

Chapter 2

General Relativity: Subtle Is the Lord

A certain king had a beautiful garden, and in the garden stood a tree which bore golden apples. These apples were always counted, and about the time when they began to grow ripe it was found that every night one of them was gone. [...] As the clock struck twelve he heard a rustling noise in the air, and a bird came flying that was of pure gold; and as it was snapping at one of the apples with its beak, the gardener's son jumped up and shot an arrow at it [...] it dropped a golden feather from its tail, and then flew away. [...] it was worth more than all the wealth of the kingdom: but the king said, 'One feather is of no use to me, I must have the whole bird.

The Golden Bird, The Brothers Grimm

In this chapter, we introduce the main ideas of Einstein's theory of General Relativity. We make precise some important terms that have been mentioned in the previous chapter, such as black hole and event horizon. For later use, we also introduce Penrose diagrams and anti-de Sitter spacetime. Finally we will briefly discuss the AdS/CFT correspondence, and the role of general relativity in that context. Some parts of this chapter consist of the author's own, perhaps biased and rather philosophical, opinion, on some aspects of general relativity.

2.1 What is General Relativity?

Devoting one's life to general relativity is definitely a labor of love, an almost irresponsible calling.

—Pedro G. Ferreira, “The Perfect Theory”.

General relativity is a theory of gravity. It models space and time together—an entity called “spacetime”—as a 4-dimensional smooth manifold equipped with a metric tensor (this can be generalized to arbitrary dimensions). The metric is equipped with a Lorentzian signature $(-, +, +, +)$, where “ $-$ ” denotesthe time direction.

The sign convention $(-, +, +, +)$, also dubbed the “east coast metric,” is usually preferred by relativists and mathematicians, whereas particle physicists mostly prefer to use $(+, -, -, -)$, also called the “west coast metric.” Physics is invariant under the choice of convention.¹ Of course, in pure mathematics, one can also consider a metric with arbitrary number of “temporal” and “spatial” directions: $(-, \dots, -, +, \dots, +)$. This is called a “semi-Riemannian” metric in general, although sometimes a Lorentzian metric is also referred to as “semi-Riemannian” (it is a special case of the latter). If the signs are all the same, the geometry is called Riemannian geometry in mathematics. Confusingly, physicists often call Lorentzian “Riemannian,” and call Riemannian “Euclidean.” In mathematics, having a Euclidean geometry would mean that it has no curvature at all, i.e., the metric can take the form $g = \delta_{ab} dx^a dx^b$, where δ_{ab} is the Kronecker delta, taking value 1 if $a = b$, and 0 otherwise.

Let us start with some quick review of basic differential geometry. This review also serves to set our notations. It is, however, not meant to be a self-contained introduction to differential geometry. Readers who are well-versed in differential geometry can feel free to skip to Sect. 2.1.2.

2.1.1 Differential Geometry in a Nutshell

Philosophy is written in this grand book the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures, without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth.

—Galileo Galilei

Consider a smooth, d -dimensional manifold M (not yet equipped with a metric tensor). We can define a smooth function f on M (Fig. 2.1).

Denote the set of all smooth functions on M by $\mathcal{F}(M)$. An alternative notation that is often used is $C^\infty(M)$. At each point $p \in M$, we can define a *tangent vector* to M at p as a real-valued function $V : \mathcal{F}(M) \rightarrow \mathbb{R}$ such that the map is

- (1) \mathbb{R} -linear: $V(af + bg) = aV(f) + bV(g)$, and
- (2) Leibnizian: $V(fg)|_p = V(f)g(p) + f(p)V(g)$,

for all $f, g \in \mathcal{F}(M)$, and $a, b \in \mathbb{R}$. Note that $\mathcal{F}(M)$ is therefore a commutative ring. The reason we have to define a tangent vector so abstractly is because the usual notion of tangent vector as little pointing arrow sticking out from a curve or surface does not make sense anymore if M is all there is, and there is no “outside” of M for the vector to “stick out into”.

