &= -\mathrm{d}\widetilde{t}^{\,2} + \mathrm{d}\widetilde{r}^{\,2} + \sin^2 \widetilde{r} \mathrm{d}\Omega_{d-1}^2 \\ &= \left[4 \cos^2 \left(\tfrac{\widetilde{t}+\widetilde{r}}{2} \right) \cos^2 \left(\tfrac{\widetilde{t}-\widetilde{r}}{2} \right) \right] \eta, \end{aligned}$$

so f is conformal.

Thus the conformal completion of Minkowski spacetime is the region depicted in Fig. 3.2b – a triangle, *including* its border, 'wrapped' onto $\mathbb{R} \times \mathbb{S}^d$. For lack of a better representation, the figure depicts this as a triangle wrapped onto a cylinder; the axial direction is $\widehat{t}$, and the angle about the cylinder represents $\widehat{\phi}_d$. The triangle is wrapped from $\widehat{\phi}_d = 0$ to π as well as mirrored

around the opposite side of the cylinder. This depicts, schematically at least, the fact that $\widehat{\phi}_d = 0$ and $\widehat{\phi}_d = \pi$ are single points on $\mathbb{S}^d$ (the other angles $\widehat{\phi}_n$ are ill-defined at these points, as per the construction in Appendix A). At other values $0 < \widehat{\phi}_d < \pi$, this degeneracy does not occur; such values are not points but submanifolds within $\mathbb{S}^d$.

The point i^0, which is $\left(\widehat{t}, \widehat{\phi}_d\right) = (0, \pi)$, is the 'image' of the limit $r \to \infty$ for finite t. It is therefore the spatial infinity of Minkowski spacetime. Similarly, the points i^+ and i^- represent future and past timelike infinities respectively, while the lines $\mathcal{I}^+$ and $\mathcal{I}^-$ are the future and past null infinities.

Since the spatial infinity i^0 is at the degenerate point $\widehat{\phi}_d = \pi$ of the spherical coordinates, it is in fact a single point on $\mathbb{R} \times \mathbb{S}^d$. So, much like for the Euclidean space $\mathbb{E}^d$, the conformal completion of Minkowski space $\mathbb{R}^{1,d}$ adds a only single point at the spatial infinity; spatial slices of Minkowski space are compactified to have the same topology as $\mathbb{S}^d$.

More specifically, a $t = 0$ constant time slice of Minkowski spacetime – which looks just like Euclidean space $\mathbb{E}^d$ – gets conformally completed to a sphere $\mathbb{S}^d$, just like $\mathbb{E}^d$ does.

3.2.3 Conformal completion of anti-de Sitter spacetime AdS_{d+2}

Anti-de Sitter spacetime is the maximally symmetric spacetime of negative curvature (Carroll 2004). The $(d+2)$-dimensional anti-de Sitter spacetime AdS_{d+2} can be represented by a quadric Q isometrically embedded in a flat spacetime $\mathbb{R}^{2,d+1}$. Say $\mathbf{X} = \left(X^0, \ldots, X^{d+2}\right)$ are rectangular coordinates in $\mathbb{R}^{2,d+1}$, so that the metric on $\mathbb{R}^{2,d+1}$ is

$$\eta_{2,d+1} = -\left(\mathrm{d}X^0\right)^2 - \left(\mathrm{d}X^{d+2}\right)^2 + \sum_{i=1}^{d+1} \left(\mathrm{d}X^i\right)^2 .$$

Then we define Q as the surface

$$-\left(X^0\right)^2 - \left(X^{d+2}\right)^2 + \sum_{i=1}^{d+1} \left(X^i\right)^2 = -R^2, \tag{3.10}$$

for some $R > 0$.

Let us (somewhat sloppily) identify the flat manifold $\mathbb{R}^{2,d+1}$ with its tangent space at a point. $\mathrm{SO}(p,q)$ is the group of linear maps $\Lambda \in \mathrm{GL}\left(\mathbb{R}^{p,q}\right)$ such that $\det \Lambda = 1$ and $\eta_{p,q}(\Lambda\mathbf{X}, \Lambda\mathbf{Y}) = \eta_{p,q}(\mathbf{X}, \mathbf{Y})$ for all $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{p,q}$. In other words, it is the group of linear isometries on $\mathbb{R}^{p,q}$ with unit determinant.

We see that the surface (3.10) defining Q can be written as

$$\eta_{2,d+1}(\mathbf{X}, \mathbf{X}) = -R^2.$$

This makes it clear that Q is $\mathrm{SO}(2, d+1)$-invariant, i.e. that if $\mathbf{X} \in Q$ then $\Lambda\mathbf{X} \in Q$ for any $\Lambda \in \mathrm{SO}(2, d+1)$.

We call the $\mathbf{X}$ coordinates of AdS the *embedding* coordinates. Let us define a new set of coordinates $\left(\tau, \chi, \lambda^1, \ldots, \lambda^{d+1}\right)$ for this $Q \subseteq \mathbb{R}^{2,d+1}$. Define an angular coordinate τ by

$$\sin\tau = \frac{X^0}{\sqrt{\left(X^0\right)^2 + \left(X^{d+2}\right)^2}},$$

and

$$\cos\tau = \frac{X^{d+2}}{\sqrt{\left(X^0\right)^2 + \left(X^{d+2}\right)^2}}. \tag{3.11}$$

On Q, (3.10) gives that $\left(X^0\right)^2 + \left(X^{d+2}\right)^2 = R^2 + \sum_i \left(X^i\right)^2 \geq R^2$, so that $0 < R/\sqrt{\left(X^0\right)^2 + \left(X^{d+2}\right)^2} \leq 1$. Therefore, we can also define $\chi \in \left[0, \frac{\pi}{2}\right)$ by

$$\cos\chi := \frac{R}{\sqrt{\left(X^0\right)^2 + \left(X^{d+2}\right)^2}}, \tag{3.12}$$

so that $\left(X^0\right)^2 + \left(X^{d+2}\right)^2 = R^2 \sec^2\chi$. Substituting this, (3.10) becomes

$$\sum_{i=1}^{d+1} \left(X^i\right)^2 = R^2 \tan^2\chi.$$

The case $\chi = 0$ corresponds to $X^i = 0$ for all i. Otherwise, set

$$\lambda^i := \frac{X^i}{R\tan\chi}.$$

Then the quadratic defining Q is simply

$$\sum_{i=1}^{d+1} \left(\lambda^i\right)^2 = 1. \tag{3.13}$$

Piecing all this together, we can find the inverse transformation:

$$\begin{aligned} X^0 &= R\sec\chi\sin\tau, & X^{d+2} &= R\sec\chi\cos\tau, \\ X^i &= R\lambda^i\tan\chi & &\text{for all } i \in \{1, \ldots, d+1\}. \end{aligned}$$

This allows us to change the metric on $\mathbb{R}^{2,d+1}$ into the $(\tau, \chi, \lambda^1, \ldots, \lambda^{d+1})$ coordinates:

$$\eta_{2,d+1} = R^2\sec^2\chi\left[-\mathrm{d}\tau^2 + \left(1 + \left[\sum_i \left(\lambda^i\right)^2 - 1\right]\sec^2\chi\right)\mathrm{d}\chi^2 + \left(\sum_i \lambda^i \mathrm{d}\lambda^i\right) 2\tan\chi\mathrm{d}\chi + \sin^2\chi\sum_i \left(\mathrm{d}\lambda^i\right)^2\right].$$

Moreover, if we let $\iota : Q \hookrightarrow \mathbb{R}^{2,d+1}$ denote the natural inclusion map, then from (3.13) we have $\iota^*\left[\sum_i \left(\lambda^i\right)^2\right] = 1$, $\iota^*\left[\sum_i \lambda^i \mathrm{d}\lambda^i\right] = 0$ and $\iota^*\left[\sum_i \left(\mathrm{d}\lambda^i\right)^2\right] = g_{\mathbb{S}^d}$. Therefore, by changing λ^i to spherical coordinates (see Appendix A) we can obtain:

$$g_Q := \iota^* \eta_{2,d+1} = R^2 \sec^2\chi \left[-\mathrm{d}\tau^2 + \mathrm{d}\chi^2 + \sin^2\chi \, \mathrm{d}\Omega_d^2\right].$$

In this quadric, the timelike coordinate τ is 2π-periodic: for fixed χ and λ^i's, the points on Q specified by τ and $\tau + 2\pi$ are identical. Therefore by letting τ vary while holding the other coordinates fixed, we obtain a closed timelike curve on Q. The existence of such curves is undesirable in a model of spacetime.

To obtain what is commonly regarded as anti-de Sitter spacetime, we can 'unwrap' this periodic coordinate (Aharony, Gubser, et al. 2000), i.e. let $\tau \in \mathbb{R}$ take any value without identifying τ and $\tau + 2\pi$. Technically, this means that AdS_{d+2} is the covering space of the quadric Q constructed above.

Therefore, AdS_{d+2} is described by coordinates $\tau \in (-\infty, \infty)$, $\chi \in [0, \frac{\pi}{2})$ and angular coordinates $\boldsymbol{\Omega} = (\phi_d, \ldots, \phi_1)$ where $\phi_1 \in [0, 2\pi)$ while $\phi_i \in [0, \pi]$ for $2 \leq i \leq d$. It inherits the metric of Q:

$$g_{\mathrm{AdS}_{d+2}} = R^2 \sec^2\chi \left[-\mathrm{d}\tau^2 + \mathrm{d}\chi^2 + \sin^2\chi \mathrm{d}\Omega_d^2\right],$$

where once again, $\mathrm{d}\Omega_d^2$ is the metric on the d-sphere in angular coordinates $\boldsymbol{\Omega}$. These coordinates $(\tau, \chi, \boldsymbol{\Omega})$ are called the *global* coordinates of AdS.

Much like the Minkowski spacetime example of Section 3.2.2, these global coordinates suggest the desired conformal completion: we can embed the AdS into an Einstein static universe, $\mathbb{R} \times \mathbb{S}^{d+1}$ with metric $g_{\mathbb{R}\times\mathbb{S}^{d+1}} = -\mathrm{d}\widehat{t}^2 + \mathrm{d}\widehat{\Omega}_{d+1}^2$. In particular, we can use conformal embedding f defined by

$$(\tau, \chi, \phi_d, \ldots, \phi_1) \overset{f}{\mapsto} \left(\widehat{t}, \widehat{\phi}_{d+1}, \widehat{\phi}_d, \ldots, \widehat{\phi}_1\right) = (t, \chi, \phi_d, \ldots, \phi_1)$$

Since $0 \leq \chi < \pi/2$ only, the AdS only covers 'half' of the $\mathbb{R} \times \mathbb{S}^{d+1}$; more precisely, no points in AdS_{d+2} are mapped to the $\widehat{\phi}_{d+1} \geq \pi/2$ region of $\mathbb{R} \times \mathbb{S}^{d+1}$. Note, though, that all the other coordinates have their full ranges covered by the AdS.

The conformal boundary of AdS_{d+2} is then seen to be the submanifold of the $\mathbb{R} \times \mathbb{S}^{d+1}$ given by $\widehat{\phi}_{d+1} = \pi/2$. This extends infinitely in the timelike $\widehat{t}$-direction, but the spatial dimensions are compactified into a $(d+1)$-sphere.

Often the explicit conformal embedding f above is not written; rather than the new coordinates $\widehat{t}$ and $\widehat{\phi}_i$, the coordinates τ, χ and ϕ_i are kept, with the range of χ values extended from $[0, \pi/2)$ to $[0, \pi/2]$ so that $\chi = \pi/2$ is the conformal boundary of AdS in global coordinates.

On a unit sphere $\mathbb{S}^{d+1}$ with angular coordinates $(\phi_{d+1}, \ldots, \phi_1)$ as per Appendix A, it can been seen from (A.3) that the submanifold described by

$\phi_{d+1} = \pi/2$ is just the intersection of the $\mathbb{S}^{d+1}$ with a hyperplane through its centre. This is nothing but a lower-dimensional unit sphere, $\mathbb{S}^d$.

By extension, the submanifold of the $\mathbb{R} \times \mathbb{S}^{d+1}$ described by $\widehat{\phi}_{d+1} = \pi/2$ is exactly the manifold $\mathbb{R} \times \mathbb{S}^d$. Therefore, the conformal boundary $\partial^c(\mathrm{AdS}_{d+2})$ is an Einstein static universe manifold of dimension $d+1$, i.e. $\mathbb{R} \times \mathbb{S}^d$.

For calculations of holographic entanglement entropy, we will limit our attention to the constant-time slice $\tau = 0$ of AdS. On the conformal boundary, this slice is just a cross section at $\widehat{t} = 0$ through the cylinder $\mathbb{R} \times \mathbb{S}^d$. So, the constant-time slice of the conformal boundary of AdS_{d+2} is in fact a sphere $\mathbb{S}^d$.

Poincaré coordinates of AdS

The global coordinate system $(\tau, \chi, \mathbf{\Omega})$ above is all that is needed to demonstrate the properties of the conformal boundary of AdS. However, we take this opportunity to introduce another coordinate system that we will use in calculations.

Let us use X^{d+1} and X^{d+2} of the embedding coordinates to define light-cone coordinates u and v,

$$u = \frac{X^{d+2} - X^{d+1}}{R^2} \qquad \text{and } v = \frac{X^{d+2} + X^{d+1}}{R^2},$$

and rescale the remaining embedding coordinates by u^{-1} (with factors of the AdS radius R for convenience):

$$x^\mu = \frac{X^\mu}{Ru} = \frac{RX^\mu}{X^{d+2} - X^{d+1}} \qquad \text{for } \mu \in \{0, 1, \ldots, d\}\,.$$

Of course, this cannot be done when $u = 0$; therefore, these new coordinates will not cover the entire quadric Q or its covering space AdS, but only a region of it.

Assuming $u \neq 0$, this map can be inverted:

$$\begin{aligned} X^\mu &= Rux^\mu \qquad \text{for } \mu \in \{0, 1, \ldots, d\}\,, \\ X^{d+1} &= \frac{R^2}{2}(v - u), \\ X^{d+2} &= \frac{R^2}{2}(v + u). \end{aligned} \tag{3.14}$$

Let $\eta^{1,d}_{\mu\nu}$ be the usual components of a Minkowski metric tensor with d spatial dimensions, i.e. the components of of a $(d+1) \times (d+1)$ diagonal matrix $\mathrm{diag}(-1, +1, \ldots, +1)$. Then $\eta^{1,d}_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -(\mathrm{d}x^0)^2 + \sum_{i=1}^d (\mathrm{d}x^i)^2$, and we define the shorthand $x^2 := \eta^{1,d}_{\mu\nu} x^\mu x^\nu = -(x^0)^2 + \sum_{i=1}^d (x^i)^2$. Then, after some algebraic manipulation, the metric $\eta_{2,d+1}$ on $\mathbb{R}^{2,d+1}$ can be written as

$$\eta_{2,d+1} = R^2 u^2 \left(\eta^{1,d}_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu\right) + R^2 \left[-R^2 \mathrm{d}v + \mathrm{d}\left(ux^2\right)\right] \mathrm{d}u. \tag{3.15}$$

The equation (3.10) defining the quadric Q can be rewritten in the new coordinates as

$$R^2 uv - u^2 x^2 = 1.$$

From this, we see that Q (or rather, the part of Q covered by the new coordinates) can be written as a graph of the function

$$v(x^\mu, u) = \frac{1}{R^2}\left(\frac{1}{u} + ux^2\right). \tag{3.16}$$

The inclusion map $\iota : Q \hookrightarrow \mathbb{R}^{2,d+1}$ is simply $\iota(x^\mu, u) = (x^\mu, u, v(x^\mu, u))$, so

$$\iota^* \mathrm{d}v = \frac{1}{R^2}\left(-\frac{1}{u^2}\mathrm{d}u + \mathrm{d}(ux^2)\right).$$

With this and (3.15), we find the metric on Q and its covering space AdS in the new coordinates:

$$g_{\mathrm{AdS}_{d+2}} = \iota^* \eta_{2,d+1} = R^2\left[u^2\left(\eta^{1,d}_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu\right) + \frac{1}{u^2}\mathrm{d}u^2\right].$$

Since we have already excluded $u = 0$, we can trivially change from u to $z := 1/u$ to get

$$\begin{aligned} g_{\mathrm{AdS}_{d+2}} &= \frac{R^2}{z^2}\left[\eta^{1,d}_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu + \mathrm{d}z^2\right] \\ &= \frac{R^2}{z^2}\left[-(\mathrm{d}x^0)^2 + \sum_{i=1}^{d}(\mathrm{d}x^i)^2 + \mathrm{d}z^2\right]. \end{aligned} \tag{3.17}$$

These coordinates (x^μ, z) of AdS are called the *Poincaré* coordinates.

While each x^μ can take any real value, the coordinate $z = 1/u$ must be non-zero. Therefore, these coordinates cover two disconnected portions of AdS: one with $z > 0$ and one with $z < 0$. It is common when using Poincaré coordinates to restrict further to only one of these regions, so we hereafter assume $z > 0$.

The conformal boundary lies at $\chi = \pi/2$ in global coordinates. In Poincaré coordinates, (3.14) and (3.16) give

$$X^{d+2} = \frac{1}{2z}\left[x^2 + z^2 + R^2\right],$$

so from (3.12),

$$R^2 \sec^2\chi = (X^0)^2 + (X^{d+2})^2 = \frac{1}{4z^2}\left[(x^2 + z^2 + R^2)^2 + 4R^2(x^0)^2\right].$$

At the conformal boundary $\chi = \pi/2$, the above must be infinite. One way that this can arise is if $z = 0$, so the plane $z = 0$ in Poincaré coordinates forms part of the conformal boundary.

Note that there are other (non-finite) values of the Poincaré coordinates that can correspond to $\chi = \pi/2$ above. For instance, if x^0 is finite then $z \to \infty$ or $x^i \to \infty$ for any $i \in \{1, \ldots, d\}$ also correspond to $\chi = \pi/2$ (Ballón Bayona and Braga 2007). Just as the Poincaré coordinates do not cover the whole of AdS, the $z = 0$ plane of Poincaré coordinates does not cover the whole conformal boundary of AdS.

Consider the constant-time slice $\tau = 0$ in global coordinates. From (3.11), global coordinate τ is zero when embedding coordinate X^0 is zero, which in turn corresponds to Poincaré coordinate $x^0 = 0$. Therefore, the $\tau = 0$ constant-time slice of AdS is conveniently represented in Poincaré coordinates as the $x^0 = 0$ slice.

This constant-time slice is therefore described by $(x^1, \ldots, x^d, z) \in \mathbb{R}^{d+1}$ with $z > 0$ and the following metric tensor, found by pulling (3.17) back onto $x^0 = 0$:

$$g = \frac{R^2}{z^2}\left[\sum_{i=1}^{d}(\mathrm{d}x^i)^2 + \mathrm{d}z^2\right]. \tag{3.18}$$

Such a manifold is well-known to mathematicians as the Poincaré half-space model of $(d+1)$-dimensional hyperbolic space $\mathbb{H}^{d+1}$. For this reason, we label the metric above as $g_{\mathbb{H}^{d+1}}$.

The case of particular interest for us in computations of holographic entanglement entropy will be $d = 2$. In this case, we rename $x^0 = t$, $x^1 = x$ and $x^2 = y$ so that the Poincaré coordinates are (t, x, y, z) with $z > 0$, and the metric is

$$g_{\mathrm{AdS}_4} = \frac{R^2}{z^2}\left[-\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right].$$

The $t = 0$ constant-time slice is the 3-dimensional hyperbolic space with metric

$$g_{\mathbb{H}^3} = \frac{R^2}{z^2}\left[\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right].$$

Chapter 4

Calculations of entanglement entropy using the Ryu-Takayanagi prescription

Having defined what is meant by entanglement entropy, and the boundary at infinity of a spacetime, we can return to the Ryu-Takayanagi formula (1.1) for holographic entanglement entropy in the AdS/CFT correspondence.

To begin, let us restate the Ryu-Takayanagi prescription using the language established in the previous chapters. Say we have a holographic theory with a $(d+2)$-dimensional bulk spacetime $\mathcal{M}$ whose $(d+1)$-dimensional conformal boundary we denote by $\partial^c\mathcal{M}$. As per the holographic duality, there is a conformal field theory defined on $\partial^c\mathcal{M}$. Let $\mathcal{N}$ be a constant-time slice of $\mathcal{M}$, and therefore a $(d+1)$-dimensional spatial submanifold of $\mathcal{M}$. Then the d-dimensional conformal boundary $\partial^c\mathcal{N}$ is the corresponding constant-time slice of $\partial^c\mathcal{M}$

Let $A \subset \partial^c\mathcal{N}$ be a d-dimensional submanifold in the constant-time slice of the boundary. Then the entanglement entropy of region A in the boundary conformal field theory is given by

$$S(A) = \frac{\text{Area}(\gamma_A)}{4G_N^{(d+2)}},$$

where γ_A is the d-dimensional submanifold isometrically embedded in $\mathcal{N}$, with the least possible area such that $\partial^c\gamma_A = \partial A$.

In this chapter, we demonstrate some applications of the Ryu-Takayanagi formula in AdS/CFT, where the bulk spacetime is the 4-dimensional anti-de Sitter space AdS_4. In particular, we find the area-minimising surfaces γ_A and holographic entanglement entropies $S(A)$ associated with some simple regions A on the conformal boundary of AdS_4. We do so using the Poincaré coordinate system described in the previous chapter; recall that the constant-

time slice in this coordinate system is represented as the half-space model of hyperbolic space.

4.1 Volume forms on (pseudo-)Riemannian manifolds

In order to construct surfaces with minimal area in a hyperbolic space, we first need a sufficiently general definition of area. Hyperbolic spaces and submanifolds embedded in them are Riemannian, and therefore have metric tensor g which gives us a notion of distance on the manifold. It is then expected that areas and volumes should also be determined by the metric tensor g.

We also expect the volume of a manifold to be given by an integration over the manifold – after all, upon dividing the manifold up into many smaller parts, the volume of the whole should be the sum of the volumes of the parts, just as in integration. Any lower-dimensional piece of a manifold should not contribute to its volume, just as sets of measure zero do not contribute to integrals.

The only thing we can integrate over an m-dimensional manifold is an m-form. So, we seek an m-form ω on m-dimensional Riemannian manifold $(\mathcal{M}, g)$ that is fully specified by g. Then we can define the volume of $\mathcal{M}$ as $\mathrm{Vol}(\mathcal{M}) = \int_{\mathcal{M}} \omega$.

Definition 4.1 (Riemannian or pseudo-Riemannian volume form)**.** The *(pseudo-)Riemannian volume form* ω_g of (pseudo-)Riemannian manifold $(\mathcal{M}, g)$ of dimension m, in a local coordinate patch $(x^1, x^2, \ldots, x^m)$, is given by the m-form

$$\omega_g = \sqrt{|\det g|}\, \mathrm{d}x^1 \wedge x^2 \wedge \ldots \wedge \mathrm{d}x^m.$$

Here, $\det g$ is the determinant of the matrix $g_{\mu\nu}$ of components representing the 2-form $g = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$ in the chosen coordinates.

While we are able to successfully define many geometric quantities in a coordinate-free manner, we have had to define ω_g in terms of a chosen coordinate patch. We must nonetheless ensure that our definition is independent of the particular choice of coordinates.

Proposition 4.1. *ω_g is invariant under an orientation-preserving change of coordinates $(x_1, \ldots, x_m) \mapsto (y_1, \ldots, y_m)$.*

Proof. In $\mathbf{x}$ and $\mathbf{y}$ coordinates, the metric tensor is given by

$$g = g_{\mu\nu}(\mathbf{x})\mathrm{d}x^\mu \mathrm{d}x^\nu = g'_{\mu\nu}(\mathbf{y})\mathrm{d}y^\mu \mathrm{d}y^\nu,$$

where, as per (3.5),

$$g_{\mu\nu}(\mathbf{x}) = g'_{\sigma\lambda}(\mathbf{y}(\mathbf{x}))\frac{\partial y^\sigma}{\partial x^\mu}\frac{\partial y^\lambda}{\partial x^\nu},$$

From this,

$$\det g = \left(\det g'\right) \left(\det \frac{\partial y}{\partial x}\right)^2,$$

so

$$\sqrt{|\det g|} = \left|\det \frac{\partial y}{\partial x}\right| \sqrt{|\det g'|}.$$

Since $\mathbf{x} \mapsto \mathbf{y}$ is a change of coordinates, it is invertible: the map $\mathbf{y} \mapsto \mathbf{x}$ exists. From this inverse map,

$$\mathrm{d}x^1 \wedge \ldots \wedge \mathrm{d}x^m = \left(\det \frac{\partial x}{\partial y}\right) \mathrm{d}y^1 \wedge \ldots \wedge \mathrm{d}y^m.$$

Altogether, this gives

$$\omega_g = \sqrt{|\det g|}\, \mathrm{d}x^1 \wedge \ldots \wedge \mathrm{d}x^m = \left(\det \frac{\partial x}{\partial y}\right) \left|\det \frac{\partial y}{\partial x}\right| \sqrt{|\det g'|}\, \mathrm{d}y^1 \wedge \ldots \wedge \mathrm{d}y^m.$$

Now, since the matrix inverse of the Jacobian, $(\partial y/\partial x)^{-1}$, is the Jacobian of the inverse map, $\partial x/\partial y$, we note that

$$\left(\det \frac{\partial y}{\partial x}\right) \left|\det \frac{\partial x}{\partial y}\right| = \operatorname{sgn}\left(\det \frac{\partial y}{\partial x}\right).$$

Recall from the discussion at the end of Section 3.1 that a smooth map is orientation-preserving if its Jacobian determinant is positive. The map $\mathbf{x} \mapsto \mathbf{y}$ is orientation-preserving, so $\det(\partial y/\partial x) > 0$. Therefore,

$$\omega_g = \sqrt{|\det g'|}\, \mathrm{d}y^1 \wedge \ldots \wedge \mathrm{d}y^m = \sqrt{|\det g|}\, \mathrm{d}x^1 \wedge \ldots \wedge \mathrm{d}x^m.$$

□

It is interesting to note the necessity of such a coordinate-based, or at least local, definition of ω_g. A manifold with a globally defined, nowhere-zero top-degree form (i.e. a general, not necessarily Riemannian, volume form) is necessarily orientable, and a particular selection of such a volume form in fact gives the manifold an orientation. So if there were a global coordinate-free definition of ω_g that reduced to the above local expression in coordinates, then (since $\omega_g \neq 0$ everywhere for a non-degenerate g) this would imply that $\mathcal{M}$ is oriented. However, not all (pseudo-)Riemannian manifolds are necessarily orientable – for instance, one can isometrically embed a non-orientable Möbius strip into $\mathbb{E}^3$ to obtain a non-orientable Riemannian manifold.

4.2 Explicit calculations of area-minimising surfaces

The Riemannian volume form will allow us to express and compute the area of the minimal surface γ_A of the Ryu-Takayanagi prescription. We now present some of the simpler calculations of area-minimising surfaces employed to find holographic entanglement entropies. These will be limited to cases where the surfaces can be described using a single coordinate chart, and to which the calculus of variations can be applied straightforwardly to determine γ_A.

Let us restrict our geometry of interest to $t = 0$ constant-time slices of AdS_4. Recall from Section 3.2.3 the Poincaré coordinates (t, x, y, z) of AdS_4, with $z > 0$ in the bulk and $z = 0$ denoting the conformal boundary; the metric tensor is given by

$$g_{\mathrm{AdS}_4} = \frac{R^2}{z^2}\left(-\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right),$$

where $R > 0$ is the AdS radius. The $t = 0$ submanifold has induced metric

$$g_{\mathbb{H}^3} = \frac{R^2}{z^2}\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right),$$

equivalent to that of the half-plane model of hyperbolic 3-space.

Now, consider some region A in the $z = 0$ conformal boundary plane. We seek a surface γ_A in the bulk ($z > 0$) region with minimal area subject to the condition that its (conformal) boundary is identical to the boundary of A.

For the simple calculations here, we will assume surface γ_A can be described as the graph of a non-negative function $z(x, y)$ for $(x, y) \in A$, i.e. by points $\big\{(x, y, z(x, y)) \in \mathbb{H}^3 : (x, y) \in A\big\}$. This is not necessarily always possible; for instance, a more complicated surface γ_A may cross particular coordinates (x, y) more than once, so that any such $z(x, y)$ would be multivalued.

We can determine the induced metric on such a graph γ_A by pulling back the metric of AdS_4 (or rather, the slice $\mathbb{H}^3$) onto it. Since $\mathrm{d}z$ is pulled back to $z_x\mathrm{d}x + z_y\mathrm{d}y$, we have induced metric

$$\begin{aligned} g_{\gamma_A} &= \frac{R^2}{z^2}\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \left[z_x\mathrm{d}x + z_y\mathrm{d}y\right]^2\right) \\ &= \frac{R^2}{z^2}\left(\left[1 + z_x^2\right]\mathrm{d}x^2 + \left[1 + z_y^2\right]\mathrm{d}y^2 + 2z_xz_y\mathrm{d}x\mathrm{d}y\right) \\ &= g^{\gamma_A}_{ij}\mathrm{d}x^i\mathrm{d}x^j. \end{aligned}$$

From this, we can find the Riemannian volume form on γ_A,

$$\omega_{\gamma_A} = \sqrt{|\det g^{\gamma_A}|}\,\mathrm{d}x \wedge \mathrm{d}y = \frac{R^2}{z^2}\sqrt{1 + z_x^2 + z_y^2}\,\mathrm{d}x \wedge \mathrm{d}y,$$

and therefore the area of γ_A:

$$\mathrm{Area}(\gamma_A) = \int_A \omega_{\gamma_A} = \int_A \frac{R^2}{z^2}\sqrt{1 + z_x^2 + z_y^2}\,\mathrm{d}x\,\mathrm{d}y. \tag{4.1}$$

For some regions A, it will be more convenient to replace the rectangular Poincaré coordinates (x, y) with polar coordinates (r, θ), as per the usual $(x, y) = r(\cos\theta, \sin\theta)$ prescription. In these coordinates, the AdS metric becomes

$$g_{\mathrm{AdS}_4} = \frac{R^2}{z^2}\left(-\mathrm{d}t^2 + \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + \mathrm{d}z^2\right).$$

As above, we use this to find the induced metric on surface γ_A described by $z(r, \theta)$, and hence the Riemannian volume form ω_{γ_A}; the result is that

$$\mathrm{Area}(\gamma_A) = \int_A \frac{R^2}{z^2}\sqrt{1 + z_r^2 + \tfrac{1}{r^2}z_\theta^2}\, r\mathrm{d}r\,\mathrm{d}\theta. \tag{4.2}$$

To find γ_A, we must find the function $z(x, y)$ or $z(r, \theta)$ that minimises $\mathrm{Area}(\gamma_A)$. In the examples below, we will do so using familiar techniques from variational calculus and classical mechanics. We require that the boundary of γ_A matches the boundary of A. In terms of our chosen coordinates, this means that we require $z(x, y) = 0$ when (x, y) is on the boundary of A, or similar in terms of the polar coordinates (r, θ).

4.2.1 The disc

Consider the example of a disc-shaped region A; in particular, in the polar coordinates above take A to be the region $r \leq \rho$ for some radius ρ on the $z = 0$ plane.

To find the associated function $z(r, \theta)$ describing γ_A, note first that this example possesses rotational symmetry about the centre of the disc A, and so is independent of the angle θ. Therefore, $z = z(r)$ only and the formula (4.2) for the area of the surface γ_A reduces to

$$\mathrm{Area}(\gamma_A) = 2\pi R^2 \int_{r=0}^{\rho} \mathrm{d}r \frac{r}{z^2}\sqrt{1 + (z')^2},$$

where $z'(r) = \frac{\mathrm{d}z}{\mathrm{d}r}$. Let

$$L(r, z, z') := \frac{r}{z^2}\sqrt{1 + (z')^2}.$$

We therefore seek $z(r)$ that minimises

$$\frac{\mathrm{Area}(\gamma_A)}{2\pi R^2} = \int_{r=0}^{\rho} \mathrm{d}r L(r, z, z').$$

This is analogous to the familiar problem of extremising an action constructed from a particular Lagrangian L, where our coordinate r is playing the role of a time coordinate. The extremising solution $z(r)$ satisfies the Euler-Lagrange equation:

$$\frac{\partial L}{\partial z} - \frac{\mathrm{d}}{\mathrm{d}r}\frac{\partial L}{\partial z'} = 0.$$

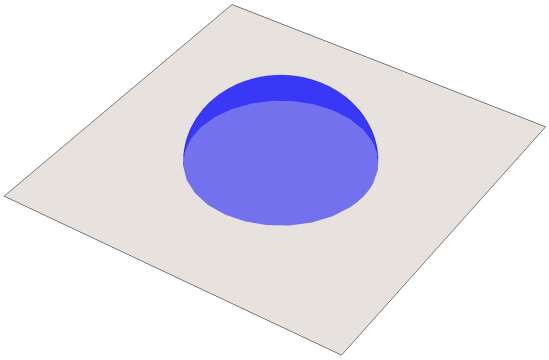


Figure 4.1: Hemispherical minimal surface γ_A associated with a disc A on the conformal boundary $z = 0$ plane in Poincaré coordinates.

After some simplification, this can be written as

$$\frac{z''}{1 + (z')^2} + \frac{z'}{r} + \frac{2}{z} = 0. \tag{4.3}$$

Equation (4.3) applies to all minimal surfaces that can be described as the graph of a function $z(r)$ in these coordinates, so this is applicable to many rotationally-symmetric (θ-independent) examples.

We are looking in particular for the solution whose boundary coincides with that of the disc $r \leq \rho$ at $z = 0$. To obtain this, we must solve (4.3) subject to boundary condition $z(\rho) = 0$. Since (4.3) is a second order ODE, we require another boundary condition; owing to the rotational symmetry, we can impose $z'(0) = 0$.

The solution of (4.3) with these boundary conditions is (Ryu and Takayanagi 2006b)

$$z(r) = \sqrt{\rho^2 - r^2}; \tag{4.4}$$

thus γ_A is a hemisphere of radius ρ centred at $r = 0$, as shown in Fig. 4.1.

Having found the minimal surface, we must compute its area in order to find the entanglement entropy. Of course, this area diverges as the surface γ_A extends towards the conformal boundary. To obtain a non-diverging result, we impose a cut-off at $z = \epsilon \ll 1$: i.e. we only calculate the area of γ_A with $z \geq \epsilon$. This cut-off corresponds to the UV cut-off or lattice spacing cut-off procedures described in Section 2.4; as per that discussion, we expect to obtain an entanglement entropy $S(A)$ parameterised by ϵ that diverges in the limit $\epsilon \to 0$.

Let ρ_* be defined as the value of r when $z = \epsilon$, so $\rho_* = \sqrt{\rho^2 - \epsilon^2}$. The area of γ_A with cut-off ϵ is then

$$\text{Area}(\gamma_A) = 2\pi R^2 \int_{r=0}^{\rho_*} \mathrm{d}r \frac{r}{z^2}\sqrt{1+(z')^2} = 2\pi R^2 \int_{z=\rho}^{\epsilon} \mathrm{d}z \frac{1}{z'}\frac{r}{z^2}\sqrt{1+(z')^2}$$

By rearranging (4.4), we have

$$r = \sqrt{\rho^2 - z^2}, \qquad \text{so} \qquad z' = \frac{-r}{\sqrt{\rho^2 - r^2}} = -\frac{\rho^2 - z^2}{z}.$$

Inserting this into the above, we find

$$\text{Area}(\gamma_A) = 2\pi R^2 \int_{z=\epsilon}^{\rho} \mathrm{d}\left(\frac{z}{\rho}\right)\left(\frac{\rho}{z}\right)^2 = 2\pi R^2 \left(\frac{\rho}{\epsilon} - 1\right).$$

Therefore, the holographic entanglement entropy of the disc of radius ρ is given, as per (1.1), by

$$S(A) = \frac{2\pi R^2}{4G_N^4}\left(\frac{\rho}{\epsilon} - 1\right). \tag{4.5}$$

4.2.2 The infinite strip

As another simple example, take the region A to be an infinitely long strip of width l. In rectangular coordinates, we can describe A as the region at $z = 0$ with $x \in [-l/2, l/2]$.

This geometry also possesses convenient symmetries that we may exploit: it is invariant under translations in the y direction, and under the reflection $x \mapsto -x$. Therefore, we know that $z = z(x)$ only, and this function is even, $z(x) = z(-x)$. From (4.1) we have:

$$\text{Area}(\gamma_A) = \int_A \frac{R^2}{z^2}\sqrt{1+(z')^2}\,\mathrm{d}x\,\mathrm{d}y = R^2 \int_{-\infty}^{\infty}\mathrm{d}y \int_{-l/2}^{l/2}\mathrm{d}x \frac{\sqrt{1+(z')^2}}{z^2},$$

where now $z'(x) = \frac{\mathrm{d}z}{\mathrm{d}x}$.

Clearly, besides the already familiar divergence of the area as γ_A approaches the conformal boundary, the area of this surface also diverges due to its infinite extent in the y-direction. Nonetheless, we can find γ_A such that the area per unit length in the y-direction is minimised.

Let $\lambda(\gamma_A)$ denote this area per unit length of γ_A, and define a function $L(x, z, z') = \frac{1}{z^2}\sqrt{1+(z')^2}$, so that

$$\lambda(\gamma_A) = R^2 \int_{-l/2}^{l/2} \mathrm{d}x L(x, z, z').$$

Again, this is analogous to extremising an action with Lagrangian L and time coordinate x. In this case, L has no explicit dependence on x. We can exploit this to obtain a first-order ODE for z rather than a second-order one. If L plays the role of a Lagrangian where x is a 'time' variable, we can construct the Hamiltonian H associated with L by a Legendre transform:

$$H = z' \frac{\partial L}{\partial z'} - L. \tag{4.6}$$

Then observe that

$$\frac{\mathrm{d}H}{\mathrm{d}x} = z'' \frac{\partial L}{\partial z'} + z' \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial L}{\partial z'} - \left[z' \frac{\partial L}{\partial z} + z'' \frac{\partial L}{\partial z'} + \frac{\partial L}{\partial x} \right] = z' \left[\frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial L}{\partial z'} - \frac{\partial L}{\partial z} \right] - \frac{\partial L}{\partial x}.$$

From the Euler-Lagrange equations, the first term in the above vanishes on extremal solutions for z. So the fact that L has no explicit x-dependence $\partial L/\partial x = 0$ means that

$$\frac{\mathrm{d}H}{\mathrm{d}x} = 0,$$

on extremal solutions.

Therefore, H is independent of x. Evaluating (4.6), we find that

$$H = \frac{-1}{z^2 \sqrt{1 + (z')^2}}.$$

Since $z(x)$ is a smooth even function, $z'(0) = 0$. Let $z_* := z(0)$. Then evaluating H at $x = 0$ gives $H = -1/z_*^2$, and the fact that H is constant in x means

$$\frac{-1}{z_*^2} = \frac{-1}{z^2 \sqrt{1 + (z')^2}},$$

or

$$(z')^2 = \frac{z_*^4 - z^4}{z^4}. \tag{4.7}$$

for all $x \in [-l/2, l/2]$. The boundary condition that this first order differential equation must satisfy for γ_A to meet the boundary of A at infinity is simply $z(l/2) = 0$.

From (4.7), we see that $z' = 0$ only when $z = z_*$. Therefore, if $z' > 0$ for x infinitesimally greater than 0, then z will remain strictly increasing, $z' > 0$, for all $x > 0$. Likewise if $z' < 0$ for x infinitesimally greater than 0, then z will be strictly decreasing, $z' < 0$, for all $x > 0$. Since z must reach 0 at $x = l/2$, we conclude that $z' < 0$ for $x \in (0, l/2)$. Therefore, since $z(x)$ is even, (4.7) gives

$$\frac{\mathrm{d}z}{\mathrm{d}x} = -\mathrm{sgn}(x) \frac{\sqrt{z_*^4 - z^4}}{z^2}. \tag{4.8}$$

We can fix z_* in terms of the width l of the strip:

$$\frac{l}{2} = \int_{x=0}^{l/2} \mathrm{d}x = \int_{z=z_*}^{0} \mathrm{d}z \frac{1}{z'} = \int_{z=0}^{z_*} \mathrm{d}z \frac{z^2}{\sqrt{z_*^4 - z^4}} = \sqrt{\pi} \frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)} z_*,$$

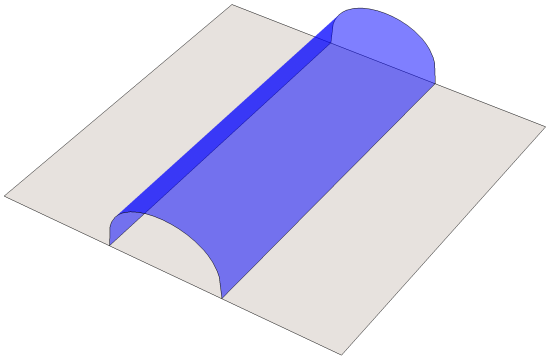


Figure 4.2: Minimal surface γ_A associated with infinite strip A on the conformal boundary $z = 0$ plane in Poincaré coordinates.

so

$$z_* = \frac{\Gamma\left(\frac{1}{4}\right)}{\Gamma\left(\frac{3}{4}\right)} \frac{l}{2\sqrt{\pi}}. \tag{4.9}$$

We can calculate the area per unit length of γ_A without explicitly solving the ODE (4.8):

$$\lambda(\gamma_A) = 2R^2 \int_0^{l/2} \mathrm{d}x \frac{\sqrt{1+(z')^2}}{z^2} = 2R^2 \int_{z_*}^0 \mathrm{d}z \frac{1}{z'} \frac{\sqrt{1+(z')^2}}{z^2}.$$

In fact, this integration must be regulated by a cut-off at $z = \epsilon \ll 1$. Substituting (4.8) into the expression, we find

$$\lambda(\gamma_A) = 2R^2 \int_\epsilon^{z_*} \mathrm{d}z \frac{1}{\sqrt{z_*^4 - z^4}} \left(\frac{z_*}{z}\right)^2 = \frac{2R^2}{z_*} \int_{\epsilon/z_*}^1 \mathrm{d}u \frac{1}{u^2\sqrt{1-u^4}}.$$

The integral can be evaluated in terms of a hypergeometric function,

$$\begin{aligned} \int_a^1 \mathrm{d}u \frac{1}{u^2\sqrt{1-u^4}} &= \left(\frac{1}{a}\right) {}_2F_1\left(-\tfrac{1}{4}, \tfrac{1}{2}; \tfrac{3}{4}; a^4\right) - \sqrt{\pi}\frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)} \\ &= \frac{1}{a} - \sqrt{\pi}\frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)} - \frac{a^3}{6} + \mathcal{O}(a^7). \end{aligned}$$

So with $a = \epsilon/z_*$ and recalling the expression (4.9) for z_*, we reach

$$\lambda(\gamma_A) = R^2 \left\{ \frac{2}{\epsilon} - \left(\frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)}\right)^2 \frac{4\pi}{l} + \mathcal{O}\left(\frac{\epsilon}{l}\right)^3 \right\}.$$

Figure 4.2 depicts the profile of γ_A; from the area per unit length $\lambda(\gamma_A)$, the Ryu-Takayanagi formula gives that the strip A has entanglement entropy

$$S(A) = \frac{LR^2}{4G_N^4}\left\{\frac{2}{\epsilon} - \left(\frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)}\right)^2 \frac{4\pi}{l} + \mathcal{O}\left(\frac{\epsilon}{l}\right)^3\right\}, \tag{4.10}$$

where $L \to \infty$ is the length of the strip.

4.2.3 The annulus

There are other regions A for which the minimal surface γ_A can be found, although they are not necessarily as straightforward as the previous two examples.

Let A be the the annulus given in the polar form of Poincaré coordinates by $\rho_1 \leq r \leq \rho_2$. Dekel and Klose (2013), Drukker and Fiol (2006), Fonda, Giomi, et al. (2015), and Krtouš and Zelnikov (2014) have examined the minimal surface associated with such an annulus, the former two papers in the context of Wilson loops rather than entanglement entropy. The entanglement entropy associated with an annulus is also studied in the context of other field theories and bulk geometries by Nakaguchi and Nishioka (2015).

The annulus, like the disc studied above, has rotational symmetry and thus the surface is θ-independent. The ODE determining $z(r)$ is therefore (4.3), just as for the disc. However, the solution must satisfy different boundary conditions: $z(\rho_1) = z(\rho_2) = 0$. Solving this boundary value problem is complicated by the fact that, for some values of ρ_1 and ρ_2, the solution does not admit a description as a single-valued function $z(r)$.

Note first that a solution composed of two disconnected hemispheres, i.e. two copies of the solution to the disc problem, with radii ρ_1 and ρ_2 respectively will satisfy the minimal surface condition. We call this surface of two disconnected components γ_A^{dis}. Evidently, γ_A^{dis} does not admit a description as the graph of a single valued function $z(r)$. From (4.5),

$$\text{Area}(\gamma_A^{\text{dis}}) = 2\pi R^2\left(\frac{\rho_1 + \rho_2}{\epsilon} - 2\right). \tag{4.11}$$

Define $\eta := \rho_1/\rho_2$. Then for some values of η, other minimal surface solutions for the annulus exist.

Specifically, there are two different hemi-torus ('half-doughnut-shaped') minimal surfaces; both of these being connected, we denote them as $\gamma_A^{\text{con},i}$ for $i = 1, 2$. Still, neither of these admit a description as a single-valued $z(r)$; they 'lean' inward towards $r = 0$, as shown in Fig. 4.3, so that they are multivalued for $r < \rho_1$. As a work-around, Fonda, Giomi, et al. (2015) artificially divide the surface into two parts each admitting a single-valued description in a particular coordinate system, and impose appropriate conditions to join

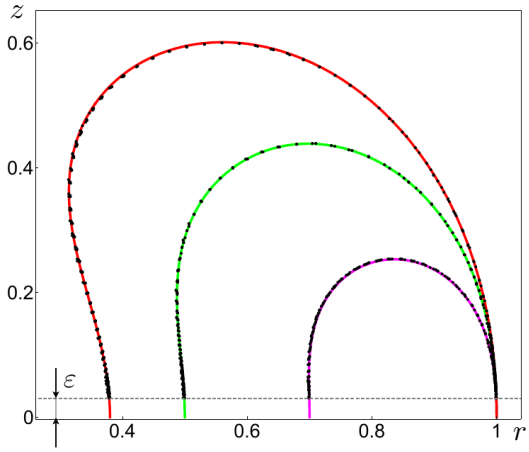


Figure 4.3: Radial profiles of the connected minimal surface solutions for the annulus problem, with three different values of η: 0.38 (red), 0.5 (green) and 0.7 (magenta). Figure reproduced from Fonda, Giomi, et al. (2015). For some r-values, the profile has two corresponding z-values, and so cannot be expressed as a graph $z(r)$.

the two parts. In Chapter 5, we will introduce a coordinate-independent description of locally area-minimising surfaces that is unaffected by such issues.

Fonda, Giomi, et al. (2015) find that when $\eta > \eta_*$ the connected solutions exist and have area given by

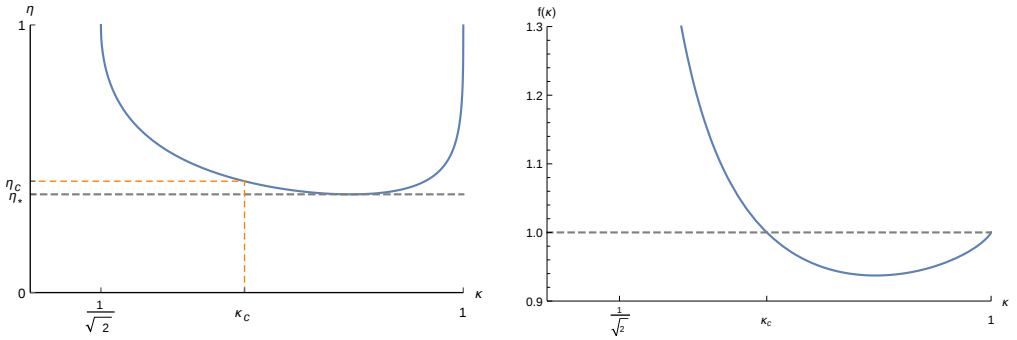
$$\text{Area}(\gamma_A^{\text{con},i}) = 2\pi R^2 \left[\frac{\rho_1 + \rho_2}{\epsilon} - 2f(\kappa_i) + \mathcal{O}(\epsilon) \right], \tag{4.12}$$

where

$$f(\kappa) = \frac{E(\kappa^2) - (1-\kappa^2)K(\kappa^2)}{\sqrt{2\kappa^2 - 1}}$$

and $\kappa_i = \kappa_i(\eta)$ are solutions of

$$\log \eta = 2\kappa \sqrt{\frac{1-2\kappa^2}{\kappa^2 - 1}} \left(K(\kappa^2) - \Pi(1-\kappa^2, \kappa^2) \right). \tag{4.13}$$



(a) Plot of η vs. κ. The value η_* is the minimum η for which a corresponding κ exists.

(b) Plot of f vs. κ. The critical value κ_c marks where $f(\kappa)$ crosses unity.

Figure 4.4: Relationship between the ratio $\eta = \rho_1/\rho_2$ and the finite term f in the area of the connected minimal surface solution for the annulus $\rho_1 < r < \rho_2$.

Here, K, E and Π are the complete elliptic integrals of the first, second and third kind:

$$K(m) := \int_0^{\frac{\pi}{2}} \frac{\mathrm{d}\theta}{\sqrt{1 - m\sin^2\theta}},$$
$$E(m) := \int_0^{\frac{\pi}{2}} \sqrt{1 - m\sin^2\theta}\ \mathrm{d}\theta,$$
$$\Pi(n, m) := \int_0^{\frac{\pi}{2}} \frac{\mathrm{d}\theta}{\left(1 - n\sin^2\theta\right)\sqrt{1 - m\sin^2\theta}}.$$

The value η_* is the smallest for which solutions κ_i in (4.13) exist. Figure 4.4a shows that for $\eta < \eta_*$ no such solutions exist, while for $\eta > \eta_*$ there are two possible solutions, which we label κ_1 and κ_2 with $\kappa_1 < \kappa_2$, corresponding to the two possible connected minimal surfaces $\gamma_A^{\mathrm{con},1}$ and $\gamma_A^{\mathrm{con},2}$.

The multiple solutions for the annulus that exist when $\eta \geq \eta_*$ all *locally* minimise area. The appropriate surface for use in calculating entanglement entropy will have the *globally* minimal area; in other words, of the locally area-minimising solutions, it will be the one with the least area.

To find which of γ_A^{dis} or $\gamma_A^{\mathrm{con},i}$ this is, compare the expressions for their areas (4.11) and (4.12). The connected solutions $\gamma_A^{\mathrm{con},i}$ only have less area than the disconnected γ_A^{dis} when $f(\kappa) > 1$. As Fig. 4.4b illustrates, this only occurs for κ less than some critical $\kappa_c = 0.823$, corresponding to $\eta_c = 0.416$, found numerically. Comparing to Fig. 4.4a we see that this allows only $\gamma_A^{\mathrm{con},1}$ with the smaller value κ_1.

From (1.1), we therefore find the entanglement entropy of the annulus A of inner and outer radii ρ_1 and ρ_2 to be

$$S(A) = \frac{2\pi R^2}{4G_N^4}\left(\frac{\rho_1+\rho_2}{\epsilon} - 2F\left(\frac{\rho_1}{\rho_2}\right) + \mathcal{O}(\epsilon)\right), \tag{4.14}$$

where

$$F(\eta) = \begin{cases} 1 & \text{for } \eta \leq \eta_c, \\ f(\kappa_1(\eta)) & \text{for } \eta > \eta_c, \end{cases} \tag{4.15}$$

for f, κ_1 and η_c as detailed above.