Given a coordinate system $\xi = (x^1, x^2, \dots, x^d) : U \subset M \longrightarrow \mathbb{R}^d$ defined on a neighborhood of p , and $f \in \mathcal{F}(M)$, we define the derivation

¹This is, however, not necessarily true if the manifold is non-orientable; see Chap. 1, Sect. 7 of [1].

$$\frac{\partial f}{\partial x^a}(p) := \frac{\partial(f \circ \xi^{-1})}{\partial x^a}(\xi_p). \quad (2.1)$$

(Essentially, we are defining how to do calculus on a manifold, via what we already know—how to do calculus in \mathbb{R}^d .)

Then, the map

$$\frac{\partial}{\partial x^a} \Big|_p : \mathcal{F}(M) \longrightarrow \mathbb{R}, \quad (2.2)$$

such that

$$\frac{\partial}{\partial x^a} \Big|_p : f \longmapsto \frac{\partial f}{\partial x^a}(p), \quad (2.3)$$

is a tangent vector to M at point p .

Let $T_p M$ denote the set of all tangent vectors to M at p . It is called the *tangent space*, and it is a vector space over \mathbb{R} . If (x^1, x^2, \dots, x^d) is a coordinate system in some open set $U \subset M$ at p , then its coordinate vectors

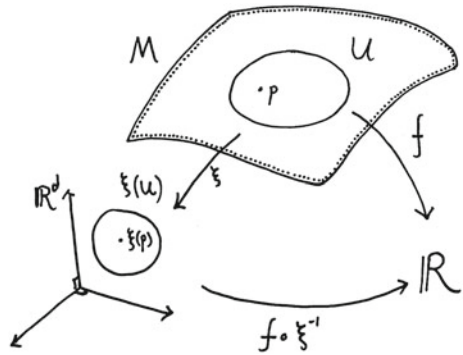
$$\left\{ \frac{\partial}{\partial x^1} \Big|_p, \frac{\partial}{\partial x^2} \Big|_p, \dots, \frac{\partial}{\partial x^n} \Big|_p \right\} \quad (2.4)$$

form a basis for the tangent space at $T_p M$, and one may expand

$$V = \sum_{i=1}^d V(x^i) \frac{\partial}{\partial x^i} \Big|_p, \quad \text{for all } V \in T_p M. \quad (2.5)$$

We often denote $V(x^a)$ by V^a . The tangent bundle TM is simply the disjoint union of all tangent spaces.

Fig. 2.1 An illustration of a smooth function f defined on a manifold M . Here ξ is a homeomorphism of an open set U of the manifold M onto an open set $\xi(U)$ in \mathbb{R}^d ; it maps a point $p \in U$ to $\xi(p) = (x^1(p), x^2(p), \dots, x^d(p))$



A map $f : M \rightarrow N$ between two manifolds induces the *pushforward* $f_* : T_p M \rightarrow T_{f(p)} N$, defined by

$$(f_*|_p V)(g) = V|_p(g \circ f), \quad \forall g \in \mathcal{F}(N). \quad (2.6)$$

Similarly, f also induces the *pullback* $f^* : T_{f(p)}^* N \rightarrow T_p^* M$.

Given any finite dimensional vector space \mathcal{V} , it has a *dual space* defined by

$$\mathcal{V}^* = \{\omega : \mathcal{V} \rightarrow \mathbb{R} \mid \omega \text{ is linear}\} = \text{Hom}(\mathcal{V}, \mathbb{R}). \quad (2.7)$$

Every member $\omega \in \mathcal{V}^*$ is called a 1-form, or traditionally a “covariant vector.” The dual of the tangent space at p is the *cotangent space*, denoted by $T_p^* M$. We have, for $\omega \in T_p^* M$, the (symmetric) scalar product, which does not require the notion of a metric:

$$\langle X, \omega \rangle := \omega(X) = X(\omega) \in \mathbb{R}, \quad (2.8)$$

where $X \in T_p M$.

Given a coordinate system $\xi = (x^1, x^2, \dots, x^d)$, one has an expansion for a 1-form

$$\omega = \sum_{i=1}^d \left\langle \omega, \frac{\partial}{\partial x^i} \right\rangle dx^i|_p, \quad \text{for all } \omega \in T_p^* M. \quad (2.9)$$

We often denote $\langle \omega, \frac{\partial}{\partial x^a} \rangle$ as ω_a . Note that $\langle dx^a, \frac{\partial}{\partial x^b} \rangle = \delta_b^a$, the Kronecker delta.

The disjoint union of all the cotangent spaces on M is the *cotangent bundle*, denoted by $T^* M$.

A type (r, s) tensor at $p \in M$ is a multilinear map

$$T : \underbrace{T_p^* M \times \dots \times T_p^* M}_r \times \underbrace{T_p M \times \dots \times T_p M}_s \longrightarrow \mathbb{R}. \quad (2.10)$$

We write $T \in \mathfrak{T}_s^r$.

We define the *tensor product* as follows:

$$T = T_1 \otimes T_2 \in \mathfrak{T}_q^p \otimes \mathfrak{T}_{q'}^{p'} \quad (2.11)$$

is an element of $\mathfrak{T}_{q+q'}^{p+p'}$ given by

$$T(\omega_1, \dots, \omega_p, \xi_1, \dots, \xi_{p'}; X_1, \dots, X_q, Y_1, \dots, Y_{q'}) \quad (2.12)$$

$$= T_1(\omega_1, \dots, \omega_p, X_1, \dots, X_q) T_2(\xi_1, \dots, \xi_{p'}, Y_1, \dots, Y_{q'}), \quad (2.13)$$

where $\{\omega_1, \dots, \omega_p, \xi_1, \dots, \xi_{p'}\}$ are 1-forms and $\{X_1, \dots, X_q, Y_1, \dots, Y_{q'}\}$ are vectors.

A (smooth) *vector field* X is a smooth assignment of each point $p \in M$ to a vector $V \in T_p M$. Let X, Y be two vector fields on $U \subset M$. Let φ_t be the flow with respect to X . Covector fields and tensor fields are defined similarly.

We define the *Lie derivative* of Y with respect to X by

$$\mathfrak{L}_X Y := \lim_{t \rightarrow 0} \frac{(\phi_{-t})_* Y - Y}{t} = \left. \frac{d}{dt} \right|_{t=0} (\phi_t^* Y). \quad (2.14)$$

The Lie derivative can be generalized to other tensor fields.

A *metric tensor* g on M is a symmetric bilinear non-degenerate² $(0, 2)$ -tensor field on M . More specifically, given any open subset $U \subset M$ and any smooth vector fields X, Y on U , the metric is the assignment

$$g(X, Y)(p) = g_p(X_p, Y_p) \in \mathbb{R}. \quad (2.15)$$

We write

$$ds^2 = g = g_{ab} dx^a \otimes dx^b = g_{ab} dx^a dx^b, \quad (2.16)$$

where

$$dx^a dx^b := \frac{1}{2} (dx^a \otimes dx^b + dx^b \otimes dx^a) \quad (2.17)$$

is the *symmetrized tensor product*.

We have used the “Einstein Summation Convention,”³ in which repeated indices—such that each index occurs once in a superscript and once in a subscript—are summed over.

In terms of a coordinate basis, g_{ab} can be written as a square $n \times n$ matrix, with inverse g^{ab} . One may use these to “raise” and “lower” indices of other tensors, e.g., $g^{ab} T_{bc} = T^a_c$, and $g_{ab} S^b_c = S_{ac}$.

An important difference between Riemannian and Lorentzian geometry is that the latter comes equipped with the notion of *causality*. A vector V is

- (a) *timelike*, if $g(V, V) < 0$,
- (b) *null*, or *lightlike*, if $g(V, V) = 0$, and
- (c) *spacelike*, if $g(V, V) > 0$.