4.3 Discussion

In the preceding chapter, we have seen some particular examples of the computation of entanglement entropy via the Ryu-Takayanagi formula. As a final note, we observe from the solutions (4.5), (4.10) and (4.14) that the leading divergent terms all have the form

$$\frac{1}{4G_N^{(d+2)}}\frac{\text{Area}(\partial A)}{\epsilon^{d-1}}. \tag{4.16}$$

This feature extends to cases with $d \geq 2$, and was observed in CFTs before the Ryu-Takayanagi holographic prescription was formulated (Srednicki 1993).

The choice of regularisation of the entanglement entropy is to some extent arbitrary. We have elected to find the minimal surface γ_A with exact boundary conditions $\partial^c\gamma_A = \partial A$ and then impose the cut-off at $z = \epsilon \ll 1$ in calculating $\text{Area}(\gamma_A)$. However, other choices are also possible; for instance, when working in a different coordinate system one may use a different cut-off adapted to those coordinates. Indeed, although the $z = \epsilon$ cut-off may seem on the surface to be a simple cut-off on distance from the boundary, such a notion requires more care to define in an appropriate coordinate-invariant way (Engelhardt 2017). Another regularisation scheme would be to actually impose the boundary conditions at $z = \epsilon$, as well as only calculating the area up to that z-value. Drukker, Gross, and Ooguri (1999) demonstrate in some explicit examples that using such an alternative regularisation only gives rise to changes at $\mathcal{O}(\epsilon)$ in the surface area of γ_A, and therefore in the entanglement entropy. Since we think of the regularisation in the limit $\epsilon \to 0$, these terms make no difference.

With this in mind, the leading divergent terms (4.16) are effectively independent of the precise choice of regularisation scheme. In the simple cases we have been considering in $d = 2$, the entanglement entropy thus takes the form

$$S(A) = \frac{1}{4G_N^4}\left(\frac{\text{Area}(\partial A)}{\epsilon} + h(A) + \mathcal{O}(\epsilon)\right), \tag{4.17}$$

for some function h of the geometry of A.

Of course in higher dimensions, there can be sub-leading divergent terms $\mathcal{O}(\epsilon^{-n})$ for $0 < n < d-1$. In addition, logarithmically divergent terms $\sim \log[h(A)/\epsilon]$ can arise in some cases, such as for the disc-shaped region A when dimension d is odd (Ryu and Takayanagi 2006a). In still other cases, logarithmically divergent terms arise even for $d = 2$, such as when the shape A has a non-smooth boundary as demonstrated by Casini and M Huerta (2007) in a direct CFT calculation and Hirata and Takayanagi (2007) using the Ryu-Takayanagi prescription.

Chapter 5

Just enough minimal surface theory

In the last chapter, we used techniques from variational calculus to minimise the area functional for a surface γ_A in a bulk manifold associated with a region A on the conformal boundary. These methods require that γ_A be described as the graph of a function; in the Poincaré coordinate system used above, for instance, γ_A was the graph of $z(x, y)$ or $z(r, \theta)$.

As mentioned in the case of the annulus, though, such descriptions are not always straightforward, and may require the surface to be covered by multiple separate coordinate patches. In addition, the total areas of the surfaces found in the previous chapter are all infinite, precisely because they extend to the conformal boundary at which their boundary conditions are prescribed. It is therefore not immediately clear in what sense their areas are minimised: any other, slightly different surface sharing the same boundary conditions would also have infinite total area.

In this chapter, we review some of the mathematical theory of area-minimising submanifolds. Although we do not give a full proof of the characterisation of an area-minimising submanifold, we review the abstract mathematical objects that are involved in such a characterisation: the Second Fundamental Form and the mean curvature. Once these are defined in the abstract context of a general Riemannian manifold, we give their simplifications in the particular context relevant to the Ryu-Takayanagi prescription for entanglement entropy.

5.1 Surfaces embedded in a Riemannian manifold

Let $(\widetilde{\mathcal{M}}, \widetilde{g})$ be a Riemannian manifold with submanifold $(\mathcal{M}, g)$. Specifically, let $\iota : \mathcal{M} \to \widetilde{\mathcal{M}}$ be an isometric embedding of $\mathcal{M}$ into $\widetilde{\mathcal{M}}$. This means that the metric g on $\mathcal{M}$ is identical to that induced on $\mathcal{M}$ by $\widetilde{g}$, i.e. $g = \iota^* \widetilde{g}$. We will throughout this chapter treat $\mathcal{M}$ as a subset of $\widetilde{\mathcal{M}}$; in effect, we identify

$\mathcal{M}$ with $\iota(\mathcal{M})$. Thus when regarding a point $p \in \mathcal{M}$, we actually refer to its image $\iota(p)$; likewise, when regarding a vector $X_p \in T_p\mathcal{M}$, we actually refer to its pushforward $\iota_* X_p$. We will call $\mathcal{M}$ the submanifold, and $\widetilde{\mathcal{M}}$ the ambient manifold.

5.1.1 Vectors tangent and normal to an embedded surface

For each $p \in \mathcal{M}$, we have both the usual tangent space $T_p\mathcal{M}$ of vectors tangent to $\mathcal{M}$ at p, as well the larger ambient tangent space $T_p\widetilde{\mathcal{M}}$ of vectors tangent to $\widetilde{\mathcal{M}}$ at p. In fact, we can view $T_p\mathcal{M}$ as a subspace of $T_p\widetilde{\mathcal{M}}$, and decompose $T_p\widetilde{\mathcal{M}}$ orthogonally into a direct sum:

$$T_p\widetilde{\mathcal{M}} = T_p\mathcal{M} \oplus N_p\mathcal{M}. \tag{5.1}$$

Here, $N_p\mathcal{M} := (T_p\mathcal{M})^\perp$ is the space of ambient tangent vectors normal to $\mathcal{M}$ at p. If $\mathcal{M}$ has dimension m and $\widetilde{\mathcal{M}}$ has dimension $\widetilde{m}$, then the normal vector space $N_p\mathcal{M}$ has dimension $\widetilde{m} - m$.

Let $\pi_p^\top : T_p\widetilde{\mathcal{M}} \to T_p\mathcal{M}$ and $\pi_p^\perp : T_p\widetilde{\mathcal{M}} \to N_p\mathcal{M}$ be the orthogonal projection operators that project onto the different components of the direct sum (5.1). We define a shorthand notation: for ambient tangent vector $X_p \in T_p\widetilde{\mathcal{M}}$, set $X_p^\top := \pi_p^\top X_p \in T_p\mathcal{M}$ and $X_p^\perp := \pi_p^\perp X_p \in N_p\mathcal{M}$. With this, we can always write $X_p = X_p^\top + X_p^\perp$.

In fact, this decomposition into tangent and normal vectors can be extended to the whole bundle of ambient tangent vectors. The disjoint unions

$$T\widetilde{\mathcal{M}}|_{\mathcal{M}} = \coprod_{p\in\mathcal{M}} T_p\widetilde{\mathcal{M}} \qquad \text{and} \qquad N\mathcal{M} = \coprod_{p\in\mathcal{M}} N_p\mathcal{M}$$

are both smooth vector bundles over $\mathcal{M}$; we call them the ambient tangent bundle and normal bundle over $\mathcal{M}$ respectively. Referring to Lee (1997, pg. 132–133) for a more detailed discussion, one can construct in a neighbourhood $\widetilde{U} \subset \widetilde{\mathcal{M}}$ of any $p \in \mathcal{M}$ an adapted orthonormal frame $(E_1, \ldots, E_{\widetilde{m}})$ such that $(E_1, \ldots, E_m)$ can be restricted to $\mathcal{M}$ to give an orthonormal frame for $T\mathcal{M}$. Then $(E_{m+1}, \ldots, E_{\widetilde{m}})$ can be used as a local trivialisation of $N\mathcal{M}$ to demonstrate that it is indeed a smooth bundle.

Define maps $\pi^\top : T\widetilde{\mathcal{M}}|_{\mathcal{M}} \to T\mathcal{M}$ and $\pi^\perp : T\widetilde{\mathcal{M}}|_{\mathcal{M}} \to N\mathcal{M}$ on the ambient tangent bundle to simply act pointwise with the orthogonal projections $\pi_p^\top$ and $\pi_p^\perp$ respectively. In other words, say $(p, X_p) \in T\widetilde{\mathcal{M}}|_{\mathcal{M}}$; then define $\pi^\top(p, X_p) := (p, X_p^\top)$ and $\pi^\perp(p, X_p) = (p, X_p^\perp)$.

For convenience, we denote the space of smooth sections of the bundle $T\mathcal{M}$ as $\mathcal{T}\mathcal{M}$, so if $X \in \mathcal{T}\mathcal{M}$ then X is a smooth tangent vector field on $\mathcal{M}$. Likewise, denote the spaces of smooth sections of $T\widetilde{\mathcal{M}}$, $T\widetilde{\mathcal{M}}|_{\mathcal{M}}$ and $N\mathcal{M}$ as $\mathcal{T}\widetilde{\mathcal{M}}$, $\mathcal{T}\widetilde{\mathcal{M}}|_{\mathcal{M}}$ and $\mathcal{N}\mathcal{M}$ respectively.

Say $X \in \mathcal{T}\widetilde{\mathcal{M}}|_{\mathcal{M}}$, i.e. X is an ambient tangent vector field on $\mathcal{M}$. Then the above projection operators can be applied to give $X^\top := \pi^\top \circ X \in \mathcal{T}\mathcal{M}$ and $X^\perp := \pi^\perp \circ X \in \mathcal{N}\mathcal{M}$, with $X = X^\top + X^\perp$. By expressing $\pi^\top$ and $\pi^\perp$ locally in an adapted orthonormal frame, one can show them to map smooth sections to smooth sections (Lee 1997, pg. 133).

5.1.2 On the locality of covariant derivatives

Let ∇ and $\widetilde{\nabla}$ denote the Riemannian (i.e. Levi-Civita) connections on $(\mathcal{M}, g)$ and $(\widetilde{\mathcal{M}}, \widetilde{g})$ respectively. By definition, $\widetilde{\nabla}$ is a map $\widetilde{\nabla} : \mathcal{T}\widetilde{\mathcal{M}} \times \mathcal{T}\widetilde{\mathcal{M}} \to \mathcal{T}\widetilde{\mathcal{M}}$ so for any sections $\widetilde{X}, \widetilde{Y} \in \mathcal{T}\widetilde{\mathcal{M}}$, the covariant derivative $\widetilde{\nabla}_{\widetilde{X}}\widetilde{Y}$ is again a section in $\mathcal{T}\widetilde{\mathcal{M}}$

Recall that, as expected of a derivative, $(\widetilde{\nabla}_{\widetilde{X}}\widetilde{Y})|_p$ depends only on the value of $\widetilde{X}$ at p and the values of $\widetilde{Y}$ on a curve $\gamma : (-\epsilon, \epsilon) \to \widetilde{\mathcal{M}}$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = X_p$, where ϵ can be arbitrarily small (Lawson 1978).

Therefore, for sections $X, Y \in \mathcal{T}\mathcal{M}$ of the submanifold's tangent bundle $T\mathcal{M}$, the covariant derivative $\widetilde{\nabla}_X Y$ with respect to the *ambient* Riemannian connection is well-defined everywhere on $\mathcal{M}$ even though X and Y are not ambient vector fields. We can express this more technically as follows. Let $\widetilde{X}, \widetilde{Y} \in \mathcal{T}\widetilde{\mathcal{M}}$ and $\widetilde{X}', \widetilde{Y}' \in \mathcal{T}\widetilde{\mathcal{M}}$ be arbitrary extensions of vector fields X, Y to $T\widetilde{\mathcal{M}}$, so that upon restricting to the submanifold, $\widetilde{X}|_{\mathcal{M}} = X = \widetilde{X}'|_{\mathcal{M}}$ and likewise for Y. Then

$$\left(\widetilde{\nabla}_{\widetilde{X}}\widetilde{Y}\right)_p = \left(\widetilde{\nabla}_{\widetilde{X}'}\widetilde{Y}'\right)_p .$$

for all $p \in \mathcal{M}$. This means that the value of this covariant derivative on $\mathcal{M}$ is independent of the particular choice of extension of X and Y to the ambient manifold. Therefore, we can sensibly define $\widetilde{\nabla}_X Y$ as $\widetilde{\nabla}_{\widetilde{X}}\widetilde{Y}\big|_{\mathcal{M}}$ for *any* arbitrary extensions $\widetilde{X}$ and $\widetilde{Y}$ to the ambient manifold. In fact, we can extend this slightly further to allow Y to be any *ambient* tangent vector field: for any $X \in \mathcal{T}\mathcal{M}$, the quantity $\widetilde{\nabla}_X Y$ is a well-defined section in $\mathcal{T}\widetilde{\mathcal{M}}|_{\mathcal{M}}$ for any $Y \in \mathcal{T}\widetilde{\mathcal{M}}|_{\mathcal{M}}$.

5.2 The Second Fundamental Form, mean curvature and minimal surfaces

With the brief reminders above, we can now begin to define in abstract the Second Fundamental Form and the associated mean curvature; these are the quantities that characterise area-minimising surfaces in a coordinate-independent way. While the abstract definitions we present in Section 5.2.1 provide a good structural understanding of these quantities, they are not trivial to relate to practical calculations. In Sections 5.2.2 and 5.2.3 we specialise to cases that permit expressions for the Second Fundamental Form and mean curvature which are more easily related to direct calculations.

5.2.1 General definitions

Take ambient and embedded Riemannian manifolds $(\widetilde{\mathcal{M}}, \widetilde{g})$ and $(\mathcal{M}, g)$ as above, and let $X, Y \in \mathcal{T}\mathcal{M}$ be vector fields tangent to the embedded submanifold.

The covariant derivative $\widetilde{\nabla}_X Y$ is a section of $T\widetilde{\mathcal{M}}|_{\mathcal{M}}$, so as per Section 5.1.1 we may decompose it into sections normal and tangent to $\mathcal{M}$:

$$\widetilde{\nabla}_X Y = (\widetilde{\nabla}_X Y)^\top + (\widetilde{\nabla}_X Y)^\perp.$$

Lemma 5.1. *With sections $X, Y \in \mathcal{T}\mathcal{M}$ as above,*

$$\nabla_X Y = (\widetilde{\nabla}_X Y)^\top.$$

Proof. We follow the argument of Lee (1997, Lemma 5.1 and Theorem 8.2). Since $(\widetilde{\nabla}_X Y)^\top = \pi^\top \circ \widetilde{\nabla}_X Y$, we must show that $\pi^\top \circ \widetilde{\nabla} = \nabla$. For convenience of notation, define $\widetilde{\nabla}^\top := \pi^\top \circ \widetilde{\nabla}$.

Recall that the Riemannian connection ∇ on $(\mathcal{M}, g)$ is unique. Therefore, if we can show that $\widetilde{\nabla}^\top$ is a Riemannian connection on $(\mathcal{M}, g)$, then necessarily $\widetilde{\nabla}^\top = \nabla$.

Let us first show that $\widetilde{\nabla}^\top$ is indeed a connection on $\mathcal{M}$. As discussed in Section 5.1.2, the operator $\widetilde{\nabla}$ can unambiguously take arguments from $\mathcal{T}\mathcal{M}$, so that $\widetilde{\nabla}^\top : \mathcal{T}\mathcal{M} \times \mathcal{T}\mathcal{M} \to \mathcal{T}\mathcal{M}$ is well-defined. Since both $\widetilde{\nabla}$ and $\pi^\top$ map smooth sections to smooth sections, so does $\widetilde{\nabla}^\top$.

Additionally, $\widetilde{\nabla}_X Y$ is $\mathbb{R}$-linear in Y and $C^\infty(\widetilde{\mathcal{M}})$-linear in X, and the projection $\pi^\top$ is $C^\infty(\mathcal{M})$-linear, so $\widetilde{\nabla}^\top_X Y$ is $\mathbb{R}$-linear in Y and $C^\infty(\mathcal{M})$-linear in X.

To show that $\widetilde{\nabla}^\top$ is a connection, it remains only to show that it obeys the product rule. Take $f \in C^\infty(\mathcal{M})$, and extend it arbitrarily to a smooth function on $\widetilde{\mathcal{M}}$, just as we did with vector fields X, Y in Section 5.1.2. Then from the product rule of $\widetilde{\nabla}$,

$$\begin{aligned}
\widetilde{\nabla}^\top_X (fY) &= \pi^\top \circ \left[f\ \widetilde{\nabla}_X Y + (Xf) Y \right] \\
&= f\ \pi^\top \circ \widetilde{\nabla}_X Y + (Xf)\ \pi^\top \circ Y \\
&= f \widetilde{\nabla}^\top_X Y + (Xf) Y,
\end{aligned}$$

as needed. Implicitly, we have used that Xf does not depend on the extension of f to $\widetilde{\mathcal{M}}$, since X is tangent to $\mathcal{M}$; this is similar to the independence of $\widetilde{\nabla}_X Y$ on the particular extension of Y to $\mathcal{T}\widetilde{\mathcal{M}}|_{\mathcal{M}}$.

To show that $\widetilde{\nabla}^\top$ is a Riemannian connection in particular, we must show that it is compatible with the metric g and is torsion-free (i.e. symmetric).

The torsion-free property of $\widetilde{\nabla}$ gives that

$$\widetilde{\nabla}_X Y - \widetilde{\nabla}_Y X = [X, Y].$$

From this, it follows that

$$\widetilde{\nabla}^{\top}_X Y - \widetilde{\nabla}^{\top}_Y X = \pi^{\top} \circ (\widetilde{\nabla}_X Y - \widetilde{\nabla}_Y X) = [X, Y]^{\top}.$$

However, since X and Y are sections in $\mathcal{T}\mathcal{M}$, so too is $[X, Y]$; therefore, $[X, Y]^{\top} = [X, Y]$. Thus $\widetilde{\nabla}^{\top}$ is torsion-free.

It remains only to prove that $\widetilde{\nabla}^{\top}$ is compatible with the metric g. The $\widetilde{g}$-compatibility of $\widetilde{\nabla}$ gives that, for any $X, Y, Z \in \mathcal{T}\mathcal{M}$,

$$X\ \widetilde{g}(Y, Z) = \widetilde{g}(\widetilde{\nabla}_X Y, Z) + \widetilde{g}(Y, \widetilde{\nabla}_X Z).$$

Since Z is tangent to $\mathcal{M}$, $\widetilde{g}(\widetilde{\nabla}_X Y, Z) = \widetilde{g}((\widetilde{\nabla}_X Y)^{\top}, Z) = \widetilde{g}(\widetilde{\nabla}^{\top}_X Y, Z)$. The same holds for Y, so

$$X\ \widetilde{g}(Y, Z) = \widetilde{g}(\widetilde{\nabla}^{\top}_X Y, Z) + \widetilde{g}(Y, \widetilde{\nabla}^{\top}_X Z)$$

Now recall that when the arguments of $\widetilde{g}$ are tangent to $\mathcal{M}$, then we can replace $\widetilde{g}$ with g, since $\mathcal{M}$ is isometrically embedded in $\widetilde{\mathcal{M}}$. Thus we have

$$X\ g(Y, Z) = g(\widetilde{\nabla}^{\top}_X Y, Z) + g(Y, \widetilde{\nabla}^{\top}_X Z),$$

so $\widetilde{\nabla}^{\top}$ is compatible with the metric g. Therefore, it is the Riemannian connection associated with g on $\mathcal{M}$. □

We are now equipped to define in abstract an important object in the study of Riemannian submanifolds:

Definition 5.1 (Second Fundamental Form)**.** For submanifold $(\mathcal{M}, g)$ in ambient Riemannian manifold $(\widetilde{\mathcal{M}}, \widetilde{g})$, the *Second Fundamental Form* is the map $I\!I : \mathcal{T}\mathcal{M} \times \mathcal{T}\mathcal{M} \to \mathcal{N}\mathcal{M}$, defined by

$$I\!I(X, Y) = (\widetilde{\nabla}_X Y)^{\perp} = \widetilde{\nabla}_X Y - \nabla_X Y,$$

for any $X, Y \in \mathcal{T}\mathcal{M}$

Lemma 5.2. *$I\!I$ is symmetric, so $I\!I(X, Y) = I\!I(Y, X)$ for all $X, Y \in \mathcal{T}\mathcal{M}$.*

Proof. Using the torsion-free property $\widetilde{\nabla}_X Y - \widetilde{\nabla}_Y X = [X, Y]$ of the ambient Riemannian connection $\widetilde{\nabla}$,

$$\begin{aligned} I\!I(X, Y) - I\!I(Y, X) &= (\widetilde{\nabla}_X Y)^{\perp} - (\widetilde{\nabla}_Y X)^{\perp} \\ &= \pi^{\perp} \circ (\widetilde{\nabla}_X Y - \widetilde{\nabla}_Y X) \\ &= \pi^{\perp} \circ [X, Y] = [X, Y]^{\perp}. \end{aligned}$$

Since X, Y are both sections of $T\mathcal{M}$, the Lie bracket $[X, Y]$ is also a section of $T\mathcal{M}$. Therefore, $[X, Y]$ is purely tangential to $\mathcal{M}$ and so has no component normal to the submanifold: $[X, Y]^{\perp} = 0$. □

Recalling the discussion on the locality of the covariant derivative in Section 5.1.2, it is clear that $I\!I(X,Y)|_p$ depends on the value of X only at p and nowhere else, for any $p \in \mathcal{M}$. Additionally, since $\widetilde{\nabla}_X Y$ is $C^\infty(\mathcal{M})$-linear in X, so is $I\!I(X,Y)$. But now by the symmetry property of Lemma 5.2, these same facts must be true of the dependence of $I\!I(X,Y)$ on Y.

So $I\!I(X,Y)$ is symmetric and $C^\infty(\mathcal{M})$-linear in both its arguments, and $I\!I(X,Y)|_p$ depends only on X_p and Y_p for any $p \in \mathcal{M}$. These properties are exactly those that characterise a symmetric 2-form, i.e. symmetric $(0,2)$-tensor field. Therefore the Second Fundamental Form is, as the name suggests, a symmetric normal-vector-valued 2-form.

At an intuitive level, Definition 5.1 tells us that Second Fundamental Form describes the difference between the geometry of the submanifold by itself and the geometry it has once embedded into the ambient manifold. Therefore, the Second Fundamental Form tells us about the extrinsic geometric properties of the submanifold, i.e. those geometric properties that it possesses due to its embedding in the ambient manifold. As such, it is the object that will allow us to determine whether an embedded submanifold has (locally) minimal area, in a suitably general and coordinate-independent way.

Definition 5.2 (Mean curvature vector)**.** The *mean curvature vector at a point* $p \in \mathcal{M}$ is the normal vector

$$K_p = \frac{1}{m} \mathrm{tr}_g I\!I_p,$$

where m is the dimension of the embedded submanifold.

The *mean curvature vector field* K is the field of normal vectors (i.e. the section of $N\mathcal{M}$) which takes value K_p at each $p \in \mathcal{M}$.

This definition at last allows a general, coordinate-free characterisation of area-minimising submanifolds: a submanifold with $K = 0$ everywhere has locally minimal area. This means that, if we pick any small patch on the submanifold and perturb the submanifold smoothly only on that patch, then all such perturbations increase the area of that patch and thus the total area of the submanifold.

We do not prove this core result here, but rather refer the reader to Lawson (1978, Chapter 1) and restrict ourselves to understanding the mean curvature quantity, from its abstract definition above to the reformulations below that can be applied in direct calculations.

A submanifold with zero mean curvature is called a *minimal* submanifold, precisely because it has locally minimum area. Thus the surfaces of interest in the Ryu-Takayanagi description of entanglement area are minimal surfaces. The zero mean curvature characterisation clarifies how the surfaces γ_A in the Ryu-Takayanagi formula, which have infinite total area, can be regarded as area-minimising: perturbations of the surface in any small, local patch of the surface will increase the area of that patch.

5.2.2 Defining mean curvature relative to a particular unit normal field

We will find it useful below to work with a scalar-valued version of the Second Fundamental Form, rather than the normal-vector-valued $I\!I$ above. To accomplish this, we must first choose particular normal vector field $N \in \mathcal{NM}$ that has unit norm at all points: $|N_p|_{\widetilde{g}} = 1$ for all $p \in \mathcal{M}$. In fact, such a field can only be defined globally if $\mathcal{M}$ is orientable. In the case that $\mathcal{M}$ is not orientable, we should treat the following statements as local, since we can always restrict to orientable neighbourhoods.

Definition 5.3. Say $N \in \mathcal{NM}$ is a normal vector field with unit norm. We define $h_N : \mathcal{TM} \times \mathcal{TM} \to \mathbb{R}$ by

$$h_N(X, Y) = \widetilde{g}\left(I\!I(X, Y), N\right),$$

for any $X, Y \in \mathcal{TM}$.

From the properties of $I\!I$ it follows that h_N is a smooth scalar-valued $C^\infty(\mathcal{M})$-bilinear symmetric form, i.e. a symmetric $(0, 2)$-tensor field, on $\mathcal{M}$. An alternative expression, sometimes more useful for explicit calculation, can be found for $h_N(X, Y)$ as follows.

Lemma 5.3 (The Weingarten Equation)**.** *For all $X, Y \in \mathcal{TM}$ and $N \in \mathcal{NM}$ with unit norm,*

$$h_N(X, Y) = \widetilde{g}(-\widetilde{\nabla}_X N, Y).$$

Proof. Since $Y \in \mathcal{TM}$ and $N \in \mathcal{NM}$, we know that $\widetilde{g}(N, Y) = 0$ everywhere on $\mathcal{M}$. Therefore, from the $\widetilde{g}$-compatibility of $\widetilde{\nabla}$, we have:

$$\begin{aligned} 0 = X\ \widetilde{g}(N, Y) &= \widetilde{g}(\widetilde{\nabla}_X N, Y) + \widetilde{g}(N, \widetilde{\nabla}_X Y) \\ &= \widetilde{g}(\widetilde{\nabla}_X N, Y) + \widetilde{g}(N, I\!I(X, Y)) + \widetilde{g}(N, \nabla_X Y). \end{aligned}$$

The last equality uses Definition 5.1. Since $\nabla_X Y$ is tangent to $\mathcal{M}$, the last term above vanishes, leaving us with

$$\widetilde{g}(N, I\!I(X, Y)) = \widetilde{g}(-\widetilde{\nabla}_X N, Y).$$

□

Using the Weingarten equation of Lemma 5.3, and recalling that Y is tangent to $\mathcal{M}$, we see that

$$h_N(X, Y) = \widetilde{g}(-(\widetilde{\nabla}_X N)^\top - (\widetilde{\nabla}_X N)^\perp, Y) = \widetilde{g}(-(\widetilde{\nabla}_X N)^\top, Y),$$

for all $X, Y \in \mathcal{TM}$. Since both Y and $(\widetilde{\nabla}_X N)^\top$ are vector fields tangent to $\mathcal{M}$, and since $\mathcal{M}$ is isometrically embedded in $\widetilde{\mathcal{M}}$,

$$h_N(X, Y) = \widetilde{g}(-(\widetilde{\nabla}_X N)^\top, Y) = g(-(\widetilde{\nabla}_X N)^\top, Y), \tag{5.2}$$

for all $X, Y \in \mathcal{TM}$.