A smooth curve γ is said to be timelike, null, or spacelike, respectively, if the tangent vector to the curve is timelike, null, or spacelike, respectively, at all points on γ . A curve is *causal* if it is either timelike or null.

²Though not positive definite in the Lorentzian case.

³I have made a great discovery in mathematics; I have suppressed the summation sign every time that the summation must be made over an index which occurs twice... – Albert Einstein [2].

Given two points p and q on a manifold, *a priori* there is no way to compare the vectors in $T_p M$ and $T_q M$ since they belong to different vector spaces. A natural way to make a comparison is to define parallel translation, which is a way to bring a vector in $T_p M$ along a curve to $T_q M$. To do this, we need an affine structure on the manifold. Let $E(M)$ denote the space of vector fields on M . A *connection* is the map

$$\nabla : E(M) \times E(M) \longrightarrow E(M), \quad (2.18)$$

written as $(X, Y) \longmapsto \nabla_X Y$, called the “covariant derivative of Y in the direction of X ,” satisfying

(1) Linearity over $\mathcal{F}(M)$ in the first argument:

$$\nabla_{fX_1 + gX_2} Y = f\nabla_{X_1} Y + g\nabla_{X_2} Y. \quad (2.19)$$

(2) Leibnizian in the second argument:

$$\nabla_X(fY) = f\nabla_X Y + (Xf)Y. \quad (2.20)$$

(3) Linearity over \mathbb{R} in the second argument:

$$\nabla_X(\alpha Y + \beta Z) = \alpha\nabla_X Y + \beta\nabla_X Z. \quad (2.21)$$

The operation ∇_X can be extended to tensors of any type, by requiring that $\nabla_X f = Xf$ and compatibility with contractions. For our purpose it suffices to assume that $E(M) = TM$. In general relativity, the connection chosen is the *Levi-Civita* connection. It is the unique connection satisfying both the metric compatible ($\nabla_c g_{ab} = 0$) and torsion-free conditions ($\nabla_X Y - \nabla_Y X = XY - YX$).

Note that property (3) implies that $\nabla_X(fY) \neq f\nabla_X Y$, so ∇_X is *not* a tensor!⁴ One has to be careful in distinguishing the statement “ $\nabla_X(\text{tensor})$ is a tensor,” which is true, and the statement that “ ∇_X is a tensor,” which is false.

Let $\{E_a\}$ be a local frame for TM on an open set $U \subset M$. We can expand in the basis to get

$$\nabla_{E_b} E_c = \Gamma^a_{bc} E_a, \quad (2.22)$$

where the Γ^a_{bc} ’s are the so-called *connection coefficients*, or the *Christoffel symbols*. Although it can be written in terms of the metric tensor, we emphasize that it does not require the metric to make sense. In terms of the metric, in a given coordinate system $\{x^a\}$, the connection coefficients are, in terms of the local frame $\{\partial/\partial x^a\}$,

⁴Of course, since $\nabla_X : Y \mapsto \nabla_X Y \in TM$, it is not a tensor in the strict sense of the word; but any tensor $T : T_p M \times T_p^* M \rightarrow \mathbb{R}$ can also be viewed naturally as a map $T : T_p M \rightarrow T_p M$. It is in this sense that ∇_X is not a tensor. If one wishes to be more accurate, one could say that ∇_X is not an endomorphism of the $\mathcal{F}(M)$ -module TM .

$$\Gamma_{bc}^a = \frac{1}{2} \left(\frac{\partial g_{cd}}{\partial x^b} - \frac{\partial g_{bc}}{\partial x^d} + \frac{\partial g_{bd}}{\partial x^c} \right) g^{da}. \quad (2.23)$$

An *affine geodesic* is a curve satisfying $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$. Note that this definition only depends on the connection, and not on the metric. If λ is the affine parameter, then the *geodesic equation* is

$$\frac{d^2 x^a}{d\lambda^2} + \Gamma_{bc}^a \frac{dx^b}{d\lambda} \frac{dx^c}{d\lambda} = 0. \quad (2.24)$$