Now recall the coordinate-free definition of 'raising an index' of a tensor, via the musical isomorphism $^\sharp$ (Lee 2013, p. 342): given a covector $\omega \in T_p^*\mathcal{M}$ at a point p, we can define a vector $\omega^\sharp \in T_p\mathcal{M}$ by

$$g(\omega^\sharp, Y) := \omega(Y), \qquad \text{for all } Y \in T_p\mathcal{M}.$$

In coordinates, this is the usual raising of an index, so $\omega^\sharp = \omega^i \partial_i$ with $\omega^i = g^{ij}\omega_j$.

If we raise one index of the $(0,2)$ tensor field h_N on $\mathcal{M}$, we obtain a $(1,1)$-tensor field $h_N^\sharp$ such that

$$h_N(X,Y) = g\left(h_N^\sharp(X), Y\right) \tag{5.3}$$

for all vector fields $X, Y \in \mathcal{TM}$. Note that since h_N is a tensor field on $\mathcal{M}$ rather than on $\widetilde{\mathcal{M}}$, we must use the metric $g = \iota^*\widetilde{g}$ to raise an index.

Comparing (5.2) and (5.3), we find that

$$g\left(h_N^\sharp(X), Y\right) = g\left(-(\widetilde{\nabla}_X N)^\top, Y\right),$$

for all $X, Y \in \mathcal{TM}$. This leads us to a corollary of the Weingarten equation, Lemma 5.3.

Corollary 5.4. *For $X \in \mathcal{TM}$ and $N \in \mathcal{NM}$ any unit normal vector field of $\mathcal{M}$ embedded in $\widetilde{\mathcal{M}}$,*

$$h_N^\sharp(X) = -(\widetilde{\nabla}_X N)^\top.$$

The map $h_N^\sharp : \mathcal{TM} \to \mathcal{TM}$ is often named the *shape operator*; Corollary 5.4 gives a convenient way of directly calculating it given the appropriate unit normal field N. With the shape operator, we can define a scalar alternative to the mean curvature vector K on $\mathcal{M}$.

Definition 5.4. The *mean curvature scalar field*, or simply *mean curvature*, of m-dimensional submanifold $(\mathcal{M}, g)$ in $(\widetilde{\mathcal{M}}, \widetilde{g})$ with respect to unit normal field N, is

$$H_N = \frac{1}{m}\mathrm{tr}_g\left(h_N\right) = \frac{1}{m}\mathrm{tr}\left(h_N^\sharp\right).$$

5.2.3 Submanifolds of codimension 1

Consider now the specific case where $\mathcal{M}$ is of codimension 1 in $\widetilde{\mathcal{M}}$, or in other words $\dim \mathcal{M} = \dim \widetilde{\mathcal{M}} - 1$. Then the space of vectors normal to $\mathcal{M}$ is one-dimensional. Therefore, the unit normal vector field N is uniquely defined up to a sign on each connected component of $\mathcal{M}$, and in addition N_p spans $N_p\mathcal{M}$ at each point $p \in \mathcal{M}$. From this, we know that the vector-valued

Second Fundamental Form is proportional to N. It follows from Definition 5.3 that

$$I\!I(X, Y) = h_N(X, Y)N,$$

and hence that $K = H_N N$ as well. The zero mean curvature condition for a minimal surface can therefore be rephrased: $\mathcal{M}$ is minimal if and only if $H_N = 0$.

Additionally, the fact that N spans the normal vector space at each point gives us a way to explicitly construct the projection operators $\pi^\top$ and $\pi^\perp$ from N: for any $X \in \mathcal{T}\widetilde{\mathcal{M}}|_{\mathcal{M}}$,

$$\pi^\perp(X) = \widetilde{g}(X, N)N$$

and

$$\pi^\top(X) = X - \pi^\perp(X) = X - \widetilde{g}(X, N)N.$$

Using these expressions, we can find simplified forms of the shape operator $h_N^\sharp$ and in turn the mean curvature scalar H_N.

Lemma 5.5. *For submanifold $\mathcal{M}$ of codimension 1, isometrically embedded in $(\widetilde{\mathcal{M}}, \widetilde{g})$ with unit normal vector field N, the shape operator $h_N^\sharp$ can be written as*

$$h_N^\sharp(X) = -\widetilde{\nabla}_X N.$$

Proof. Using the explicit form of the projection operator $\pi^\perp$ above, the normal component of vector $\widetilde{\nabla}_X N$ for $X \in \mathcal{T}\mathcal{M}$ takes the form

$$(\widetilde{\nabla}_X N)^\perp = \widetilde{g}(\widetilde{\nabla}_X N, N)N.$$

But using the compatibility of the connection $\widetilde{\nabla}$ with the metric $\widetilde{g}$,

$$\widetilde{g}(\widetilde{\nabla}_X N, N) = \frac{1}{2}\left(\widetilde{g}(\widetilde{\nabla}_X N, N) + \widetilde{g}(N, \widetilde{\nabla}_X N)\right) = \frac{1}{2} X\, \widetilde{g}(N, N) = 0,$$

since $\widetilde{g}(N, N) = 1$ is constant. Thus $(\widetilde{\nabla}_X N)^\perp = 0$. From Corollary 5.4, the shape operator is then

$$h_N^\sharp(X) = -(\widetilde{\nabla}_X N)^\top = -\widetilde{\nabla}_X N + (\widetilde{\nabla}_X N)^\perp = -\widetilde{\nabla}_X N.$$

□

Recall the coordinate-free definition of the divergence operator on a Riemannian manifold $\widetilde{\mathcal{M}}$ (Carmo 1992, p. 83): for a vector field $Y \in \mathcal{T}\widetilde{\mathcal{M}}$, the divergence of Y is defined as the trace of the map $X \mapsto \widetilde{\nabla}_X Y$:

$$\operatorname{div}_{\widetilde{\mathcal{M}}} Y = \operatorname{tr}(X \mapsto \widetilde{\nabla}_X Y).$$

With this and Lemma 5.5, Definition 5.4 of the mean curvature scalar H_N simplifies remarkably for codimension 1 submanifolds.

Theorem 5.6. *For m-dimensional submanifold $\mathcal{M}$ isometrically embedded in $(m+1)$-dimensional Riemannian manifold $(\widetilde{\mathcal{M}}, \widetilde{g})$ with unit normal vector field N, the mean curvature scalar field can be written as*

$$H_N = -\frac{1}{m}\mathrm{div}_{\widetilde{\mathcal{M}}} N,$$

where $\mathrm{div}_{\widetilde{\mathcal{M}}}$ is the divergence operator in $\widetilde{\mathcal{M}}$.

Proof. Choose a local, adapted orthonormal frame $(E_1, \ldots, E_m, E_{m+1})$ in a neighbourhood $\widetilde{U} \subseteq \widetilde{\mathcal{M}}$ of a point p on the submanifold $\mathcal{M}$, so that $(E_1, \ldots, E_m)$ is a local orthonormal frame of $\mathcal{M}$ in $\widetilde{U} \cap \mathcal{M}$, and $E_{m+1} = N$ is the normal unit vector field. The associated dual frame is $(\phi_1, \ldots, \phi_{m+1})$ where $\phi_i(Y) = \widetilde{g}(E_i, Y)$ for any vector field $Y \in \mathcal{T}\widetilde{U}$.

Using this frame, the shape operator can be written explicitly in U as

$$h_N^\sharp = \sum_{i=1}^{m} h_N^\sharp(E_i)\phi_i = \sum_{i=1}^{m}(-\widetilde{\nabla}_{E_i} N)\phi_i.$$

Since the shape operator is a map on the space of tangent vector fields $\mathcal{T}\mathcal{M}$, its trace is

$$\mathrm{tr}(h_N^\sharp) = \sum_{i=1}^{m} \phi_i(h_N^\sharp(E_i)) = \sum_{i=1}^{m} \widetilde{g}(-\widetilde{\nabla}_{E_i} N, E_i),$$

since $(E_1, \ldots, E_m)$ locally spans the tangent vector fields.

Much as was found in the proof of Lemma 5.5,

$$\widetilde{g}(-\widetilde{\nabla}_N N, N) = -\frac{1}{2} N\widetilde{g}(N, N) = 0,$$

so we can write

$$\mathrm{tr}(h_N^\sharp) = \sum_{i=1}^{m} \widetilde{g}(-\widetilde{\nabla}_{E_i} N, E_i) + \widetilde{g}(-\widetilde{\nabla}_N N, N) = \sum_{i=1}^{m+1} \widetilde{g}(-\widetilde{\nabla}_{E_i} N, E_i).$$

This is nothing but the coordinate expression for a trace over the ambient tangent vector fields $\mathcal{T}\widetilde{\mathcal{M}}|_{\mathcal{M}}$ (restricted to the neighbourhood U). In coordinate free form, then,

$$\mathrm{tr}(h_N^\sharp) = \mathrm{Tr}(h_N^\sharp),$$

where we use the capitalised Tr to denote the 'enlarged' trace over the ambient tangent space.

Then from Definition 5.4 of the mean curvature scalar field,

$$H_N = \frac{1}{m}\mathrm{tr}(h_N^\sharp) = \frac{1}{m}\mathrm{Tr}(h_N^\sharp) = -\frac{1}{m}\mathrm{Tr}(X \mapsto \widetilde{\nabla}_X N) = -\frac{1}{m}\mathrm{div}_{\widetilde{\mathcal{M}}} N.$$

□

This result is directly applicable to the Ryu-Takayanagi formula for entanglement entropy in holographic theories. The Ryu-Takayanagi prescription relates a minimal submanifold γ_A to a region A on the conformal boundary of a manifold, such that $\partial^c \gamma_A = \partial A$; thus $\dim \gamma_A = \dim A$. In a $(d+1)$-dimensional constant-time slice of AdS_{d+2}, this means that A and γ_A are both d-dimensional. The submanifold γ_A is in particular a codimension 1 minimal submanifold in $\mathbb{H}^{d+1}$, and so must have $H_N = -\frac{1}{d}\mathrm{div}_{\mathbb{H}} N = 0$ for unit normal field N in the hyperbolic half-space.

5.3 Mean curvature in hyperbolic ambient manifolds.

In this section, we investigate the mean curvature of submanifolds embedded in a Poincaré half-plane model of the hyperbolic space $\mathbb{H}^{d+1}$. As we saw in Section 3.2.3, this is a description of the constant-time slice of anti-de Sitter spacetime AdS_{d+2} given in Poincaré coordinates. To start, we work with a submanifold of general codimension. However, the Ryu-Takayanagi prescription describes submanifolds γ_A of codimension 1 in the constant-time slice, so we will specialise to this case at the end.

Recall that in the half-space model of a $(d+1)$-dimensional hyperbolic space $\mathbb{H}^{d+1}$, we have coordinates $\mathbf{x} = (x^1, x^2, \ldots, x^d, z) \in \mathbb{R}^{d+1}$ with $z > 0$ and, as per (3.18), a metric tensor

$$g_{\mathbb{H}^{d+1}} = \frac{1}{z^2}\left[\sum_{i=1}^{d} \left(\mathrm{d}x^i\right)^2 + \mathrm{d}z^2\right]. \tag{5.4}$$

There is a natural embedding of the hyperbolic half-space into the Euclidean space of the same dimension, $\iota : \mathbb{H}^{d+1} \hookrightarrow \mathbb{E}^{d+1}$, given simply by $\iota(\mathbf{x}) = \mathbf{x}$. Moreover, this embedding is conformal since $\iota^* g_{\mathbb{E}^{d+1}} = z^2 g_{\mathbb{H}^{d+1}}$.

Owing to our familiarity with Euclidean geometry, it is often easiest to work with Euclidean quantities. Therefore, it will be convenient to relate our quantities in $\mathbb{H}^{d+1}$ to those in $\mathbb{E}^{d+1}$ via ι.

Consider a submanifold $\mathcal{M}$ embedded in the hyperbolic half-space $\mathbb{H}^{d+1}$; then $\iota(\mathcal{M})$ is a submanifold embedded in Euclidean space $\mathbb{E}^{d+1}$. Choose a field of normal vectors $N_{\mathbb{H}} \in \mathcal{N}\mathcal{M}$ on $\mathcal{M}$ with unit hyperbolic norm everywhere:

$$|N_{\mathbb{H}}|_{\mathbb{H}} = 1$$

at all points on $\mathcal{M}$. Using the methods of Section 5.2, we can construct mean curvature field $H_{N_{\mathbb{H}}}$ of $\mathcal{M}$ in ambient hyperbolic space $\mathbb{H}^{d+1}$ with respect to this unit normal field.

However, we can also relate this hyperbolic mean curvature $H_{N_{\mathbb{H}}}$ to a mean curvature of $\iota(\mathcal{M})$ in the Euclidean ambient space. To do so, we must

first obtain a unit normal vector field $N_{\mathbb{E}} \in \mathcal{N}[\iota(\mathcal{M})]$ in the Euclidean ambient space that is related to the unit normal field $N_{\mathbb{H}} \in \mathcal{NM}$ in the hyperbolic ambient space.

If we push the unit normal field $N_{\mathbb{H}}$ forward into the Euclidean embedding we obtain a vector field $\iota_* N_{\mathbb{H}}$, naïvely in $\mathcal{T}\mathbb{E}^{d+1}|_{\iota(\mathcal{M})}$. The embedding ι is a diffeomorphism onto its image within the Euclidean space, so that a vector field on $\mathcal{M}$ will indeed be pushed forward to a vector field on $\iota(\mathcal{M})$. However, ι is conformal and therefore preserves angles; so any vector normal to $\mathcal{M}$ in the hyperbolic space will be pushed forward to a vector normal to $\iota(\mathcal{M})$ in the Euclidean space. Likewise, any vector tangent to $\mathcal{M}$ will be pushed forward to a vector tangent to $\iota(\mathcal{M})$.

Therefore, $\iota_* N_{\mathbb{H}}$ is actually a normal vector field in $\mathcal{N}[\iota(\mathcal{M})]$. In order to build a *unit* normal vector field $N_{\mathbb{E}}$ out of $\iota_* N_{\mathbb{H}}$, we thus only need to rescale $\iota_* N_{\mathbb{H}}$ at each point on $\iota(\mathcal{M})$.

To this end, consider any vector $X \in T\mathbb{H}^{d+1}$; its norm is determined by the hyperbolic metric:

$$|X|_{\mathbb{H}} = \sqrt{g_{\mathbb{H}^{d+1}}(X, X)}.$$

The embedding ι from $\mathbb{H}^{d+1}$ into $\mathbb{E}^{d+1}$ is conformal, so

$$g_{\mathbb{H}^{d+1}}(X, X) = \tfrac{1}{z^2}\left(\iota^* g_{\mathbb{E}^{d+1}}\right)(X, X) = \tfrac{1}{z^2} g_{\mathbb{E}^{d+1}}(\iota_* X, \iota_* X),$$

and thus

$$|X|_{\mathbb{H}} = \frac{1}{z} |\iota_* X|_{\mathbb{E}}\,. \tag{5.5}$$

Since ι is a 'trivial' embedding $\iota(\mathbf{x}) = \mathbf{x}$, the $\iota_* X$ and X have identical expressions in coordinates: if $X = X^i \partial_i$ then $\iota_* X = \iota_*(X^i \partial_i) = X^i \partial_i$. For this reason, we can unambiguously drop the explicit pushforward notation, and treat X as a vector in both Euclidean and hyperbolic space.

On this basis, and in light of (5.5) in particular, define

$$N_{\mathbb{E}} := \frac{1}{z} N_{\mathbb{H}}. \tag{5.6}$$

Then

$$|N_{\mathbb{E}}|_{\mathbb{E}} = \frac{1}{z} |N_{\mathbb{H}}|_{\mathbb{E}} = |N_{\mathbb{H}}|_{\mathbb{H}} = 1,$$

at all points on $\iota(\mathcal{M})$.

So, a given hyperbolic unit normal field $N_{\mathbb{H}}$ corresponds via (5.6) to a Euclidean unit normal field $N_{\mathbb{E}} = \frac{1}{z} N_{\mathbb{H}}$. For notational convenience, define the hyperbolic mean curvature $H^{\mathbb{H}}_N := H_{N_{\mathbb{H}}}$ of $\mathcal{M}$ relative to $N_{\mathbb{H}}$, and Euclidean mean curvature $H^{\mathbb{E}}_N := H_{N_{\mathbb{E}}}$ of $\iota(\mathcal{M})$ relative to $N_{\mathbb{E}}$.

Theorem 5.7. *In the half-space model of $\mathbb{H}^{d+1}$ embedded as above in $\mathbb{E}^{d+1}$, the hyperbolic and Euclidean mean curvatures of a submanifold $\mathcal{M}$ satisfy*

$$H^{\mathbb{H}}_N = z H^{\mathbb{E}}_N + N^{d+1}_{\mathbb{E}},$$

where $N_{\mathbb{E}}^{d+1}$ is the $(d+1)$-th component of the Euclidean unit normal $N_{\mathbb{E}}$.

Proof. We present the proof given by Pacard and Pimentel (2004); an alternative argument can be found in Fonda, Giomi, et al. (2015, Appendix A). We begin by calculating the shape operators of $\mathcal{M}$ in hyperbolic ambient space and $\iota(\mathcal{M})$ in Euclidean ambient space, using Corollary 5.4. We denote the Riemannian connections associated with $g_{\mathbb{H}^{d+1}}$ and $g_{\mathbb{E}^{d+1}}$ by $\widetilde{\nabla}$ and $\overline{\nabla}$ respectively.

From (5.4), the metric $g_{\mathbb{H}^{d+1}}$ in the half-space model has components $g_{ij} = \frac{1}{z^2}\delta_{ij}$. The associated Christoffel symbols are

$$\Gamma_{ij}^k = \frac{1}{2}g^{kl}\left(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}\right) = \frac{1}{z}\left(\delta_{d+1}^k\delta_{ij} - \delta_j^k\delta_{i,d+1} - \delta_i^k\delta_{j,d+1}\right).$$

For any vector fields X, Y on $\mathbb{H}^{d+1}$, the covariant derivative $\widetilde{\nabla}_X Y$ is given in coordinates by

$$\begin{aligned}\widetilde{\nabla}_X Y &= \left(XY^k + X^iY^j\Gamma_{ij}^k\right)\partial_k \\ &= (XY^k)\partial_k + g_{\mathbb{H}^{d+1}}(X,Y)\cdot z\partial_{d+1} - \tfrac{1}{z}\left(Y^{d+1}X^k + X^{d+1}Y^k\right)\partial_k.\end{aligned}$$

Now, let $X \in \mathcal{T}\mathcal{M}$ be a vector field tangent to $\mathcal{M}$ and let $Y = N_{\mathbb{H}}$ be the chosen hyperbolic unit normal field of $\mathcal{M}$. Then X and $N_{\mathbb{H}}$ are orthogonal: $g_{\mathbb{H}^{d+1}}(X, N_{\mathbb{H}}) = 0$. So,

$$\widetilde{\nabla}_X N_{\mathbb{H}} = (XN_{\mathbb{H}}^k)\partial_k - \tfrac{1}{z}\left(N_{\mathbb{H}}^{d+1}X + X^{d+1}N_{\mathbb{H}}\right).$$

Project this onto the tangent bundle of $\mathcal{M}$ to obtain

$$(\widetilde{\nabla}_X N_{\mathbb{H}})^\top = \left[(XN_{\mathbb{H}}^k)\partial_k\right]^\top - \frac{1}{z}N_{\mathbb{H}}^{d+1}X,$$

since $N_{\mathbb{H}}^\top = 0$. Thus the hyperbolic shape operator $h^\sharp_{N_{\mathbb{H}}}$ on $\mathcal{M}$ can be written as

$$h^\sharp_{N_{\mathbb{H}}}(X) = -(\widetilde{\nabla}_X N_{\mathbb{H}})^\top = -\left[(XN_{\mathbb{H}}^k)\partial_k\right]^\top + \frac{1}{z}N_{\mathbb{H}}^{d+1}X. \tag{5.7}$$

We next construct the shape operator $h^\sharp_{N_{\mathbb{E}}}$ on $\iota(\mathcal{M})$. Using the relation (5.6) between Euclidean and hyperbolic normals $N_{\mathbb{H}}$ and $N_{\mathbb{E}}$,

$$\overline{\nabla}_X N_{\mathbb{E}} = \overline{\nabla}_X\left(\tfrac{1}{z}N_{\mathbb{H}}\right) = \tfrac{1}{z}\overline{\nabla}_X N_{\mathbb{H}} + X\left(\tfrac{1}{z}\right)N_{\mathbb{H}}.$$

After projection onto the tangent space, this becomes

$$(\overline{\nabla}_X N_{\mathbb{E}})^\top = \frac{1}{z}\left[\overline{\nabla}_X N_{\mathbb{H}}\right]^\top = \frac{1}{z}\left[(XN_{\mathbb{H}}^k)\partial_k\right]^\top,$$

since $N_{\mathbb{H}}^{\top} = 0$ and, in the last equality, the Christoffel symbols of the Riemannian connection $\overline{\nabla}$ on Euclidean space vanish. The Euclidean shape operator can therefore be written as

$$h^{\sharp}_{N_{\mathbb{E}}}(X) = -(\overline{\nabla}_X N_{\mathbb{E}})^{\top} = -\frac{1}{z}\left[(XN^k_{\mathbb{H}})\partial_k\right]^{\top}.$$

Comparing this to the expression (5.7) for the hyperbolic shape operator, we find that

$$h^{\sharp}_{N_{\mathbb{H}}}(X) = z h^{\sharp}_{N_{\mathbb{E}}}(X) + (\tfrac{1}{z}N_{\mathbb{H}})^{d+1}X = z h^{\sharp}_{N_{\mathbb{E}}}(X) + N_{\mathbb{E}}^{d+1}X,$$

or

$$h^{\sharp}_{N_{\mathbb{H}}} = z h^{\sharp}_{N_{\mathbb{E}}} + N_{\mathbb{E}}^{d+1}\mathrm{id}.$$

To find the scalar mean curvature, we take the trace of the shape operator and divide by $m = \dim \mathcal{M}$:

$$H^{\mathbb{H}}_N = \frac{1}{m}\mathrm{tr}(h^{\sharp}_{N_{\mathbb{H}}}) = z\left[\frac{1}{m}\mathrm{tr}(h^{\sharp}_{N_{\mathbb{E}}})\right] + N_{\mathbb{E}}^{d+1}\frac{\mathrm{tr}(\mathrm{id})}{m} = zH^{\mathbb{E}}_N + N_{\mathbb{E}}^{d+1}.$$

□

To find holographic entanglement entropies, we must find codimension-1 surfaces γ_A in hyperbolic space, which minimise area and thus satisfy $H^{\mathbb{H}}_N = 0$ everywhere. Theorem 5.7 allows us to shift our calculations into the more familiar Euclidean space: we seek surfaces in $\mathbb{E}^{d+1}$ that satisfy $H^{\mathbb{E}}_N = -\frac{1}{z}N_{\mathbb{E}}^{d+1}$ everywhere.

Additionally, since these surfaces are of codimension 1, Theorem 5.6 allows us to express $H^{\mathbb{E}}_N$ using the usual Euclidean divergence of the normal vector field $N_{\mathbb{E}}$. In particular, the condition $H_{\mathbb{E}} = -\frac{1}{z}N_E^{d+1}$ can be written as

$$\mathrm{div}_{\mathbb{E}} N_{\mathbb{E}} = \frac{d}{z}N_E^{d+1}. \tag{5.8}$$

Let us specialise to the $d = 2$ case of interest in the rest of this work, and use rectangular coordinates (x, y, z). Suppose we can write the surface γ_A as a level set $f(x, y, z) = 0$ for some smooth function f. Then the Euclidean normal vector field is simply

$$N_{\mathbb{E}} = \frac{\overline{\nabla} f}{|\overline{\nabla} f|} = \frac{(f_x, f_y, f_z)}{\sqrt{f_x^2 + f_y^2 + f_z^2}},$$

where $\overline{\nabla} = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ is the familiar Euclidean derivative operator, $f_x = \frac{\partial f}{\partial x}$ and likewise for y and z.

The equation (5.8) for a minimal surface $f(x, y, z) = 0$ in the hyperbolic space is therefore

$$\overline{\nabla} \cdot \left(\frac{\overline{\nabla} f}{|\overline{\nabla} f|}\right) = \frac{2}{z}\frac{f_z}{|\overline{\nabla} f|} \tag{5.9}$$

The disc and infinite-strip examples treated in Chapter 4 are special cases of this in which $f(x, y, z) = z - \xi(x, y)$, so that the level set is really the graph $z = \xi(x, y)$. In this particular case, the hyperbolic minimal surface equation (5.9) becomes

$$\overline{\nabla} \cdot \left(\frac{(-z_x, -z_y, 1)}{\sqrt{1 + z_x^2 + z_y^2}} \right) = \frac{2}{z\sqrt{1 + z_x^2 + z_y^2}},$$

or

$$\left(1 + z_x^2\right) z_{yy} - 2z_x z_y z_{xy} + \left(1 + z_y^2\right) z_{xx} = -\frac{2}{z}\left(1 + z_x^2 + z_y^2\right).$$

Alternatively, we could write this using cylindrical coordinates (r, θ, z) instead of rectangular coordinates (x, y, z). By applying the gradient and divergence operators in cylindrical coordinates to $f(r, \theta, z) = z - \xi(r, \theta)$ in (5.9), one finds

$$\left(r^2 + z_\theta^2\right) z_{rr} - 2z_r z_\theta z_{r\theta} + (1 + z_r^2) z_{\theta\theta} = -\frac{2}{z}\left(r^2 + z_\theta^2 + r^2 z_r^2\right) - \frac{2}{r} z_r z_\theta^2 - r z_r (1 + z_r^2)$$

To compare to the methods used in Chapter 4, assume that $z = z(r)$ only, i.e. $z_\theta = 0$ and $z_r = z'$. Then the above simplifies directly to the equation (4.3) found using variational calculus to describe the surface γ_A associated with disc-shaped A.