In general relativity, the trajectories of the particles that are not subjected to exterior forces are precisely the geodesics (gravity is itself not considered as a force). Given a connection, one can define the *Riemann curvature endomorphism*, $Rm : TM \times TM \times TM \rightarrow TM$ by

$$R(X, Y)Z := \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z, \quad (2.25)$$

where $[X, Y] := XY - YX$ is the Lie bracket. It is equal to the Lie derivative: $[X, Y] = \mathcal{L}_X Y$. There is a natural isomorphism with the (1,3)-tensor (also denoted by Rm): $Rm : TM \times TM \times TM \times TM^* \rightarrow \mathbb{R}$. In terms of local coordinates,

$$Rm = R_{bcd}^a \frac{\partial}{\partial x^a} \otimes dx^b \otimes dx^c \otimes dx^d, \quad (2.26)$$

that is,

$$R_{bcd}^a = dx^a \left[R \left(\frac{\partial}{\partial x^b}, \frac{\partial}{\partial x^c} \right) \frac{\partial}{\partial x^d} \right]. \quad (2.27)$$

In terms of the connection coefficients,⁵ we have

$$R_{bcd}^a = \frac{\partial \Gamma_{bd}^a}{\partial x^c} - \frac{\partial \Gamma_{bc}^a}{\partial x^d} + \Gamma_{ec}^a \Gamma_{bd}^e - \Gamma_{ed}^a \Gamma_{bc}^e. \quad (2.28)$$

Note that we do not require a metric to define the Riemann curvature tensor.⁶ Among the many symmetries of the Riemann curvature tensor, one notes from the coordinate expression (2.28) that it is skew-symmetric with respect to $c \longleftrightarrow d$.

⁵There is unfortunately no accepted convention of the sign of the curvature tensor, or even which index is the one to be put “upstairs.” Exercise extreme caution when reading the literature.

⁶Spivak’s Volume 2 [3] has a nice explanation of the Riemann curvature tensor. Essentially, it comes about from an integrability condition for the existence of solution, when trying to solve for $g(\partial/\partial x^a, \partial/\partial x^b) = \delta_{ab}$.

Theorem (Riemann, 1861): The sufficient and necessary condition for a (semi)-Riemannian manifold (M, g) to be flat is the vanishing of the Riemann curvature tensor.

From the Riemann curvature tensor, we can obtain, by contraction, the *Ricci tensor*:

$$R_{ab} = R^c_{acb}. \quad (2.29)$$

Again, we note that there is no requirement of a metric tensor to define the Ricci tensor.

The Ricci tensor can be contracted again to obtain the *Ricci scalar*, or the *scalar curvature*:

$$R = g^{ab} R_{ab} = R^a_a. \quad (2.30)$$

However, note that this time, we *do* need a metric—to raise one of the indices before summing over them.

One can also define the *torsion tensor*:

$$T(X, Y) := \nabla_X Y - \nabla_Y X - [X, Y]. \quad (2.31)$$

In general relativity, just like in Riemannian geometry, the torsion vanishes identically by the choice of the (Levi-Civita) connection.

2.1.2 The Einstein Field Equations

There was a blithe certainty that came from first comprehending the full Einstein field equations, arabesques of Greek letters clinging tenuously to the page, a gossamer web. They seemed insubstantial when you first saw them, a string of squiggles. Yet to follow the delicate tensors as they contracted, as the superscripts paired with subscripts, collapsing mathematically into concrete classical entities—potential; mass; forces vectoring in a curved geometry—that was a sublime experience. The iron fist of the real, inside the velvet glove of airy mathematics.

—Gregory Benford, “Timescape”.