This demonstrates explicitly that the mean curvature expressions presented in this chapter do indeed reduce to the same expressions found by variational methods, in the cases where the variational methods can be applied.

However, the characterisation $H_N = 0$, or more generally $K = 0$, forms the beginning to a deeper structural understanding of the kinds of surfaces employed in the Ryu-Takayanagi conjecture, and the particular coordinate-free expression (5.8) provides a concise and convenient expression of this characterisation in the geometry of interest that can be employed for any coordinate description of the submanifold.

Chapter 6

Conformal symmetry transformations

In the AdS/CFT conjecture, the field theory on the boundary of AdS is a conformal field theory; this means that its collection of spacetime symmetry transformations is made up of conformal maps on the boundary spacetime manifold. Building from the definition of a conformal map given in Chapter 3, we seek in this chapter to understand these conformal symmetries. In particular, we are interested in the conformal boundary of a constant-time slice of AdS_4, which we found in Section 3.2.3 to be $\mathbb{S}^2$.

Using these symmetries we can can map some of the regions for which we have found minimal surfaces and entanglement entropies in Chapter 4, to other regions for which we have not. This will allow us to deduce the entanglement entropies of the latter from those of the former.

6.1 Conformal symmetries of a manifold

The symmetries of a physical theory naturally arise with an algebraic structure. At a schematic level, we identify symmetries as transformations on the space of physical configurations or states in the theory. In other words, symmetries are maps from the space of states to itself. In particular, a symmetry will map any physical state to another physically equivalent state. Without rigorous definition of this term, we nonetheless expect it to guarantee that symmetries are invertible, and that a composition of two symmetry transformations is another symmetry transformation. This is enough to ensure that (with the binary operation of map composition), the set of symmetries of a theory forms a *group*.

In the conformal field theory defined on the boundary manifold of AdS, the symmetries of interest are (as the name suggests) conformal transformations of the boundary manifold. This is as opposed to, for instance, the

internal symmetries familiar from Yang-Mills gauge theories, or the Poincaré symmetries of relativistic theories on Minkowski spacetime.

The first thing to note is that, for any (pseudo-)Riemannian manifold $(\mathcal{M}, g)$, the set of conformal maps $\mathcal{M} \to \mathcal{M}$ does not, as it stands, form a group: in particular, there is no guarantee that a conformal transformation is invertible. As an example of this, consider the unit circle $\mathbb{S}^1 \subset \mathbb{E}^2$. Choose the angle $\phi \mapsto (\cos\phi, \sin\phi)$ as a coordinate parameterisation. Then the map $f : \mathbb{S}^1 \to \mathbb{S}^1$ given by $f(\phi) = 2\phi$ in this angular coordinate is easily seen to be conformal, but not injective.

So, to find the group of conformal symmetries, we need to restrict the set of conformal transformations on $\mathcal{M}$ to those that are also invertible, and whose inverses are themselves conformal transformations.

We start with the well-known group $\mathrm{Diff}(\mathcal{M})$ of diffeomorphisms on $\mathcal{M}$, i.e. smooth invertible maps $\mathcal{M} \to \mathcal{M}$ whose inverses are also smooth. Since conformal maps must be smooth, the desired group of conformal symmetries will be a subgroup of $\mathrm{Diff}(M)$.

Definition 6.1. A *conformal diffeomorphism* of (pseudo-)Riemannian manifold $(\mathcal{M}, g)$ is a diffeomorphism $f \in \mathrm{Diff}(\mathcal{M})$ that is also a conformal map. The set of all conformal diffeomorphisms of $(\mathcal{M}, g)$ is denoted $\mathrm{Conf}(\mathcal{M}, g)$.

By Definition 3.2, a conformal diffeomorphism f has some strictly positive conformal factor $\Lambda \in C^\infty(\mathcal{M})$ such that $f^*g = \Lambda^2 g$. Other names for conformal diffeomorphisms are more common in the particular field of complex analysis: for instance, Needham (1998) and others use the term *automorphism*, while Conway (1996) uses *conformal equivalence*.

Proposition 6.1. $\mathrm{Conf}(\mathcal{M}, g)$ *is a subgroup of* $\mathrm{Diff}(\mathcal{M})$.

Proof. We only need to show that $\mathrm{Conf}(\mathcal{M}, g)$ is closed under the group operation (i.e. map composition) and inversion. To do so, recall that for $f_1, f_2 \in \mathrm{Diff}(\mathcal{M})$ the composition $f_2 \circ f_1$ has pullback $(f_2 \circ f_1)^* = f_1^* f_2^*$. This follows in turn from the composition law for pushforwards of vectors. For instance, acting on the metric g at some point $p \in \mathcal{M}$, with $X, Y \in T_p\mathcal{M}$:

$$\begin{aligned}
\left[(f_2 \circ f_1)^* g\right](X, Y) &= g\left((f_2 \circ f_1)_* X, (f_2 \circ f_1)_* Y\right) \\
&= g\left((f_2)_*(f_1)_* X, (f_2)_*(f_1)_* Y\right) \\
&= f_2^* g\left((f_1)_* X, (f_1)_* Y\right) \\
&= \left[f_1^* f_2^* g\right](X, Y).
\end{aligned}$$

So, for $f_1, f_2 \in \mathrm{Conf}(\mathcal{M}, g)$ we find that

$$(f_2 \circ f_1)^* g = f_1^* f_2^* g = f_1^*(\Lambda_2^2\, g) = (\Lambda_2 \circ f_1)^2 f_1^* g = (\Lambda_2 \circ f_1)^2 \Lambda_1^2 g.$$

Since both Λ_1 and $\Lambda_2 \circ f_1$ are positive $C^\infty(\mathcal{M})$ functions, so is their product; this shows that $f_2 \circ f_1 \in \mathrm{Conf}(\mathcal{M}, g)$.

Next, take $f \in \mathrm{Conf}(\mathcal{M}, g)$. Since the conformal factor is strictly positive, we can write $g = (1/\Lambda^2) f^* g$. Then

$$\left(f^{-1}\right)^* g = \left(f^{-1}\right)^* \left[\frac{1}{\Lambda^2} f^* g\right] = \frac{1}{(\Lambda \circ f^{-1})^2} \left(f \circ f^{-1}\right)^* g = \frac{1}{(\Lambda \circ f^{-1})^2}\, g,$$

where, in the last equality, we have used that $\mathrm{id}_{\mathcal{M}}^* g = g$. This shows that $f^{-1} \in \mathrm{Conf}(\mathcal{M}, g)$.

□

Thus the group of spacetime symmetries for a conformal field theory defined on spacetime $(\mathcal{M}, g)$ is the group $\mathrm{Conf}(\mathcal{M}, g)$.

In fact, we will primarily be interested in those conformal symmetries that also preserve orientation, since they form the portion of the symmetry group connected to the identity and therefore describable using the associated Lie algebra. Although we will not use Lie algebra tools directly here, most implementations of conformal symmetry in physical theories use them extensively enough that the algebra, rather than the group, is taken to define the symmetries.

Definition 6.2. By $\mathrm{Conf}_+(\mathcal{M}, g)$ we denote the subset of $\mathrm{Conf}(\mathcal{M}, g)$ consisting of orientation-preserving maps, i.e. maps $f \in \mathrm{Conf}(\mathcal{M}, g)$ with everywhere-positive Jacobian determinant $\det J_f > 0$.

Proposition 6.2. $\mathrm{Conf}_+(\mathcal{M}, g)$ *is a subgroup of* $\mathrm{Conf}(\mathcal{M}, g)$.

Proof. Say $f, h \in \mathrm{Conf}_+(\mathcal{M}, g)$. Then by the composition law for Jacobians, $f \circ h$ has Jacobian determinant $\det(J_{f\circ h}) = \det(J_f J_h) = (\det J_f)(\det J_h) > 0$, so $f \circ h \in \mathrm{Conf}_+(\mathcal{M}, g)$. Also, $\det J_{f^{-1}} = \det(J_f)^{-1} = 1/\det J_f > 0$ so that $f^{-1} \in \mathrm{Conf}_+(\mathcal{M}, g)$. □

Below, we use the term conformal diffeomorphism to refer specifically to maps in $\mathrm{Conf}_+(\mathcal{M}, g)$; we will not be interested in maps that reverse orientation.

6.2 Conformal maps on the plane

We are interested in the (orientation-preserving) conformal symmetries of $\mathbb{S}^2$, the conformal boundary of AdS_4. In Section 3.2.1, we found $\mathbb{S}^2$ to be the conformal completion of $\mathbb{E}^2$, essentially by adding a single point at infinity to $\mathbb{E}^2$. This suggests that, with some effort to account for the extra point at infinity, conformal maps on $\mathbb{E}^2$ can be useful in studying those on $\mathbb{S}^2$.

Recall that d-dimensional Euclidean space is the Riemannian manifold $\mathbb{E}^d = \left(\mathbb{R}^d, g_{\mathbb{E}^d}\right)$, where $g_{\mathbb{E}^d}$ is the familiar dot product. This manifold is covered

by a single chart given by the rectangular coordinates $\mathbf{x} = (x^1, \ldots, x^d) \in \mathbb{R}^d$. In such coordinates,

$$g_{\mathbb{E}^d} = \delta_{ij} \mathrm{d}x^i \mathrm{d}x^j = \sum_{i=1}^{d} \left(\mathrm{d}x^i\right)^2.$$

Consider a conformal transformation f of the Euclidean plane $\mathbb{E}^2$,

$$f : \mathbb{E}^2 \to \mathbb{E}^2 \qquad \text{with} \qquad f^* g_{\mathbb{E}^2} = \Lambda^2 g_{\mathbb{E}^2}.$$

As per Proposition 3.1, we can write the tangent map as $f_*|_p = \Lambda(p) R_p$ at any point p, where R_p is a linear isometry of $T_p\mathbb{E}^2$. In this case, the set of such linear isometries of is well-known: it is the group $\mathrm{O}(2)$ of orthogonal transformations (rotations and reflections) in the plane.

Any $R \in \mathrm{O}(2)$ has $\det R = \pm 1$. In the case that $\det R = +1$, R is simply a rotation by some angle $\theta \in [-\pi, \pi]$:

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

In the case that $\det R = -1$, R can be represented as a reflection in the x-axis followed by a rotation:

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

Let us choose coordinates $\mathbf{z} = (x, y)$ on the domain of f, and $\mathbf{w} = (u, v)$ on the target space, so that $\mathbf{w} = f(\mathbf{z}) = (u(x, y), v(x, y))$. Then as per (3.2), the pushforward f_* is represented in the coordinate basis by the Jacobian of f:

$$J_f(\mathbf{z}) = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} = \Lambda(\mathbf{z}) R(\mathbf{z}),$$

where $R(\mathbf{z}) \in \mathrm{O}(2)$ for all $\mathbf{z} \in \mathbb{R}^2$.

The Euclidean plane $\mathbb{E}^2$ is a connected manifold. Hence, as per the discussion at the end of Section 3.1, either $\det J_f > 0$ everywhere if f is orientation-preserving, or $\det J_f < 0$ everywhere if f is orientation-reversing. Therefore, either $\det R(\mathbf{z}) = +1$ for all $\mathbf{z}$ or $\det R(\mathbf{z}) = -1$ for all $\mathbf{z}$. Let us start with the former case.

Theorem 6.3. *Consider $U \subseteq \mathbb{E}^2$ an open neighbourhood of the point $\mathbf{z}$, and $f : U \to \mathbb{E}^2$ a smooth map represented in rectangular coordinates as above. Then f is orientation-preserving and conformal at $\mathbf{z}$ if and only if*

$$\frac{\partial u}{\partial x}(\mathbf{z}) = \frac{\partial v}{\partial y}(\mathbf{z}) \qquad \text{and} \qquad \frac{\partial u}{\partial y}(\mathbf{z}) = -\frac{\partial v}{\partial x}(\mathbf{z}), \tag{6.1}$$

and

$$\det J_f(\mathbf{z}) = \left(\frac{\partial u}{\partial x} \frac{\partial v}{\partial y} - \frac{\partial u}{\partial y} \frac{\partial v}{\partial x} \right)\Bigg|_{\mathbf{z}} \neq 0.$$

Proof. Say f is conformal and orientation preserving at $\mathbf{z}$. Then from the local definition of conformality, Definition 3.3, $J_f(\mathbf{z}) = \Lambda(\mathbf{z})R(\mathbf{z})$ where $\Lambda(\mathbf{z})$ is the conformal factor of f and $R(\mathbf{z})$ is an orthogonal matrix. But since f is orientation-preserving at $\mathbf{z}$, $\det J_f(\mathbf{z}) > 0$. So $\det R(\mathbf{z}) = +1$, and we can write

$$J_f(\mathbf{z}) = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} = \Lambda(\mathbf{z}) \begin{pmatrix} \cos\theta(\mathbf{z}) & -\sin\theta(\mathbf{z}) \\ \sin\theta(\mathbf{z}) & \cos\theta(\mathbf{z}) \end{pmatrix}.$$

It follows that (6.1) is satisfied everywhere.

Conversely, say f satisfies (6.1) and $\det J_f(\mathbf{z}) \neq 0$. Let

$$\Lambda := \sqrt{\left(\frac{\partial u}{\partial x}(\mathbf{z})\right)^2 + \left(\frac{\partial u}{\partial y}(\mathbf{z})\right)^2}.$$

Since

$$0 \neq \det J_f = \frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x} = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2,$$

we see that $\Lambda > 0$ strictly and $\det J_f(\mathbf{z}) = \Lambda^2 > 0$ strictly; therefore f is orientation-preserving at $\mathbf{z}$.

Now, using (6.1) we find that

$$\left(\frac{J_f}{\Lambda}\right)\left(\frac{J_f}{\Lambda}\right)^t\Bigg|_{\mathbf{z}} = \frac{1}{\Lambda^2}\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ -\frac{\partial u}{\partial y} & \frac{\partial u}{\partial x} \end{pmatrix}_{\mathbf{z}} \begin{pmatrix} \frac{\partial u}{\partial x} & -\frac{\partial u}{\partial y} \\ \frac{\partial u}{\partial y} & \frac{\partial u}{\partial x} \end{pmatrix}_{\mathbf{z}} = \mathbb{1}.$$

Therefore, $R(\mathbf{z}) := J(\mathbf{z})/\Lambda(\mathbf{z})$ is an orthogonal matrix. So, according to the local definition of conformality, f is conformal at $\mathbf{z}$. □

The conditions (6.1) are the Cauchy-Riemann conditions of complex analysis. This motivates us to treat the Euclidean plane $\mathbb{E}^2$ as the complex plane $\mathbb{C}$ (essentially by artificially introducing complex multiplication). The rectangular coordinates $\mathbf{z} = (x, y)$ and $\mathbf{w} = f(\mathbf{z}) = (u, v)$ become $z = x + iy$ and $w = f(z) = u(x, y) + iv(x, y)$. Denoting by $f'(z)$ the complex derivative of f, note that $|f'(z)|^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 = \det J_f$.

Then Theorem 6.3 tells us that the conformal, orientation-preserving maps $f(z)$ on $\mathbb{C}$ are precisely those which are (complex) analytic with non-zero derivative $f'(z) \neq 0$.

Corollary 6.4. *Take a smooth map $f : U \to \mathbb{E}^2$ for U an open neighbourhood of $\mathbf{z}$ in $\mathbb{E}^2$ as above. f is conformal and orientation-reversing at $\mathbf{z}$ if and only if*

$$\frac{\partial u}{\partial x}(\mathbf{z}) = -\frac{\partial v}{\partial y}(\mathbf{z}) \qquad \textit{and} \qquad \frac{\partial u}{\partial y}(\mathbf{z}) = \frac{\partial v}{\partial x}(\mathbf{z}), \tag{6.2}$$

and $\det J_f(\mathbf{z}) \neq 0$.

This can be proved similarly to Theorem 6.3; the conditions (6.2) are often called the anti-Cauchy-Riemann conditions. If $f(z)$ obeys the anti-Cauchy-Riemann conditions then $f(\bar{z}) = f(x - iy)$ is an analytic function. In complex analysis, usually a (geometrically) conformal map is called conformal when it preserves orientation and anti-conformal when it reverses orientation.

Having identified $\mathbb{E}^2$ with the complex plane $\mathbb{C}$ and found that conformal maps are analytic functions with non-zero derivative, we can now use the tools of complex analysis to explore the group $\mathrm{Conf}_+(\mathbb{C})$ of orientation-preserving conformal diffeomorphisms of the plane.

Lemma 6.5. *All $f \in \mathrm{Conf}_+(\mathbb{C})$ are polynomials of degree one,*

$$f(z) = \alpha z + \beta$$

for some $\alpha, \beta \in \mathbb{C}$ with $\alpha \neq 0$.

Proof. For any $f \in \mathrm{Conf}_+(\mathbb{C})$, we first show that $f(z) \to \infty$ as $z \to \infty$, i.e. that f has a pole at infinity. To do so, assume that $\lim_{z\to\infty} f(z)$ is either finite (as in the case of a removable singularity at infinity) or does not exist (as in the case of an essential singularity at infinity). Then in either case there is a sequence $\{z_n\}$ in $\mathbb{C}$ such that $z_n \to \infty$ while $f(z_n) \to a \in \mathbb{C}$. In the case where $\lim_{z\to\infty} f(z)$ exists, naturally $a = \lim_{z\to\infty} f(z)$; otherwise, this is an application of Picard's Great Theorem (in any neighbourhood of an essential singularity, the function f attains every complex value barring at most one). But since f is a diffeomorphism, f^{-1} is continuous. Therefore $z_n = f^{-1}(f(z_n)) \to f^{-1}(a)$, which is in $\mathbb{C}$ and therefore not ∞. This is a contradiction; so, $\lim_{z\to\infty} f(z) = \infty$.

Thus f has a pole at infinity. From Theorem 6.3, f is analytic everywhere on $\mathbb{C}$, i.e. f is entire. But any entire function with a pole at infinity is a polynomial. Not only this, but f also has nonzero derivative everywhere. Therefore, it must be a polynomial of degree one: $f(z) = \alpha z + \beta$ for $\alpha, \beta \in \mathbb{C}$. Clearly $\alpha \neq 0$ for f to be invertible. □

6.3 Conformal maps on the Riemann sphere

Interesting as it is to have determined the form of all conformal diffeomorphisms on the plane, we are really interested in the sphere $\mathbb{S}^2$. We will exploit the fact that $\mathbb{S}^2$ is the conformal completion of the plane, in order to use the tools of complex analysis to treat conformal maps on the sphere.

With the plane taken to be the complex plane $\mathbb{C}$ as above, we denote $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$. We call $\mathbb{C}_\infty$ the Riemann sphere, or the extended complex plane. The point ∞ is to be understood intuitively as the 'preimage' under stereographic projection of the North pole $\mathbf{N}$ of the 2-sphere $\mathbb{S}^2$ into which $\mathbb{C}$ must be embedded for conformal completion. So, $\mathbb{C}_\infty$ has the topology and

smooth structure of $\mathbb{S}^2$; its metric tensor is that of the Euclidean plane $\mathbb{C}$ except at the extra point ∞ where it blows up; thus it is conformally related to $\mathbb{S}^2$. We will use conformal maps on $\mathbb{C}_\infty$ to understand those on $\mathbb{S}^2$.

Recall the stereographic projection of Section 3.2.1, which gives us a chart that covers all but the point $\mathbf{N}$ on $\mathbb{S}^2$ with the plane $\mathbb{C}$. In particular, the expression (3.9) can be specialised to a map $\phi : \mathbb{C} \to \mathbb{S}^2 \setminus \{\mathbf{N}\}$ that gives a coordinate map (i.e. inverse of a chart) for the sphere. For $z \in \mathbb{C}$ and $\mathbf{y} = (y^1, y^2, y^3) \in \mathbb{S}^2$, this map is

$$\mathbf{y} = \phi(z) = \left(\frac{2\Re(z)}{|z|^2 + 1}, \frac{2\Im(z)}{|z|^2 + 1}, \frac{|z|^2 - 1}{|z|^2 + 1} \right). \tag{6.3}$$

In general, one would need to use additional charts to cover the whole sphere; to discuss a map on $\mathbb{S}^2$, one would have to discuss its expressions in the various charts as well as the transitions between those charts. However, there is only one point on $\mathbb{S}^2$ that is not covered by the stereographic projection chart. In this case, it is common to use exclusively this single chart, and to simply prescribe additional ad hoc rules to treat functions at the excluded point.

In this vein, let $f : \mathbb{S}^2 \to \mathbb{S}^2$ be a smooth map. Using the stereographic projection ϕ, the coordinate representation $\phi^{-1} \circ f \circ \phi$ is a map on the complex plane; thus Theorem 6.3 tells us that it is conformal when it is analytic and has non-zero derivative.

We already saw in the proof of Proposition 6.1 that compositions of conformal maps are conformal, although there we used the global definition of conformality. It is easy to see from the composition property of the Jacobian $J_{f \circ g}(p) = J_f(g(p)) J_g(p)$, the local alternative: if a map g is conformal at p, and f is conformal at $g(p)$, then $f \circ g$ is conformal at p.

So when the coordinate representation $\phi^{-1} \circ f \circ \phi$ is conformal at a point on the plane, the map f is also conformal at the corresponding point on the sphere, since ϕ is conformal.

It remains to develop methods to say when $f : \mathbb{S}^2 \to \mathbb{S}^2$ is conformal at the points $\mathbf{N}$ and $f^{-1}(\mathbf{N})$ not covered by the stereographic projection ϕ.

To this end, consider the inversion map $I : \mathbb{S}^2 \to \mathbb{S}^2$ defined in stereographic coordinates by $\phi^{-1} \circ I \circ \phi(z) = 1/z$ for all $z \in \mathbb{C}_\infty \setminus \{0, \infty\}$, and by $I(\mathbf{N}) = \mathbf{S}$, $I(\mathbf{S}) = \mathbf{N}$ for those points where stereographic coordinate representation does not apply. Recall that $\mathbf{N} = (0, 0, 1)$ and $\mathbf{S} = (0, 0, -1)$ are the North and South poles of the 2-sphere. The action of I on these poles formalises the intuitive idea that, in the complex plane, $1/0 = \infty$ and $1/\infty = 0$.

Let us describe geometrically the action of I on the sphere. From the definition of I in stereographic coordinates on the plane,

$$\phi^{-1} \circ I \circ \phi(z) = \frac{1}{z} \qquad \Leftrightarrow \qquad I \circ \phi(z) = \phi\left(\frac{1}{z}\right).$$

Now recall that $\Re(1/z) = (\Re z)/|z|^2$ and $\Im(1/z) = -(\Im z)/|z|^2$. Then using (6.3), we can write I explicitly in coordinates on the sphere:

$$I\left(\frac{2\Re(z)}{|z|^2+1}, \frac{2\Im(z)}{|z|^2+1}, \frac{|z|^2-1}{|z|^2+1}\right) = \left(\frac{2\frac{\Re(z)}{|z|^2}}{\frac{1}{|z|^2}+1}, \frac{-2\frac{\Im(z)}{|z|^2}}{\frac{1}{|z|^2}+1}, \frac{\frac{1}{|z|^2}-1}{\frac{1}{|z|^2}+1}\right)$$
$$= \left(\frac{2\Re(z)}{|z|^2+1}, -\frac{2\Im(z)}{|z|^2+1}, -\frac{|z|^2-1}{|z|^2+1}\right).$$

Between this expression for points covered by the stereographic projection, and the action of I on the North and South poles of the sphere, we see that $I(y^1, y^2, y^3) = (y^1, -y^2, -y^3)$ for all $(y^1, y^2, y^3) \in \mathbb{S}^2$.

Now recall that, on the Euclidean space $\mathbb{E}^3$ in which the sphere is embedded, the rotation $R_{1,\pi}$ about the y^1-axis by π radians takes the explicit form

$$R_{1,\pi}\begin{pmatrix} y^1 \\ y^2 \\ y^3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\pi & -\sin\pi \\ 0 & \sin\pi & \cos\pi \end{pmatrix}\begin{pmatrix} y^1 \\ y^2 \\ y^3 \end{pmatrix} = \begin{pmatrix} y^1 \\ -y^2 \\ -y^3 \end{pmatrix},$$

so we see that the inversion I on the 2-sphere is nothing but the restriction to the 2-sphere of this rotation. This is clearly an isometric, and thus conformal, diffeomorphism of $\mathbb{S}^2$.

The inversion I now gives us a very simple way to characterise the behaviour of a function $f : \mathbb{S}^2 \to \mathbb{S}^2$ at $\mathbf{N}$ and its preimages on the sphere, using only expressions in the stereographic chart.

Whenever $f \circ I$ is conformal at a point $\mathbf{y}$, the map $f = f \circ I \circ I^{-1}$ is conformal at $I(\mathbf{y})$. Therefore, to establish whether f is conformal at $\mathbf{N}$, which is not covered by the coordinates, we can check whether $f \circ I$ is conformal at $\mathbf{S} = I^{-1}(\mathbf{N})$, which is covered by the coordinates. Similarly, if $I \circ f$ is conformal at a point $\mathbf{y} \in f^{-1}(\mathbf{N})$, then the map $f = I^{-1} \circ I \circ f$ is also conformal at $\mathbf{y}$. So to establish whether f is conformal at a point in $f^{-1}(\mathbf{N})$, we can check whether $I \circ f$ is conformal there instead. While $f(\mathbf{y}) = \mathbf{N}$ is not covered by the coordinates in this case, the point $I \circ f(\mathbf{y}) = \mathbf{S}$ is covered.

With this, we can move entirely to the stereographic coordinate representation: let $\tilde{f} := \phi^{-1} \circ f \circ \phi$, with the understanding that $\tilde{f} : \mathbb{C}_\infty \to \mathbb{C}_\infty$ is a map on the Riemann sphere where ∞ corresponds to the point $\mathbf{N}$ on $\mathbb{S}^2$ above. Then $\tilde{f}(z)$ is conformal at $z = \infty$ if $\tilde{f}(1/z) = \tilde{f} \circ \left(\phi^{-1} \circ I \circ \phi\right)(z)$ is conformal at $z = 0$. Likewise, $\tilde{f}(z)$ is conformal at $z \in \tilde{f}^{-1}(\infty)$ if $1/\tilde{f}(z) = \left(\phi^{-1} \circ I \circ \phi\right) \circ \tilde{f}(z)$ is conformal there.

We can now introduce an important set of maps on on the Riemann sphere.

Definition 6.3 (Möbius transformations)**.** A *Möbius transformation* f on the Riemann sphere is a map $f : \mathbb{C}_\infty \to \mathbb{C}_\infty$ of the form

$$f(z) = \frac{az+b}{cz+b},$$

for some $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$.

Proposition 6.6. *The set of Möbius transformations on $\mathbb{C}_\infty$ forms a group, with the group operation of map composition.*

Proof. Given two Möbius transformations,

$$f_1(z) = \frac{a_1 z + b_1}{c_1 z + d_1} \qquad \text{and} \qquad f_2(z) = \frac{a_2 z + b_2}{c_2 z + d_2},$$

with $a_i d_i - b_i c_i \neq 0$, their composition is

$$f_1 \circ f_2(z) = \frac{(a_1 a_2 + b_1 c_2)z + (a_1 b_2 + b_1 d_2)}{(c_1 a_2 + d_1 c_2)z + (c_1 b_2 + d_1 d_2)} = \frac{a_3 z + b_3}{c_3 z + d_3}. \tag{6.4}$$

By direct calculation, one finds that

$$a_3 d_3 - b_3 c_3 = (a_1 d_1 - b_1 c_1)(a_2 d_2 - b_2 c_2) \neq 0,$$

so $f_1 \circ f_2$ is again a Möbius transformation. The property of associativity is inherited from general map compositions. The identity map $f(z) = z$ is clearly a Möbius transformation, with $a = d = 1$ and $b = c = 0$. We can invert the map $w = f(z)$ explicitly to find

$$z = f^{-1}(w) = \frac{-dw + b}{cw - a}, \tag{6.5}$$

which is again a Möbius transformation. □

Lemma 6.7. *All Möbius transformations on $\mathbb{C}_\infty$ are conformal diffeomorphisms of $\mathbb{C}_\infty$.*

Proof. Since we have already shown that inverses of Möbius transformations are again Möbius transformations, we need only show that any Möbius transformation f is conformal at every point on $\mathbb{C}_\infty$. f is analytic at every point, since $f(z) = \frac{az+b}{cz+d}$ is independent of the complex conjugate $\bar{z}$ and thus satisfies the Cauchy-Riemann equations; so we must check that the derivative is non-vanishing.

First, consider the case where $c \neq 0$. Then $f^{-1}(\infty) = -\frac{d}{c}$. For any $z \in \mathbb{C}_\infty \setminus \{\infty, -\frac{d}{c}\}$,

$$\frac{\mathrm{d}}{\mathrm{d}z} f(z) = \frac{ad - bc}{(cz + d)^2} \neq 0.$$

At the preimage of infinity $z = -\frac{d}{c}$, $1/f(z)$ is conformal:

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{f(z)}\right)\bigg|_{z=-\frac{d}{c}} = \frac{bc - da}{(az + b)^2}\bigg|_{z=-\frac{d}{c}} \neq 0,$$

while for $z = \infty$, $f(1/z)$ is conformal at $z = 0$:

$$\frac{\mathrm{d}}{\mathrm{d}z} f\left(\frac{1}{z}\right)\bigg|_{z=0} = \frac{bc - da}{(dz + c)^2}\bigg|_{z=0} \neq 0.$$

Thus $f(z)$ is conformal everywhere in $\mathbb{C}_\infty$.

Next, consider the case where $c = 0$. Then by setting $\alpha = a/d$ and $\beta = b/d$, we have $f(z) = \alpha z + \beta$ with $\alpha \neq 0$. For all $z \in \mathbb{C}_\infty \setminus \{\infty\}$,

$$\frac{\mathrm{d}}{\mathrm{d}z} f(z) = \alpha \neq 0.$$

For $z = \infty = f^{-1}(\infty)$,

$$\frac{\mathrm{d}}{\mathrm{d}z} \frac{1}{f\left(\frac{1}{z}\right)}\bigg|_{z=0} = \frac{\alpha}{(\beta z + \alpha)^2}\bigg|_{z=0} \neq 0.$$

So again, f is conformal everywhere on $\mathbb{C}_\infty$. □

From this, we see that the Möbius transformations form a subgroup of $\mathrm{Conf}_+(\mathbb{C}_\infty)$. Using the characterisation of $\mathrm{Conf}_+(\mathbb{C})$ of orientation-preserving conformal diffeomorphisms on the plane given in Lemma 6.5, we show that all conformal diffeomorphisms of $\mathbb{C}_\infty$ are in fact Möbius transformations.

Lemma 6.8. *All conformal diffeomorphisms of $\mathbb{C}_\infty$ are Möbius transformations on $\mathbb{C}_\infty$.*

Proof. We present an adaptation of Conway (1996, Chapter 14, Corollary 1.2). Let $f \in \mathrm{Conf}_+(\mathbb{C}_\infty)$ be an conformal diffeomorphism on $\mathbb{C}_\infty$. First, if $f(\infty) = \infty$ then f restricts to a conformal diffeomorphism on the plane $\mathbb{C}$, so Lemma 6.5 applies directly to give the $f(z) = \alpha z + \beta$, which is a Möbius transformation. Otherwise, let $\omega := f(\infty) \neq \infty$ and define

$$g(z) = \frac{1}{f(z) - \omega},$$

or $g = I \circ T \circ f$ where $T(z) = z - \omega$ is a translation, and $I(z)$ is the inversion map discussed above. Then since T and I are both conformal diffeomorphisms on $\mathbb{C}_\infty$, so too is g. But $g(\infty) = \infty$, so that once again Lemma 6.5 applies, giving $(f(z) - \omega)^{-1} = g(z) = cz + d$ for some $c, d \in \mathbb{C}$ with $c \neq 0$. This can be rearranged to give

$$f(z) = \frac{az + b}{cz + d},$$

with $a = c\omega$ and $b = \omega d + 1$, so $ad - bc = -c \neq 0$. □

Together, Lemmas 6.7 and 6.8 fully characterise the group of conformal diffeomorphisms of the Riemann sphere:

Corollary 6.9. *The conformal diffeomorphism group* $\mathrm{Conf}_+(\mathbb{C}_\infty)$ *is exactly the group of Möbius transformations on* $\mathbb{C}_\infty$.

Note that the parameters a, b, c and d in Definition 6.3 of a Möbius transformation are not unique. In particular, for any non-zero $k \in \mathbb{C}$ we have

$$\frac{az+b}{cz+d} = \frac{kaz+kb}{kcz+kd}.$$

Using this, we can always scale the coefficients in a Möbius transformation so that $ad - bc = 1$; in such a case, the Möbius transformation is called *normalised*. In fact, since we can swap the signs on all four parameters a, b, c and d of a normalised Möbius transformation to obtain another normalised transformation, it is clear that any given Möbius transformation has exactly two normalised representations.

In the proof of Proposition 6.6 above, we found expressions (6.4) and (6.5) for inverses and compositions of Möbius transformations. The parameters of a, b, c, d of these transformations may look familiar. Let $z = z_1/z_2$ and $w = w_1/w_2$ for some $z_1, z_2, w_1, w_2 \in \mathbb{C}_\infty$. Then another way to write the Möbius transformation $w = f(z)$ is

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

so we can also represent Möbius transformations as matrices. Then for $f_1(z) = \frac{a_1 z+b_1}{c_1 z+d_1}$ and $f_2(z) = \frac{a_2 z+b_2}{c_2 z+d_2}$, the composition $f_1 \circ f_2$ is represented by

$$\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = \begin{pmatrix} a_1a_2 + b_1c_2 & a_1b_2 + b_1d_2 \\ c_1a_2 + d_1c_2 & c_1b_2 + d_1d_2 \end{pmatrix},$$

matching the parameters found in (6.4). Likewise, the inverse f^{-1} of Möbius transformation $f(z) = \frac{az+b}{cz+d}$ is represented by

$$\begin{pmatrix} a & b \\ c & b \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

which, once scaled throughout by $-(ad - bc)$, matches the parameters found in (6.5).

In the above matrix representation, $(z_1, z_2)^T$ is identical to $\lambda(z_1, z_2)^T$ for any non-zero $\lambda \in \mathbb{C}$, since they both represent the same complex number z. This hints at the link between conformal maps and projective geometry (Schottenloher 2008).

Since each Möbius transformation has two normalised representations, each of which can be written as a 2×2 complex matrix with unit determinant, we can identify the Möbius group as $\mathrm{SL}(2, \mathbb{C})/\mathbb{Z}_2$. Alternatively, the Möbius

group can also be identified as $\mathrm{PSL}(2,\mathbb{C})$ via the aforementioned connection to projective geometry.

Although any given Möbius transformation is specified by 4 complex numbers (8 real numbers), the ability to scale them all by an arbitrary complex number reduces the number of real degrees of freedom to 6; we will see this again in another form below. In fact, the Möbius group is a 6-dimensional Lie group.

In the context of conformal field theories, it is sometimes claimed that the group of conformal symmetries on the 2 dimensional plane is infinite-dimensional; statements such as these more precisely refer to infinitesimal transformations and therefore an infinite dimensional Lie algebra, rather than the group itself (Schottenloher 2008).

6.4 Geometric properties of Möbius transformations

A general Möbius transformation $f(z) = \frac{az+b}{cz+d}$ with $ad - bc \neq 0$ can be decomposed into a few geometrically more intuitive operations. When $c = 0$, we have $f(z) = \alpha z + \beta$, where $\alpha = a/d$ and $\beta = b/d$. The requirement that $ad - bc \neq 0$ in this case ensures that $\alpha \neq 0$, so f does not reduce to the trivial constant function. The map $z \mapsto \alpha z$ is a combination of global scaling by $|\alpha|$ combined with a global rotation by $\arg \alpha$ in the anti-clockwise directions about the origin. Therefore in this case f is simply a global scaling and rotation, followed by translation.

For the case where $c \neq 0$, the Möbius transformation can be written as

$$f(z) = \frac{a}{c} + \frac{bc - ad}{c} \cdot \frac{1}{cz + d}, \tag{6.6}$$

which is a composition of scaling/rotation with translation, followed by inversion, and subsequently scaling/rotation and translation again. Recall that inversion refers to the map $z \mapsto 1/z$, which represents a rotation of the 2-sphere to which $\mathbb{C}_\infty$ is conformally related. We see that, again, the condition $ad - bc \neq 0$ ensures that $f(z)$ is not a constant function.

It is interesting to note that, on Euclidean spaces $\mathbb{E}^d$ with $d > 2$, conformal maps are also necessarily compositions of translations, rotations, scalings and inversions in $(d-1)$-spheres, as per Liouville's Theorem of conformal maps (Dubrovin, Fomenko, and Novikov 1992, §15). However, for $d > 2$, even when restricting to small sub-regions in $\mathbb{E}^d$ (i.e. looking at local properties), the conformal maps remain constrained to such compositions. Locally, the constraints on conformal maps for $d = 2$ (analyticity, with non-zero derivative) are much less restrictive.

While the geometric interpretation of scalings, rotations and translations on the plane are all clear, we should examine the inversion map $z \mapsto 1/z$ more closely. We have already seen that it represents a simple rotation of the sphere to which the extended complex plane is related.

Proposition 6.10. *The inversion $z \mapsto 1/z$ maps straight lines and circles to other straight lines and circles on $\mathbb{C}_\infty$.*

Proof. We follow Brown and Churchill (2009, §92). Let $z = x + iy$ and $w = u + iv$ for real x, y, u and v. Then if $w = 1/z$, also $z = 1/w$ so that

$$x = \frac{u}{u^2 + v^2} \qquad \text{and} \qquad y = \frac{-v}{u^2 + v^2}. \tag{6.7}$$

In general a circle or straight line on the z-plane can be written as

$$A(x^2 + y^2) + Bx + Cy + D = 0 \qquad \text{with } B^2 + C^2 > 4AD, \tag{6.8}$$

for real constants A, B, C and D.

In the case where $A = 0$, the condition reduces to $B^2 + C^2 > 0$, so that B and C are not simultaneously zero; then (6.8) represents a straight line.

In the case where $A \neq 0$, (6.8) can be rearranged into the form

$$\left(x + \frac{B}{2A}\right)^2 + \left(y + \frac{C}{2A}\right)^2 = \left(\frac{\sqrt{B^2 + C^2 - 4AD}}{2A}\right)^2,$$

representing a circle so long as the condition on the constants is satisfied.

By simply substituting (6.7) into (6.8) we find

$$D(u^2 + v^2) + Bu - Cv + A = 0 \qquad \text{with } B^2 + C^2 > 4AD,$$

which describes either a circle ($D \neq 0$) or a line ($D = 0$) in the w-plane. □

Considered on the sphere $\mathbb{S}^2$, the inversion is simply a rotation and so clearly maps circles on the surface $\mathbb{S}^2$ to other such circles. Proposition 6.10 can thus be interpreted as evidence that the stereographic projection between $\mathbb{S}^2$ and $\mathbb{C}_\infty$ maps circles on the one to circles on the other.

Projected to the sphere $\mathbb{S}^2$, straight lines are simply circles passing through the North pole **N**. For this reason, we interpret straight lines on $\mathbb{C}_\infty$ as circles that pass through ∞, i.e. circles with infinite radius. It is therefore common when working on $\mathbb{C}_\infty$ to define the term *generalised circle* (or alternatively, *circline*) to denote both circles and straight lines. With this, Proposition 6.10 can then be more easily stated: inversions map generalised circles to generalised circles.

Since rotations, scalings and translations are all similarity transformations of the plane, they also map circles to circles on $\mathbb{C}_\infty$. Thus from the decomposition (6.6) of a general Möbius transformation, we immediately obtain a corollary.

Corollary 6.11. *Any Möbius transformation maps generalised circles to generalised circles.*

An important motivating feature of the AdS/CFT conjecture is the fact that the isometry group of AdS_{d+1} is identical to the group of conformal symmetries of the d-dimensional boundary. We have restricted our attention to constant-time slices of AdS_4. In this restricted context, it is possible to demonstrate the relationship between the isometry group of the bulk and the conformal group of the boundary directly, by explicitly constructing a group isomorphism. The construction is geometric and very visual. While we shall prove neither the preliminary results that are needed to construct the map, nor that it is indeed the group isomorphism sought, it is nonetheless worth mentioning here, even if in a non-rigorous way.

We wish to relate the conformal diffeomorphisms of $\mathbb{C}_\infty$ to the isometries of the hyperbolic space $\mathbb{H}^3$. Recall that another representation of $\mathbb{H}^3$, besides the half-plane model we have used previously, is the Poincaré ball model. In this model, $\mathbb{H}^3$ is the open unit ball centred at the origin in $\mathbb{R}^3$ and equipped with a metric $g = \frac{4}{(1-|\mathbf{x}|^2)^2} \sum_i (\mathrm{d}x^i)^2$.

Then the desired group isomorphism is called the Poincaré extension, and is as follows (Carne 2012, p. 58; Parker 2007, §5.2). It can be shown that any Möbius transformation on $\mathbb{C}_\infty$ can be written as a composition of an even number of inversions in circles. So say a given Möbius transformation M can be written as $M = \sigma_{2N} \circ \ldots \circ \sigma_1$ for some natural number N, where σ_i are inversions in 1-spheres Σ_i. Consider the Riemann sphere $\mathbb{C}_\infty$ as the unit sphere embedded in $\mathbb{R}^3_\infty = \mathbb{R}^3 \cup \{\infty\}$ (a one-point compactification of $\mathbb{R}^3$ just as $\mathbb{C}_\infty$ is a one-point compactification of $\mathbb{R}^2$); then $\mathbb{C}_\infty$ makes up the boundary of the open ball representing $\mathbb{H}^3$. We can find 2-spheres J_i orthogonal to $\mathbb{C}_\infty$ such that $J_i \cap \mathbb{C}_\infty = \Sigma_i$. Define j_i as the inversions in the spheres J_i; then each j_i is a transformation of $\mathbb{R}^3_\infty$ whose restriction to $\mathbb{C}_\infty$ is just σ_i.

The composition $\widetilde{M} = j_{2N} \circ \ldots \circ j_1$ is then the Poincaré extension of M. When restricted to $\mathbb{C}_\infty$, it reduces to M. As well as that, it maps the unit ball (bounded by $\mathbb{C}_\infty$) to itself; in other words, it maps the Poincaré ball to itself. In fact, for any Möbius transformation M, the extension $\widetilde{M}$ turns out to be an isometry of the Poincaré ball.

6.5 Constructing particular Möbius transformations

Having found the general functional form of conformal symmetry transformations on the Riemann sphere, as well as some of their particular geometric properties, we want to use these symmetries to relate various regions on the boundary at infinity of a holographic theory. This will allow us to relate the entanglement entropies of those regions.

In order to accomplish this, we will need to be able to explicitly construct particular Möbius transformations to map between the desired regions. To

this end, we now explore further some properties of Möbius transformations that will be applicable in such constructions.

6.5.1 Determining a Möbius transformation by its action on specified points

Just as a polynomial of degree n on the real line can be entirely determined by its action on $n+1$ points, so too can a Möbius transformation be determined by its action on three points in $\mathbb{C}_\infty$.

Proposition 6.12. *Say M is a Möbius transformation that maps three distinct points p, q, r to p', q', r' respectively. Then M is the unique Möbius transformation which does so.*

Proof. Say N is another Möbius transformation such that $N(p) = M(p) = p'$, $N(q) = M(q) = q'$ and $N(r) = M(r) = r'$. Then $N^{-1} \circ M$ has three fixed points at p, q, r, i.e. $N^{-1} \circ M(z) = z$ for $z \in \{p, q, r\}$. But $N^{-1} \circ M$ is itself a Möbius transformation, so we may write it as

$$N^{-1} \circ M(z) = \frac{az+b}{cz+d} \qquad \text{with} \qquad ad - bc \neq 0.$$

Assume $c \neq 0$. Then fixed points $z = N^{-1} \circ M(z)$ satisfy

$$az + b = z(cz + d).$$

This is a quadratic equation in z, and so $N^{-1} \circ M$ has at most two distinct finite fixed points; since $N^{-1} \circ M(\infty) = a/c$, ∞ is not a fixed point. Thus if $c \neq 0$, $N^{-1} \circ M$ has at most two fixed points which is a contradiction.

Therefore $c = 0$ and we may write

$$N^{-1} \circ M(z) = \alpha z + \beta \qquad \text{where} \qquad \alpha \neq 0.$$

Then $N^{-1} \circ M(\infty) = \infty$, so that ∞ is a fixed point, while finite fixed points satisfy

$$\alpha z + \beta = z,$$

which has at most one solution (again a contradiction) unless $\alpha = 1$ and $\beta = 0$ so that every point is a fixed point. Then $N^{-1} \circ M(z) = z$, i.e. $N^{-1} \circ M$ is the identity. □

Since any Möbius transformation M is uniquely identified by its action on 3 chosen points, we may for convenience describe M by its action on 0, 1 and ∞. In this way, M is uniquely specified by the three complex numbers $M(0)$, $M(1)$ and $M(\infty)$. Again we see that the group of Möbius transformations has 6 real (i.e. 3 complex) degrees of freedom.

We introduce a useful notation based on this:

Definition 6.4 (Cross ratio)**.** The *cross-ratio* denoted $[z, p, q, r]$, where $z, p, q, r \in \mathbb{C}_\infty$, is the image of z under the unique Möbius transformation $M_{p,q,r}$ that sends $p \mapsto 0$, $q \mapsto 1$ and $\mapsto \infty$.

It is easy to explicitly construct $M_{p,q,r}(z)$ above. To map p to 0 and r to ∞ we take

$$M_{p,q,r}(z) = K\frac{z-p}{z-r}.$$

Then we choose the constant K to ensure that q is mapped to 1:

$$M_{p,q,r}(z) = \frac{(z-p)(q-r)}{(z-r)(q-p)}. \tag{6.9}$$

In the cases where p, q or r is ∞, different expressions are required:

$$M_{p,q,r}(z) = \begin{cases} \frac{q-r}{z-r} & \text{if } p = \infty \\ \frac{z-p}{z-r} & \text{if } q = \infty \\ \frac{z-p}{q-p} & \text{if } r = \infty \end{cases}. \tag{6.10}$$

We can use cross-ratios to explicitly construct the unique Möbius transformation M that takes p, q, r to p', q', r' respectively. With $M_{p,q,r}(z) := [z, p, q, r]$ and $M_{p',q',r'}(z) := [z, p', q', r']$, consider the map $M^{-1}_{p',q',r'} \circ M_{p,q,r}$. It takes $p \mapsto 0 \mapsto p'$, $q \mapsto 1 \mapsto q'$ and $r \mapsto \infty \mapsto r'$, so by Proposition 6.12 it is precisely M:

$$M(z) = M^{-1}_{p',q',r'} \circ M_{p,q,r}(z),$$

or

$$[M(z), p', q', r'] = [z, p, q, r]. \tag{6.11}$$

To explicitly find the formula for a given M, one can use the formulae (6.9) and (6.10) for the cross-ratios and invert the above.

Now recall that three points are sufficient to specify a circle on the plane; the same holds when extending to the Riemann sphere. Straight lines are specified by two finite points and (usually implicitly) infinity or by three collinear finite points, so we can continue to regard them as generalised circles.

Equation (6.11) therefore gives a direct way of constructing a Möbius transformation M that maps one chosen circle to another, simply by taking p, q and r on the first and p', q' and r' on the second. Of course, this does not mean that M is the only Möbius transformation that maps the first circle to the second; we can, after all, choose *any* three distinct p, q and r from the first, and likewise any p', q' and r' from the second; different such choices will give different maps.

As a final note on the topic of cross-ratios, we show that a cross-ratio is invariant under a Möbius transformation of all the points involved.

Theorem 6.13. *If $z, p, q, r \in \mathbb{C}_\infty$ are distinct points, and T is any Möbius transformation then*

$$[z, p, q, r] = [Tz, Tp, Tq, Tr].$$

Proof. Denote $M_{p,q,r}(z) = [z, p, q, r]$. Then by construction of the cross-ratio, $M_{p,q,r} \circ T^{-1}$ maps Tp, Tq and Tr to $0, 1$ and ∞ respectively. But this means that $M_{p,q,r} \circ T^{-1} = M_{Tp,Tq,Tr}$, so

$$[Tz, Tp, Tq, Tr] = M_{Tp,Tq,Tr}(Tz) = M_{p,q,r} \circ T^{-1}(T(z)) = M_{p,q,r}(z) = [z, p, q, r].$$

□

6.5.2 Preserving the unit disc

We will also find it useful below to know the functional form of Möbius transformations that map the closed unit disc $\bar{\mathbb{D}} = \{z \in \mathbb{C}_\infty : |z| \leq 1\}$ to itself.

Proposition 6.14. *Let M be a Möbius transformation that preserves the closed unit disc, i.e. $M(\bar{\mathbb{D}}) = \bar{\mathbb{D}}$. Then*

$$M(z) = \lambda \frac{z - \alpha}{\bar{\alpha} z - 1}, \tag{6.12}$$

where $\lambda, \alpha \in \mathbb{C}$ with $|\lambda| = 1$ and $|\alpha| < 1$.

Proof. We follow the argument presented in Chapling (2015). The boundary $\partial\bar{\mathbb{D}}$ of the closed disc is unit circle $\mathbb{S} = \{z \in \mathbb{C}_\infty : |z| = 1\}$, while the interior is the open disc $\mathbb{D} = \{z \in \mathbb{C}_\infty : |z| < 1\}$; so $\bar{\mathbb{D}} = \mathbb{D} \cup \mathbb{S}$. Since Möbius transformations are homeomorphisms, they map boundaries to boundaries and interiors to interiors. So M necessarily satisfies $M(\mathbb{S}) = \mathbb{S}$ and $M(\mathbb{D}) = \mathbb{D}$. In other words, $|M(z)| = 1$ whenever $|z| = 1$ and $|M(z)| < 1$ whenever $|z| < 1$.

Start with a general form of the Möbius transformation M:

$$M(z) = \frac{az + b}{cz + d}, \qquad ad - bc \neq 0.$$

If $d = 0$ then $M(0) = \infty$, which violates the restriction that $|M(z)| < 1$ whenever $|z| < 1$. So d is non-zero; we can rescale all the coefficients above to set $d = -1$:

$$M(z) = \lambda \frac{z - \alpha}{\beta z - 1}.$$

Whenever $|z| = 1$, we require that $|M(z)| = 1$, so

$$|M(z)|^2 = |\lambda|^2 \frac{|z - \alpha|^2}{|\beta z - 1|^2} = 1,$$

or

$$|\lambda|^2\left(|z|^2 - 2\Re(\bar{\alpha}z) + |\alpha|^2\right) = |\beta|^2|z|^2 - 2\Re(\beta z) + 1.$$

Using that $|z|^2 = 1$, this can be rearranged as follows:

$$2\Re\left([\beta - |\lambda|^2\bar{\alpha}]z\right) + |\lambda|^2(1 + |\alpha|^2) - 1 - |\beta|^2 = 0.$$

This must hold for all z with $|z| = 1$; so separately we must have

$$\beta = |\lambda|^2\bar{\alpha}$$

and

$$|\lambda|^2(1 + |\alpha|^2) - 1 - |\beta|^2 = 0.$$

Inserting the former into the latter gives a quadratic equation in $|\lambda|^2$, which has two solutions:

$$|\lambda| = 1/|\alpha| \qquad \text{or} \qquad |\lambda| = 1.$$

From the requirement that $M(\mathbb{D}) = \mathbb{D}$, we need $|M(0)| = |\lambda||\alpha| < 1$. This rules out the first of the above two solutions, and leaves us with $|\lambda| = 1$, $\beta = \bar{\alpha}$ and $|\alpha| < 1$. □

We will call Möbius transformations of the above form *disc-preserving* Möbius transformations. As with general Möbius transformations in Section 6.4, it is worth investigating the the geometric interpretation of this set of transformations. In (6.12), the factor λ simply represents a rotation about the origin, since $|\lambda| = 1$. So let us consider the map $z \mapsto \frac{z-\alpha}{\bar{\alpha}z-1}$ for some $|\alpha| < 1$.

Lemma 6.15. *For any non-zero $\alpha \in \mathbb{C}$, the function*

$$T_\alpha(z) = \frac{z - \alpha}{\bar{\alpha}z - 1}$$

maps the straight line L through the points α and 0 to itself.

Proof. The line L can be identified as the collection of points $z \in \mathbb{C}$ such that $\arg z = \arg\alpha$ modulo π. Take any z on L. It is geometrically clear that $z - \alpha$ is also on L. Additionally, since $\arg\bar{\alpha}z = \arg z - \arg\alpha = 0$ modulo π, the factor $\bar{\alpha}z - 1$ is real. Therefore $\arg T_\alpha(z) = \arg(z - \alpha) = \arg\alpha$ modulo π, i.e. $T_\alpha(z)$ also lies on L. □

So the map T_α for $|\alpha| < 1$ not only preserves the unit disc, but also the line (or diameter of the disc) passing through α.

These features are more elegantly understood through a geometric construction of T_α as a composition of the following: an inversion in the circle centred at $1/\bar{\alpha}$ and orthogonal to the unit circle at the origin, and a reflection through the straight line connecting the centres of those two circles (which is the same straight line as the α-containing diameter of the unit circle). We do not prove this fact here, but rather refer the interested reader to Needham (1998, §3.IX.3).

6.6 Examples of specific Möbius transformations, and applications to entanglement entropy

We now understand the conformal symmetry group of a constant time-slice of the boundary of AdS_4 well enough to construct useful symmetry transformations. With this, we can apply symmetry principles to the calculation of entanglement entropies.

Recall the somewhat general expression (4.17) for entanglement entropy of a region A on the 2-dimensional boundary slice:

$$S(A) = \frac{1}{4G_N^4}\left(\frac{\text{Area}(\partial A)}{\epsilon} + h(A) + \mathcal{O}(\epsilon)\right),$$

where $h(A)$ is some function of the geometry of A and ϵ is a cut-off parameter. We exclude any regions A with non-smooth boundaries here, and therefore any logarithmically divergent terms in $S(A)$, for simplicity. The leading divergent term seems manifestly to break the expected conformal invariance. One of the simplest possible conformal symmetry transformations, a constant scaling $z \mapsto \lambda z$, will clearly change $\text{Area}(\partial A)$. To reconcile this, recall that the conformal symmetry transformations on the boundary manifold in this holographic theory are naturally extended (via, for instance, the Poincaré extension mentioned in Section 6.4) to isometries of the bulk. The isometries will also act on the cut-off parameter ϵ, so that $\text{Area}(\partial A)/\epsilon$ as a whole ought to be invariant. Implicit in this assumption is that the regularisation by a cut-off ϵ is in fact a 'good' regularisation, so that the expansion in ϵ should preserved by symmetry transformations.

Therefore, say a symmetry maps $A \mapsto A'$ and $\epsilon \mapsto \epsilon'$. Under the assumptions above, we expect the entanglement entropies to be equal $S(A) = S(A')$, and in particular their finite parts also should be equal $h(A) = h(A')$.

To find the entanglement entropy $S(A)$ of a region from a previously known entanglement entropy $S(A')$ of symmetry-related region A', we must therefore simply express $h(A')$ in terms of the geometric parameters of A rather than A'. In the examples below, we both demonstrate the existence of regions A symmetric to some of the regions A' for which we have calculated entanglement entropies in Chapter 4, and find suitable expressions of the finite parts h to give the entanglement entropies of the new regions.

6.6.1 The half-plane and the circle

Using the tools of Section 6.5.1, we can construct a Möbius transformation that maps a disc A' of radius ρ to a half-plane A. For instance, we can build the Möbius transformation $M_{\rho,i\rho,-\rho}$ that takes points ρ, $i\rho$, $-\rho$ on the circumference of the circle of radius ρ centred at 0, to $0, 1, \infty$ on the real line:

$$M_{\rho,i\rho,-\rho}(z) = [z, \rho, i\rho, -\rho] = \frac{(z-\rho)(i\rho+\rho)}{(z+\rho)(i\rho-\rho)} = -i\frac{z-\rho}{z+\rho}.$$

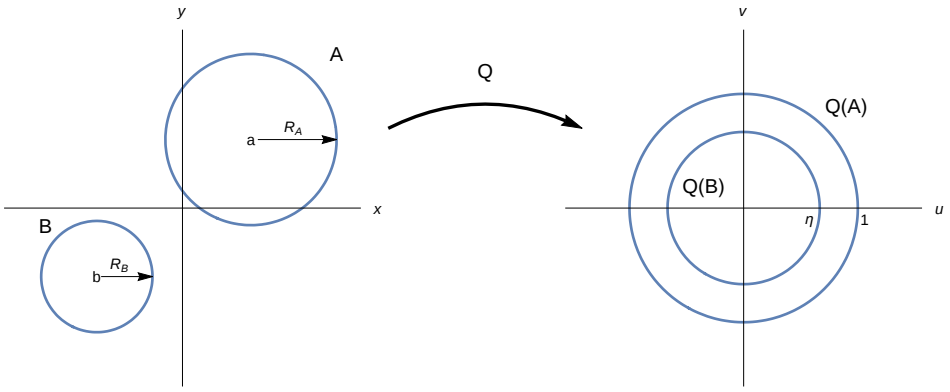


Figure 6.1: A Möbius transformation Q that maps two disjoint circles A and B to a concentric configuration, wherein the image $Q(A)$ of A is the unit circle and the image $Q(B)$ of B is a circle of smaller radius $\eta < 1$ also centred on the origin.

Since $M_{\rho,i\rho,-\rho}(0) = i$, we see that the interior of the unit disc is mapped to the upper half-plane.

In Section 4.2.1, we calculated the entanglement entropy (4.5) of a disc-shaped region of radius ρ:

$$S(A') = \frac{2\pi R^2}{4G_N^4}\left(\frac{\rho}{\epsilon'} - 1\right) = \frac{1}{4G_N^4}\left(\frac{\text{Area}(\partial A')}{\epsilon'} - 2\pi R^2\right).$$

In this case, the finite term $h(A') = 2\pi R^2$ is independent of the shape A'; therefore, the entanglement entropy of the half-plane A is simply

$$S(A) = \frac{1}{4G_N^4}\left(\frac{\text{Area}(\partial A)}{\epsilon} - 2\pi R^2\right)$$

as well. In this case the leading term is divergent not only due to the infinitesimal cut-off ϵ but also due to the infinite area of the boundary ∂A, which is an infinite straight line.

6.6.2 Disjoint discs and the annulus

A well-known application of Möbius transformations is to take two non-intersecting circles on $\mathbb{C}_\infty$ and transform them to a configuration of two concentric circles, as illustrated in Fig. 6.1. This will take the area outside the two circles and map it to an annular region.

Say we have two non-intersecting circles A, B with centres a, b and radii R_a, R_b respectively. We construct a Möbius transformation that maps A to the unit circle $\mathbb{S}^1$ and B to a circle of radius $\eta < 1$ centred at the origin. We

will use the following strategy to construct the desired Möbius transformation as a composition of three simpler ones:

1. Use a combined translation and scaling P to map A to the unit circle. Let $A' := P(A) = \mathbb{S}^1$ and $B' := P(B)$.

2. Let N be either an inversion in the unit circle $z \overset{N}{\mapsto} 1/z$ if B' lies outside $\mathbb{S}^1$, or the identity $z \overset{N}{\mapsto} z$ if B' is already inside $\mathbb{S}^1$. So $A'' := N(A') = \mathbb{S}^1$ remains fixed, and $B'' := N(B')$ is a circle somewhere inside the unit disc $\mathbb{D}$.

3. Find a disc-preserving Möbius transformation M as per Proposition 6.14 such that the circle $B''' := M(B'')$ has centre 0. The $A''' = M(A'') = \mathbb{S}^1$ is still fixed, while B''' is a smaller concentric circle of radius $\eta < 1$.

The composition $Q = M \circ N \circ P$ is the final desired Möbius transformation.

Finding M in particular requires some careful construction, as follows. Say the circle $B'' := N \circ P(B)$ has its centre at $\beta \neq 0$, where $N \circ P$ has ensured that $|\beta| < 1$. Then let L be the straight line through 0 and β. Let l_1 and l_2 be the points at which L intersects B'', and denote $u = \beta/|\beta| = \exp(i \arg \beta)$.

For any $r \in (-1, 1)$, the disc-preserving Möbius transformation

$$T_{ru}(z) = \frac{z - ru}{r\bar{u}z - 1}$$

also preserves the line L, as per Lemma 6.15. Therefore, the points $T_{ru}(l_1)$ and $T_{ru}(l_2)$ are at the intersections of L with the image circle $T_{ru}(B'')$.

If the image circle $T_{ru}(B'')$ is centred at 0, then necessarily those intersections with L must satisfy

$$T_{ru}(l_1) = -T_{ru}(l_2). \tag{6.13}$$

Imposing this condition is also sufficient to determine a particular value r_* of r; then we may take $M = T_{r_* u}$.

The procedure just described is general enough to handle any non-intersecting generalised circles A and B. This includes cases where one circle is already positioned inside the other but they are not concentric, or when B is a straight line (i.e. a circle through infinity).

For calculational simplicity, it is easiest to include in P an additional rotation so as to place the centre of $B' = P(B)$ on the positive real axis. With this, $B'' = N \circ P(B)$ will also be centred on the positive real axis; then $u = \pm 1$ and l_1 and l_2 will be real numbers, so that (6.13) will become a purely real equation for r.

Let us apply this procedure to a simple example. Let A be the circle of radius R_a centred at the origin, and let B be a circle of radius R_b centred at $b > R_a + R_b$ on the positive real axis. Evidently P only needs to be a

rescaling, $P(z) = z/R_a$, so $N \circ P(z) = R_a/z$. Let $z_1 = b - R_b$ and $z_2 = b + R_b$ be the two points on B that intersect the real axis. Then after applying $N \circ P$, we find

$$l_1 = N \circ P(z_1) = \frac{R_a}{b - R_b} \qquad \text{and} \qquad l_2 = N \circ P(z_2) = \frac{R_a}{b + R_b}. \tag{6.14}$$

As promised, l_1 and l_2 are real; further, $0 < l_2 < l_1 < 1$. Therefore, the direction u to use in the disc-preserver T_{ru} for the third map M is simply $u = +1$. To find M that centres the circle $N \circ P(B)$ on the origin while leaving the unit circle $N \circ P(A)$ in place, we must therefore solve

$$T_r(l_1) = -T_r(l_2),$$

or

$$\frac{l_1 - r}{rl_1 - 1} = -\frac{l_2 - r}{rl_2 - 1}$$

for $r \in (-1, 1)$.

This is a quadratic equation in r, with solutions

$$r = \left[\frac{1 + l_1 l_2}{l_1 + l_2}\right] \pm \sqrt{\left[\frac{1 + l_1 l_2}{l_1 + l_2}\right]^2 - 1}. \tag{6.15}$$

Now note that that, since $0 < l_2 < l_1 < 1$,

$$0 < (1 - l_1)(1 - l_2) = 1 - l_1 - l_2 + l_1 l_2,$$

so

$$\frac{1 + l_1 l_2}{l_1 + l_2} > 1.$$

Therefore, only one of the two solutions (6.15) is actually between -1 and 1:

$$r_* = \left[\frac{1 + l_1 l_2}{l_1 + l_2}\right] - \sqrt{\left[\frac{1 + l_1 l_2}{l_1 + l_2}\right]^2 - 1}. \tag{6.16}$$

This establishes the existence of the appropriate $M = T_{r_*}$ to give a Möbius transformation $Q = M \circ N \circ P$ that maps the original configuration of disjoint circles into concentric ones.

We can also adapt this procedure to the case where B is a straight line. With A as above, take B to be the vertical line given by $x = z_1 > R_a$. Then with $N \circ P(z) = R_a/z$ as above, the image of straight line B under $N \circ P$ is a circle passing through the origin, with its centre on the positive real axis. This circle intersects the positive real axis at R_a/z_1. Altogether, the case where B is a straight line is similar to the case where B is a circle, taking z_1 to be the position of the line and $z_2 = \infty$ so that $l_1 = R_a/z_1$ and $l_2 = 0$. The

third map $M = T_{r_*}$ is determined by the same formula (6.16) for r_*, although it greatly simplifies in this case.

We could proceed to substitute the values (6.14) of l_1 and l_2 into (6.16) to find the explicit forms of M and thus Q in terms of the original parameters of the circles (or circle and line). However, if we wish to use this Möbius transformation Q to relate the entanglement entropy $S(AB)$ of the two disjoint (generalised) discs to that of the concentric (annular) configuration, this will not be necessary. Recall from Section 4.2.3, and (4.14) in particular, that the finite term in the entanglement entropy of an annulus depends only on the ratio of its inner and outer radii. Since the outer radius of the annulus to which we have transformed is unity, this ratio is given simply by the inner radius η.

We must therefore find an expression for η in terms of the original geometric parameters (radii, centres) of A and B. We can use the invariance of the cross-ratio, Theorem 6.13, to determine η quite simply, as done by Nakaguchi and Nishioka (2015).

Consider the cross-ratio of the points $-R_a, R_a, z_1$ and z_2 where circles A and B intersect the real axis; by the invariance of the cross-ratio,

$$[-R_a, R_a, z_1, z_2] = [Q(-R_a), Q(R_a), Q(z_1), Q(z_2)].$$

We have already found $l_1 = N \circ P(z_1)$ and $l_2 = N \circ P(z_2)$, and by construction, $N \circ P(-R_a) = -1$ and $N \circ P(R_a) = 1$. So, since $Q = M \circ N \circ P$,

$$[-R_a, R_a, z_1, z_2] = [M(-1), M(1), M(l_1), M(l_2)]. \tag{6.17}$$

From the expression $M = T_{r_*}$, we see that $M(1) = -1$ and $M(-1) = 1$. We know that $M(l_1)$ is mapped to either $-\eta$ or η, and that $M(l_2) = -M(l_1)$; so to fix $M(l_1)$ and $M(l_2)$, we must simply determine the sign of one of them.

To do so, we rearrange (6.16) as follows:

$$r_* = \frac{1}{l_1 + l_2}\left(1 + l_1 l_2 - \sqrt{(1 - l_1^2)(1 - l_2^2)}\right).$$

Then since $0 < l_2 < l_1 < 1$,

$$\begin{aligned}
&\sqrt{1 - l_1^2} < \sqrt{1 - l_2^2}, \\
\implies &\sqrt{1 - l_1^2}\left(\sqrt{1 - l_1^2} - \sqrt{1 - l_2^2}\right) < 0, \\
\implies &1 - l_1^2 - \sqrt{(1 - l_1^2)(1 - l_2^2)} < 0, \\
\implies &1 + l_1 l_2 - \sqrt{(1 - l_1^2)(1 - l_2^2)} < l_1(l_1 + l_2),
\end{aligned}$$

the last step showing that $r_* < l_1$. By a similar argument, one can show that $r_* > l_2$ so that $0 < l_2 < r_* < l_1 < 1$.

From this, it is clear that $l_1 - r > 0$ and $rl_1 - 1 < 0$, so that $M(l_1) = (l_1 - r)/(rl_1 - 1) < 0$. Thus $M(l_1) = -\eta$, and $M(l_2) = +\eta$. Returning to the cross-ratio expression (6.17), this means that

$$[1, -1, -\eta, \eta] = [-R_a, R_a, z_1, z_2]. \tag{6.18}$$

Inserting (6.9) for the cross-ratios, we find

$$\frac{-4\eta}{(1-\eta)^2} = \frac{2R_a(z_2 - z_1)}{(z_1 + R_a)(z_2 - R_a)},$$

or with $z_1 = b - R_b$ and $z_2 = b + R_b$.

$$\frac{\eta}{(1-\eta)^2} = \frac{R_a R_b}{b^2 - (R_a + R_b)^2}. \tag{6.19}$$

This is quadratic in η, but has only one solution satisfying $0 < \eta < 1$:

$$\eta = \left[\frac{b^2 - R_a^2 - R_b^2}{2R_a R_b}\right] - \sqrt{\left[\frac{b^2 - R_a^2 - R_b^2}{2R_a R_b}\right]^2 - 1}. \tag{6.20}$$

We can also treat the case where B is a straight vertical line at $x = z_1$. Inserting $z_2 = \infty$ into (6.18) and recalling the expressions (6.10) for cross-ratios involving infinity, we find

$$\frac{\eta}{(1-\eta)^2} = \frac{1}{2(\frac{z_1}{R_a} - 1)}, \tag{6.21}$$

whose only solution for η between 0 and 1 is:

$$\eta = \left[\frac{z_1}{R_a}\right] - \sqrt{\left[\frac{z_1}{R_a}\right]^2 - 1}. \tag{6.22}$$

With the radius η fixed in terms of the parameters of the original geometric configuration of A and B, we can write the entanglement entropy of this configuration directly using the expression (4.14) found for the annulus:

$$S(AB) = \frac{1}{4G_N^4}\left(\frac{\text{Area}(\partial(A \cup B))}{\epsilon} - 4\pi R^2 F(\eta) + \mathcal{O}(\epsilon)\right),$$

where $F(\eta)$ is given by (4.15) and $\text{Area}(\partial(A \cup B)) = \text{Area}(\partial A) + \text{Area}(\partial B)$.

Knowing both $S(AB)$ of the region $A \cup B$, and $S(A)$ and $S(B)$ of the individual disjoint discs A and B, or disc A and half-plane B, we can construct the mutual information $I(A : B)$ between A and B:

$$I(A : B) = S(A) + S(B) - S(AB).$$

While the entanglement entropies are divergent, the mutual information has the benefit of being a finite physical quantity. Using our explicit expressions for $S(AB)$, $S(A)$ and $S(B)$, we have

$$\begin{aligned} I(A:B) &= \frac{1}{4G_N^4}\left[\left(\frac{\text{Area}(\partial A)}{\epsilon} - 2\pi R^2\right) + \left(\frac{\text{Area}(\partial B)}{\epsilon} - 2\pi R^2\right)\right. \\ &\qquad \left. - \left(\frac{\text{Area}(\partial A) + \text{Area}(\partial B)}{\epsilon} - 4\pi R^2 F(\eta) + \mathcal{O}(\epsilon)\right)\right] \\ &= \frac{4\pi R^2}{4G_N^4}\left[F(\eta) - 1\right]. \end{aligned}$$

The divergent terms evidently cancel so that the limit $\epsilon \to 0$ can be taken. When A and B are discs with radii R_a and R_b, and distance b between their centres, the parameter η is given by (6.20). When A is a disc of radius R_a and B is a half-plane a distance z_1 from the centre of A, the parameter η is given by (6.22).

The function $F(\eta)$ defined by (4.15) takes value 1 for $\eta \leq \eta_c \approx 0.416$; thus for $\eta \leq \eta_c$, the mutual information $I(A:B)$ vanishes.

Rearranging (6.19), we find that

$$b^2 = (R_a + R_b)^2 + \frac{(1-\eta)^2}{\eta} RaRb.$$

Since $(1-\eta)^2/\eta$ is monotonically decreasing in $\eta \in (0,1)$, the mutual information $I(A:B)$ between discs A and B vanishes for

$$b^2 \geq (R_a + R_b)^2 + \frac{(1-\eta_c)^2}{\eta_c} RaRb \approx (R_a + R_b)^2 + 0.8 R_a R_b.$$

Likewise, in the case of a disc and a half-plane we can rearrange (6.21) to give

$$\frac{z_1}{R_a} = 1 + \frac{1}{2}\frac{(1-\eta)^2}{\eta},$$

so that again the mutual information $I(A:B)$ between disc A and half-plane B vanishes when

$$\frac{z_1}{R_a} \geq 1 + \frac{1}{2}\frac{(1-\eta_c)^2}{\eta_c} \approx 1.4.$$

Three or more circles

Since we can move two non-intersecting circles into a convenient concentric configuration, it is reasonable to ask whether it would be possible to do the same for three or more circles. The answer, unfortunately, is that it is not.

When the circles bound three or more disjoint discs on the Riemann sphere, as shown in Fig. 6.2, this fact is a consequence of continuity (rather

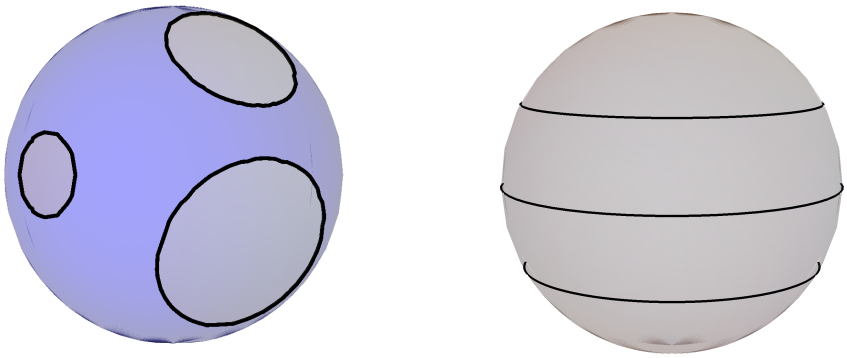


Figure 6.2: Comparison of a configuration of three circles bounding disjoint discs on the 2-sphere, and a configuration of three concentric circles on the same sphere. The blue region in the disjoint configuration is adjacent to all three disc regions, but there is no region in the concentric configuration that could be coloured blue that would be adjacent to all other regions.

than, for instance, conformality) of the Möbius transformations. Say we were to attempt to map n disjoint discs labelled D_1 to D_n into a concentric configuration. Then the initial configuration is of $n+1$ distinct regions in the Riemann sphere: the n discs as well as the single region exterior to all of them, which we label $R := \mathbb{C}_\infty \setminus \cup_{i=1}^{n} D_i$. Each of these is a connected region. The desired final configuration also consists of $n+1$ distinct, connected regions: two 'caps' of the Riemann sphere, and $n-1$ nested annular regions.

Recall, though, that a continuous transformation will map connected regions to connected regions, and any connected regions adjacent to each other in the initial configuration will again be adjacent in the image. In our initial configuration, each D_i is adjacent only to R. But in the desired final configuration, there is no single one of the $n+1$ target regions of interest that is adjacent to all others unless $n < 3$. So, no continuous transformation of $\mathbb{C}_\infty$ could map 3 or more disjoint discs into a concentric configuration.

If the initial configuration is of $n \geq 3$ circles that are already nested, then there is no obstruction on the grounds of continuity. Instead, we can compare the number of degrees of freedom available in the set of Möbius transformations to the number of conditions that need to be enforced to achieve a concentric configuration. Applying the necessary scaling and translation (operation P in the procedure described for two circles above) we can arrange the outermost circle as the unit circle centred at the origin. Then any subsequent Möbius transformations to arrange the smaller circles to be centred at the origin must

be disc-preserving. This means the only parameters we can tune in order to centre all $n-1$ remaining circles on the origin are the complex numbers λ and α of Proposition 6.14, where $|\lambda| = 1$ and $|\alpha| < 1$. But λ only adjusts an overall rotation about the origin, and our desired configuration is completely rotationally symmetric; so altering λ will not help. In effect we only have one complex parameter α, or equivalently two real parameters, to tune in order to centre $n - 1 \geq 2$ circles on the origin.

However, we already had to use both of these available real degrees of freedom to centre one circle – we had to take $\arg \alpha$ equal to the argument of the centre of the circle, and set $|\alpha|$ as per (6.13). So we have no remaining degrees of freedom available to tune the positions of any more circles.

Of course, this argument from degrees of freedom in the Möbius transformation does not guarantee that no cases exists where $n > 3$ nested circles can be made concentric. It simply shows that it is not generally possible for all such configurations.

Chapter 7

Conclusion

In this work, we have pursued a detailed understanding of the Ryu-Takayanagi formula for holographic entanglement entropy. In particular, we have prioritised mathematically clear definitions of the many objects and concepts that go into it: first entanglement, then entropy, and then entanglement entropy, restricted to a finite-dimensional quantum system so as to facilitate rigorous definition. We next presented the 'boundary at infinity' of a spacetime manifold via a conformal embedding, and constructed this embedding explicitly for AdS and other examples. While this background was sufficient for some simple applications of the Ryu-Takayanagi formula, these cases were heavily dependent on convenient choices of coordinate systems.

For a more general understanding of area-minimising surfaces, we turned to the mathematical theory of minimal surfaces embedded in general Riemannian ambient manifolds. We constructed the Second Fundamental Form of a submanifold in the most general case as the normal projection of the ambient covariant derivative; this allowed the definition of mean curvature via the trace of the Second Fundamental Form, which in turn allowed a general characterisation of locally area-minimising submanifolds as those with zero mean curvature. By reformulating the mean curvature in the particular context of the Ryu-Takayanagi prescription, we confirmed that this somewhat abstract characterisation does indeed generalise the more naïve approach with which we began.

Finally, we sought to understand the role of symmetries in the holographic entanglement entropy. Having found from our exploration of the boundary at infinity of anti-de Sitter spacetime that its constant-time slices are spheres, we sought to characterise the conformal diffeomorphisms that are the symmetry transformations of the boundary CFT. Using the tools of complex analysis on the extended complex plane, which is itself conformally related to the sphere, we showed these symmetries to be Möbius transformations. Utilising arguments based on such symmetries, we determined the entanglement en-

tropies of some new regions related by symmetry to those whose entanglement entropies could be calculated directly using the Ryu-Takayanagi formula.

While we have attempted to provide a thorough exposition of the structures involved in the holographic entanglement entropy, there nonetheless remain several pieces of missing information. In defining entanglement entropies, we restricted our attention to finite dimensional Hilbert spaces; however, quantum field theories (in particular, boundary components of holographic theories – the realm of the Ryu-Takayanagi prescription) have infinite-dimensional Hilbert spaces. It is clear from the divergence of entanglement entropies calculated in quantum field theories that no fully well-defined construction of entanglement entropies would be possible by simply extending the definitions to an infinite dimensional Hilbert space: the divergences in this physical quantity indicate that, fundamentally, the mathematical framework of the finite-dimensional case must be supplemented or replaced by some additional, as yet unknown, structure to be extended to the case of quantum field theory.

Our description of the boundary at infinity, or conformal boundary, of a spacetime was fairly operational in its approach, built on direct calculations in particular spacetimes of interest. In so doing, we left some important underlying issues unexamined. We saw, with the case of the Euclidean plane as example, that the conformal completion of a manifold seems sometimes to be related to the topological notion of compactification; in the case of anti-de Sitter spacetime, the conformal completion is not a full compactification of the entire spacetime, but only compactifies the spacelike directions. We note here that there are many different ways to compactify a topological space, such as the Stone-Čech compactification or the one-point compactification seen in the Euclidean case. In order for the notion of a conformal boundary to have sensible physical meaning, it must be unique: any given bulk spacetime should have one and only one conformal boundary manifold. Thus, we should ask what, precisely, the conditions are that we require of a particular conformal embedding of one manifold into another, in order to call it 'the' conformal completion of the first.

Throughout this work, we have only considered a pure AdS bulk, but there are many other instances of the AdS/CFT correspondence with different bulk spacetimes, and indeed there are other holographic theories entirely. We have attempted to define the concepts involved in the Ryu-Takayanagi prescription in a sufficiently general way that they can be applied in any context, but further insight may be gained by directly studying such examples.

In the course of our study of the role of symmetries, we referred to the link between isometries in the AdS bulk and conformal transformations on the boundary, and a particular method of constructing the appropriate group isomorphism in the special case of a constant-time slice of AdS via the Poincaré extension. A more thorough understanding of this link is needed,

not only without the restriction to particular time slices, but also in the context of other bulk spacetime manifolds as mentioned above.

We have studied primarily the constituents of the Ryu-Takayanagi prescription, rather than its physical implications, in this work. Clearly the latter is also of great interest, and is currently a topic of much research. It is the hope of the author that the current work will form a robust foundation from which to engage in this field.

Appendix A

Spherical coordinates in $\mathbb{E}^d$

Apart from the standard rectangular coordinates, Euclidean spaces $\mathbb{E}^d$ have another common family of coordinate systems. The 2-dimensional $\mathbb{E}^2$ has polar coordinates, while 3-dimensional $\mathbb{E}^3$ has both cylindrical and spherical polar coordinates. In this appendix, we present the generalisation of these coordinate systems to $\mathbb{E}^d$ for arbitrary dimension d.

Recall that with rectangular coordinates $\left(x^1, \ldots, x^d\right)$ of $\mathbb{E}^d$, each coordinate can take any real value: $x^i \in (-\infty, \infty)$ for all $i \in \{1, \ldots, d\}$; the metric is

$$g_{\mathbb{E}^d} = \sum_{i=1}^{d} (\mathrm{d}x^i)^2.$$

If we hold all x^i constant for $i \leq d - n$, the remaining n coordinates span an n-dimensional hyperplane, itself a copy of $\mathbb{E}^n$ isometrically embedded into $\mathbb{E}^d$.

We construct *n-spherical coordinates* $(x^1, \ldots, x^{d-(n+1)}, \rho_n, \phi_n, \ldots, \phi_1)$, where $x^i \in (-\infty, \infty)$ are identical to the respective rectangular coordinates, $\rho_n \in [0, \infty)$ is called the radial coordinate, and $\phi_1 \in [0, 2\pi)$ and $\phi_i \in [0, \pi]$ for $i \in \{2, \ldots, n\}$ are the angular coordinates. These are constructed so that, when all the rectangular coordinates x^i and the radial coordinate ρ_n are held constant, the angular coordinates ϕ_j span out an n-dimensional sphere of radius ρ_n. In particular, if $\rho_n = 1$ then this is the unit n-sphere $\mathbb{S}^n$.

Note that since the ranges of possible values of the radial coordinate ρ_n and angular coordinates ϕ_i are not open sets, the n-spherical coordinates are not formal coordinate maps (i.e. inverse maps of charts) of the manifold $\mathbb{E}^d$. To be precise, one could restrict the possible ranges of values of the radial and angular coordinates to obtain proper coordinate maps; in this case, one would need multiple such coordinate systems to cover $\mathbb{E}^d$ fully.

In this language, the polar coordinates of $\mathbb{E}^2$ are 1-spherical coordinates, while in $\mathbb{E}^3$ the cylindrical coordinates are 1-spherical coordinates and the usual spherical coordinates are 2-spherical coordinates.

We construct these coordinate systems inductively in n. First, we transform from rectangular coordinates to 1-spherical coordinates. Then, given n-spherical coordinates for $n < d-1$, we transform to $(n+1)$-spherical coordinates.

From rectangular to 1-spherical coordinates

Let

$$\rho_1 := \sqrt{(x^{d-1})^2 + (x^d)^2}. \tag{A.1}$$

If $\rho_1 = 0$ then the angular coordinate ϕ_1 is ill-defined – this is a direct consequence of 1-spherical coordinates not being well-defined formal coordinate maps of the manifold. Assuming $\rho_1 \neq 0$, define

$$\theta := \arccos\left(\frac{x^{d-1}}{\rho_1}\right),$$

so that $\theta \in [0, \pi]$ and $x^{d-1} = \rho_1 \cos\theta$. From (A.1), this gives $\left(x^d\right)^2 = (\rho_1 \sin\theta)^2$, or $\left|x^d\right| = \rho_1 \sin\theta$.

Whereas $\sin\theta$ is always non-negative because $\theta \in [0, \pi]$, the rectangular coordinate x^d may be positive or negative. Thus, together with ρ_1, the angle $\theta \in [0, \pi]$ is insufficient to fully specify x^d. To account for this, define

$$\phi_1 := \begin{cases} \theta & \text{if } x^d \geq 0, \\ 2\pi - \theta & \text{if } x^d < 0, \end{cases}$$

so that $\phi_1 \in [0, 2\pi)$, and

$$x^{d-1} = \rho_1 \cos\phi_1 \qquad \text{and} \qquad x^d = \rho_1 \sin\phi_1.$$

The Euclidean metric $g_{\mathbb{E}^d}$ in the 1-spherical coordinates becomes

$$\begin{aligned} g_{\mathbb{E}^d} &= \sum_{i=1}^{d} \left(\mathrm{d}x^i\right)^2 \\ &= \sum_{i=1}^{d-2} \left(\mathrm{d}x^i\right) + \left[\frac{\partial x^{d-1}}{\partial \rho_1}\mathrm{d}\rho_1 + \frac{\partial x^{d-1}}{\partial \phi_1}\mathrm{d}\phi_1\right]^2 + \left[\frac{\partial x^{d}}{\partial \rho_1}\mathrm{d}\rho_1 + \frac{\partial x^{d}}{\partial \phi_1}\mathrm{d}\phi_1\right]^2 \\ &= \sum_{i=1}^{d-2} \left(\mathrm{d}x^i\right) + \mathrm{d}\rho_1^2 + \rho_1^2 \mathrm{d}\phi_1^2. \end{aligned}$$

For the sake of the inductive construction, let us define notation $\mathrm{d}\Omega_1^2 := \mathrm{d}\phi_1^2$. In the inductive step below we will find a recursive formula for a symmetric covariant 2-tensor $\mathrm{d}\Omega_n^2$. Afterwards, we will look more closely at exactly what this tensor represents.

From $(n-1)$-spherical coordinates to n-spherical coordinates

Say we have $(n-1)$-spherical coordinates $\left(x^1, \ldots, x^{d-n}, \rho_{n-1}, \phi_{n-1}, \ldots, \phi_1\right)$, and say the metric tensor in such coordinates is

$$g_{\mathbb{E}^d} = \sum_{i=1}^{d-n} \left(\mathrm{d}x^i\right)^2 + \mathrm{d}\rho_{n-1}^2 + \rho_{n-1}^2 \mathrm{d}\Omega_{n-1}^2,$$

for some symmetric covariant 2-tensor $\mathrm{d}\Omega_{n-1}^2$.

Then define

$$\rho_n := \sqrt{\left(x^{d-n}\right)^2 + \left(\rho_{n-1}\right)^2}. \tag{A.2}$$

If $\rho_n \neq 0$, we can define the angular coordinate

$$\phi_n := \arccos\left(\frac{x^{d-n}}{\rho_n}\right),$$

so $\phi_n \in [0, \pi]$. Special cases to be aware of here are when $\phi_n = 0$ or $\phi_n = \pi$; both of these cases indicate that $\rho_{n-1} = 0$, or $x^i = 0$ for all $i > d-n$, which necessarily means that all ϕ_i for $i < n$ are ill-defined.

If $\rho_n = 0$ then necessarily $\rho_{n-1} = 0$ as well; this tells us that not only is ϕ_n ill-defined, but all ϕ_i for $i < n$ are again ill-defined as well.

From this and (A.2), we find

$$x^{d-n} = \rho_n \cos\phi_n \qquad \text{and} \qquad \rho_{n-1} = \rho_n \sin\phi_n. \tag{A.3}$$

Unlike the $n = 1$ case, here there is no sign ambiguity in the latter since we know $\rho_{n-1} \geq 0$; therefore, there is no need to redefine the angular coordinate over a larger range of values.

With this, we may transform the Euclidean metric from $(n-1)$-spherical coordinates to n-sphericals:

$$\begin{aligned} g_{\mathbb{E}^d} &= \sum_{i=1}^{d-(n+1)} \left(\mathrm{d}x^i\right)^2 + \left[\frac{\partial x^{d-n}}{\partial \rho_n}\mathrm{d}\rho_n + \frac{\partial x^{d-n}}{\partial \phi_n}\mathrm{d}\phi_n\right]^2 + \\ &\qquad + \left[\frac{\partial \rho_{n-1}}{\partial \rho_n}\mathrm{d}\rho_n + \frac{\partial \rho_{n-1}}{\partial \phi_n}\mathrm{d}\phi_n\right]^2 + \rho_{n-1}^2 \mathrm{d}\Omega_{n-1}^2 \\ &= \sum_{i=1}^{d-(n+1)} \left(\mathrm{d}x^i\right)^2 + \mathrm{d}\rho_n^2 + \rho_n^2\left[\mathrm{d}\phi_n^2 + \sin^2\phi_n \mathrm{d}\Omega_{n-1}^2\right]. \end{aligned}$$

If we define $\mathrm{d}\Omega_n^2 := \mathrm{d}\phi_n^2 + \sin^2\phi_n \mathrm{d}\Omega_{n-1}^2$, we have a concise expression for the metric in n-spherical coordinates:

$$g_{\mathbb{E}^d} = \sum_{i=1}^{d-(n+1)} \left(\mathrm{d}x^i\right)^2 + \mathrm{d}\rho_n^2 + \rho_n^2 \mathrm{d}\Omega_n^2.$$

This recursive construction can continue at most until $n = d - 1$. The $(d-1)$-spherical (or simply *spherical* in the more usual terminology) coordinates of $\mathbb{E}^d$ are then $(r, \phi_{d-1}, \ldots, \phi_1)$, where we have have used the conventional notation $r := \rho_{d-1}$ for the radial coordinate. The metric tensor is then written as

$$g_{\mathbb{E}^d} = \mathrm{d}r^2 + r^2 \mathrm{d}\Omega_{d-1}^2.$$

The metric tensor on the n-sphere $\mathbb{S}^n$

We have defined a symmetric covariant 2-tensor $\mathrm{d}\Omega_n^2$ recursively in n:

$$\mathrm{d}\Omega_n^2 := \mathrm{d}\phi_n^2 + \sin^2\phi_n \mathrm{d}\Omega_{n-1}^2 \qquad \text{and} \qquad \mathrm{d}\Omega_1^2 := \mathrm{d}\phi_1^2. \tag{A.4}$$

To see exactly what $\mathrm{d}\Omega_n^2$ represents, consider a n-sphere of radius R in $\mathbb{E}^d$. In n-spherical coordinates $\left(x^1, \ldots, x^{d-(n+1)}, \rho_n, \phi_n, \ldots, \phi_1\right)$, this surface is simply described by the constraints

$$\rho_n = R > 0 \qquad \text{and} \qquad x^i = x_0^i \quad \text{for all } i \in \{1, \ldots, d-(n+1)\},$$

where R and x_0^i are constants.

If we denote the natural inclusion of this surface into $\mathbb{E}^d$ by ι, then $\iota^*\left(\mathrm{d}x^i\right) = \iota^*\left(\mathrm{d}\rho_n\right) = 0$, so

$$\iota^* g_{\mathbb{E}^d} = R^2 \mathrm{d}\Omega_n^2.$$

In particular, when $R = 1$ this describes the unit n-sphere $\mathbb{S}^n$. Thus, in n-spherical coordinates, the metric of $\mathbb{S}^n$ is

$$g_{\mathbb{S}^n} = \iota^* g_{\mathbb{E}^d} = \mathrm{d}\Omega_n^2.$$

From rectangular to n-spherical coordinates

For completeness, we combine the transformations between rectangular and 1-spherical, and $(n-1)$-spherical and n-spherical coordinates, into a single transformation between rectangular and n-spherical coordinates. From (A.3), we can find that

$$\rho_i = \rho_n \prod_{j=i+1}^{n} \sin\phi_j,$$

so

$$x^{d-i} = \rho_n \cos\phi_i \prod_{j=i+1}^{n} \sin\phi_j \qquad \text{for all } i \in \{1, \ldots, n\},$$

and

$$x^d = \rho_n \prod_{j=1}^{n} \sin\phi_j.$$

where the 'empty' product is taken to be unity.

From (A.4), one can write the non-recursive formula for the metric tensor of $\mathbb{S}^n$ in n-spherical coordinates:

$$\mathrm{d}\Omega_n^2 = \sum_{i=1}^{n} \left(\prod_{j=i+1}^{n} \sin^2 \phi_j \right) \mathrm{d}\phi_i^2.$$

Bibliography

Aharony, O, Gubser, SS, Maldacena, J, Ooguri, H, and Oz, Y (2000). "Large N field theories, string theory and gravity". In: *Physics Reports* 323.3-4, pp. 183–386. DOI: 10.1016/S0370-1573(99)00083-6. arXiv: hep-th/9905111.

Allais, A and Mezei, M (2015). "Some results on the shape dependence of entanglement and Rényi entropies". In: *Physical Review D* 91.4, p. 046002. DOI: 10.1103/PhysRevD.91.046002. arXiv: 1407.7249.

Baez, J and Huerta, J (2010). "The algebra of grand unified theories". In: *Bulletin of the American Mathematical Society* 47.3, pp. 483–552. DOI: 10.1090/s0273-0979-10-01294-2.

Ballón Bayona, CA and Braga, NRF (2007). "Anti-de Sitter boundary in Poincaré coordinates". In: *General Relativity and Gravitation* 39.9, pp. 1367–1379. DOI: 10.1007/s10714-007-0446-y. arXiv: hep-th/0512182.

Bengtsson, I and Życzkowski, K (2017). *Geometry of Quantum States: An Introduction to Quantum Entanglement*. 2nd ed. Cambridge University Press. 637 pp. ISBN: 9781107026254. arXiv: 1612:07747. Preprint of Chap. 17.

Brown, JW and Churchill, RV (2009). *Complex Variables and Applications*. 8th ed. McGraw-Hill Higher Education. ISBN: 9780073051949.

Calabrese, P and Cardy, J (2009). "Entanglement entropy and conformal field theory". In: *Journal of Physics A: Mathematical and Theoretical* 42.50, p. 504005. DOI: 10.1088/1751-8113/42/50/504005. arXiv: 0905.4013.

Carmi, D (2015). "On the shape dependence of entanglement entropy". In: *Journal of High Energy Physics* 2015.12. DOI: 10.1007/JHEP12(2015)043. arXiv: 1506.07528.

Carmo, MP do (1992). *Riemannian Geometry*. Trans. by F Flaherty. 2nd ed. Mathematics: Theory & Applications. Birkhäuser Boston. ISBN: 9780817634902.

Carne, TK (2012). "Geometry and groups lecture notes". URL: https://www.dpmms.cam.ac.uk/~tkc/GeometryandGroups/GeometryandGroups.pdf (visited on 02/13/2017).

Carroll, SM (2004). *Spacetime and Geometry: An Introduction to General Relativity*. Addison Wesley. ISBN: 9780805387322.

Casini, H and Huerta, M (2007). "Universal terms for the entanglement entropy in 2+1 dimensions". In: *Nuclear Physics B* 764.3, pp. 183–201. DOI: 10.1016/j.nuclphysb.2006.12.012. arXiv: hep-th/0606256.

Casini, H, Huerta, M, and Myers, RC (2011). "Towards a derivation of holographic entanglement entropy". In: *Journal of High Energy Physics* 2011.5. DOI: 10.1007/JHEP05(2011)036. arXiv: 1102.0440.

Chapling, R (2015). "Automorphisms of the unit disc". URL: http://people.ds.cam.ac.uk/rc476/complexanalysis/autD.pdf (visited on 02/13/2017).

Conway, JB (1996). *Functions of One Complex Variable*. Vol. 2. 2 vols. Graduate Texts in Mathematics. Springer-Verlag. ISBN: 9780387944609.

Cover, TM and Thomas, JA (2006). *Elements of Information Theory*. 2nd ed. John Wiley & Sons. ISBN: 9780471241959.

Dekel, A and Klose, T (2013). "Correlation function of circular Wilson loops at strong coupling". In: *Journal of High Energy Physics* 2013.11. DOI: 10.1007/jhep11(2013)117. arXiv: 1309.3203.

Drukker, N and Fiol, B (2006). "On the integrability of Wilson loops in $AdS_5 \times S^5$: some periodic ansatze". In: *Journal of High Energy Physics* 2006.01, p. 56. DOI: 10.1088/1126-6708/2006/01/056. arXiv: hep-th/0506058.

Drukker, N, Gross, DJ, and Ooguri, H (1999). "Wilson loops and minimal surfaces". In: *Physical Review D* 60.12. DOI: 10.1103/physrevd.60.125006. arXiv: hep-th/9904191.

Dubrovin, BA, Fomenko, AT, and Novikov, SP (1992). *Modern Geometry - Methods and Applications*. Vol. 1: *The Geometry of Surfaces, Transformation Groups, and Fields*. Trans. by RG Burns. 2nd ed. 3 vols. Graduate Texts in Mathematics. Springer. ISBN: 9780387976631.

Engelhardt, N (2017). "Into the bulk: a covariant approach". In: *Physical Review D* 95.6, p. 066005. DOI: 10.1103/PhysRevD.95.066005. arXiv: 1610.08516 [hep-th].

Fonda, P, Giomi, L, Salvio, A, and Tonni, E (2015). "On shape dependence of holographic mutual information in AdS_4". In: *Journal of High Energy Physics* 1502.2, p. 5. DOI: 10.1007/JHEP02(2015)005. arXiv: 1411.3608.

Fonda, P, Seminara, D, and Tonni, E (2015). "On shape dependence of holographic entanglement entropy in AdS_4/CFT_3". In: *Journal of High Energy Physics* 2015.12, p. 037. DOI: 10.1007/JHEP12(2015)037. arXiv: 1510.03664.

Freedman, M and Headrick, M (2016). "Bit threads and holographic entanglement". In: *Communications in Mathematical Physics*. DOI: 10.1007/s00220-016-2796-3. arXiv: 1604.00354.

Gubser, SS, Klebanov, IR, and Polyakov, AM (1998). "Gauge theory correlators from non-critical string theory". In: *Physics Letters B* 428 (1–2),

pp. 105–114. DOI: 10.1016/S0370-2693(98)00377-3. arXiv: hep-th/9802109.

Hayden, P, Headrick, M, and Maloney, A (2013). "Holographic mutual information is monogamous". In: *Physical Review D* 87.4, p. 046003. DOI: 10.1103/PhysRevD.87.046003. arXiv: 1107.2940.

Headrick, M (2014). "General properties of holographic entanglement entropy". In: *Journal of High Energy Physics* 2014.3, p. 085. DOI: 10.1007/JHEP03(2014)085. arXiv: 1312.6717.

Headrick, M, Hubeny, VE, Lawrence, A, and Rangamani, M (2014). "Causality & holographic entanglement entropy". In: *Journal of High Energy Physics* 12.12, p. 162. DOI: 10.1007/JHEP12(2014)162. arXiv: 1408.6300.

Headrick, M and Takayanagi, T (2007). "A holographic proof of the strong subadditivity of entanglement entropy". In: *Physical Review D* 76.10. DOI: 10.1103/PhysRevD.76.106013. arXiv: 0704.3719.

Hirata, T and Takayanagi, T (2007). "AdS/CFT and strong subadditivity of entanglement entropy". In: *Journal of High Energy Physics* 2007.02, p. 42. DOI: 10.1088/1126-6708/2007/02/042. arXiv: hep-th/0608213.

Hubeny, VE, Rangamani, M, and Takayanagi, T (2007). "A covariant holographic entanglement entropy proposal". In: *Journal of High Energy Physics* 2007.07, pp. 062–062. DOI: 10.1088/1126-6708/2007/07/062. arXiv: 0705.0016.

Jaynes, ET (1957). "Information theory and statistical mechanics". In: *Physical Review* 106.4 (4), pp. 620–630. DOI: 10.1103/PhysRev.106.620.

Klebanov, IR, Nishioka, T, Pufu, SS, and Safdi, BR (2012). "On shape dependence and RG flow of entanglement entropy". In: *Journal of High Energy Physics* 2012.7, p. 001. DOI: 10.1007/JHEP07(2012)001. arXiv: 1204.4160.

Krtouš, P and Zelnikov, A (2014). "Minimal surfaces and entanglement entropy in anti-de Sitter space". In: *Journal of High Energy Physics* 2014.10. DOI: 10.1007/jhep10(2014)077. arXiv: 1406.7659.

Lawson Jr., HB (1978). *Lectures on Minimal Submanifolds*. Vol. 1. Monografias de matematica. Impa. ISBN: 9780914098188.

Lee, JM (1997). *Riemannian Manifolds: An Introduction to Curvature*. Graduate Texts in Mathematics. Springer New York. ISBN: 9780387982717.

— (2013). *Introduction to Smooth Manifolds*. 2nd ed. Graduate Texts in Mathematics. Springer New York. ISBN: 9781441999825.

Liu, H and Mezei, M (2013). "A refinement of entanglement entropy and the number of degrees of freedom". In: *Journal of High Energy Physics* 2013.4. DOI: 10.1007/jhep04(2013)162. arXiv: 1202.2070.

Maldacena, JM (1999). "The large N limit of superconformal field theories and supergravity". In: *International Journal of Theoretical Physics* 38.4, pp. 1113–1133. DOI: 10.1023/A:1026654312961. arXiv: hep-th/9711200.

Munkres, JR (2000). *Topology*. 2nd ed. Prentice Hall. ISBN: 9780131816299.
Nakaguchi, Y and Nishioka, T (2015). "Entanglement entropy of annulus in three dimensions". In: *Journal of High Energy Physics* 2015.4. DOI: 10.1007/JHEP04(2015)072. arXiv: 1501.01293.
Needham, T (1998). *Visual Complex Analysis*. Clarendon Press. ISBN: 9780198534464.
Nielsen, M and Chuang, I (2010). *Quantum Computation and Quantum Information*. 10th Anniversary Edition. Cambridge University Press. ISBN: 9780511994005.
Pacard, F and Pimentel, FAA (2004). "Attaching handles to constant-mean-curvature-1 surfaces in hyperbolic 3-space". In: *Journal of the Institute of Mathematics of Jussieu* 3.3, pp. 421–459. DOI: 10.1017/s147474800400012x.
Parker, JR (2007). "Hyperbolic spaces - the Jyväskylä notes". URL: http://maths.dur.ac.uk/~dma0jrp/img/HSjyvaskyla.pdf (visited on 02/13/2017).
Reed, M and Simon, B (1981). *Methods of Modern Mathematical Physics*. Vol. 1: *Functional Analysis*. Revised and enlarged edition. 4 vols. Academic Press, Inc. ISBN: 9780125850506.
Ryu, S and Takayanagi, T (2006a). "Aspects of holographic entanglement entropy". In: *Journal of High Energy Physics* 2006.08. DOI: 10.1088/1126-6708/2006/08/045. arXiv: hep-th/0605073.
— (2006b). "Holographic derivation of entanglement entropy from the anti-de Sitter space/conformal field theory correspondence". In: *Physical Review Letters* 96.18. DOI: 10.1103/PhysRevLett.96.181602. arXiv: hep-th/0603001.
Schottenloher, M (2008). *A Mathematical Introduction to Conformal Field Theory*. 2nd ed. Lecture Notes in Physics. Springer Berlin Heidelberg. ISBN: 9783540686255.
Shannon, CE (1948). "A mathematical theory of communication". In: *Bell System Technical Journal* 27.3, pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.
Srednicki, M (1993). "Entropy and area". In: *Physical Review Letters* 71.5, pp. 666–669. DOI: 10.1103/PhysRevLett.71.666. arXiv: hep-th/9303048.
Streater, RF and Wightman, AS (1964). *PCT, Spin and Statistics, and All That*. W.A. Benjamin, Inc.
Witten, E (1998). "Anti de Sitter space and holography". In: *Advances in Theoretical and Mathematical Physics* 2.2, pp. 253–291. DOI: 10.4310/ATMP.1998.v2.n2.a2. arXiv: hep-th/9802150.