Mathematically, general relativity is just Lorentzian geometry subjected to the constraint of the *field equations*:

$$\boxed{R_{ab} - \frac{1}{2}g_{ab}R + \Lambda g_{ab} = \frac{8\pi G}{c^4}T_{ab}}, \quad (2.32)$$

where R_{ab} are the components⁷ of the Ricci tensor, g_{ab} are components of the metric tensor, R is the scalar curvature, Λ is a possibly nonzero cosmological constant term (which we have not included in Chap. 1 when we first showed the field equations.), and T_{ab} are the components of the energy–momentum tensor. Mathematicians usually prefer to write the field equations as

$$\text{Ric}_g - \frac{R}{2}g + \Lambda g = 0. \quad (2.33)$$

One often assumes that the energy–momentum tensor should satisfy some nice, “realistic,” properties. These are known as the *energy conditions*:

- (1) **Weak Energy Condition (WEC)** For all future-pointing timelike vector fields V , the matter density observed by the corresponding observer is always non-negative,

$$T_{ab}V^aV^b \geq 0, \quad (2.34)$$

- (2) **Strong Energy Condition (SEC)** For all future-pointing timelike vector fields V ,

$$\left(T_{ab} - \frac{1}{2}Tg_{ab}\right)V^aV^b \geq 0, \quad (2.35)$$

- (3) **Null Energy Condition (NEC)** For all future-pointing null vector fields K ,

$$T_{ab}K^aK^b \geq 0, \quad (2.36)$$

- (4) **Dominant Energy Condition (DEC)** In addition to requiring that the NEC holds, one also requires that for every future-pointing causal vector field (i.e., either timelike or null) Y , the vector field $-T_b^aY^b$ must be a future-pointing causal vector.

In terms of a fluid with energy density ρ and pressure p , these conditions read, respectively,

- (1) WEC: $\rho \geq 0$, $\rho + p \geq 0$,
- (2) SEC: $\rho + p \geq 0$, $\rho + (d-1)p \geq 0$,
- (3) NEC: $\rho + p \geq 0$, and
- (4) DEC: $\rho \geq |p|$.

Among all the four, the NEC is the weakest energy condition. Note also that the SEC does *not* imply the WEC. All these energy conditions are known to be

⁷Sometimes, we carelessly refer to R_{ab} as *the* Ricci tensor, or that g_{ab} is the metric tensor, instead of the *components* of these tensors in a particular basis. In the “abstract-index notation,” they are actually referring to the tensors themselves. However, such practice can be confusing to beginners. For example, it may cause people to ask whether “coordinate x^a is a vector” (c.f. V^a , the components of a vector $V = V^a \frac{\partial}{\partial x^a}$). My humble opinion is: learn the geometric objects properly, and then go ahead and abuse the notations all you want—but not before you know what you are doing!

violated, especially by quantum systems. However, there are other weaker, averaged versions of the energy conditions. We shall not go into the details here. The readers are encouraged to refer to [4] for detailed discussions.

Usually, in physics, one prefers to derive the equation of motion from the “action.” An action is a curious beast.⁸ In classical mechanics, the total energy of a physical system is the sum of its kinetic energy T and potential energy V ; one could also define the difference between the kinetic energy and potential energy, and it is called the “*Lagrangian*”: $L(x, \dot{x}, t) = T - V$. The action is then the integral

$$S = \int L \, dt. \quad (2.37)$$

A variation of the action $\delta S = 0$ then gives the Euler–Lagrange equation,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0, \quad (2.38)$$

which in this case is equivalent to Newton’s force law $F = ma$. In the case of general relativity, the action is known as the *Einstein–Hilbert action*. It is given by (restoring c and G),

$$S = \frac{c^4}{16\pi G} \int d^4x \sqrt{-\det(g)} (R - 2\Lambda), \quad (2.39)$$

where Λ is a possibly nonzero cosmological constant term. Note that the overall sign of the action depends on the choice of sign convention. From this action, one can derive the Einstein field equations.

Although not often emphasized in a typical course in general relativity (hereinafter, “GR”), it is crucial—at least for someone who wants to work in gravitation—to understand the layers of mathematical structures used in GR: Briefly, we have, starting from the most basic structure,

- (1) **Topological Manifold** Only topology is introduced at this level. Note, in particular, that there is not yet the concept of a metric. (See, e.g. [5])
- (2) **Differentiable/Smooth Manifold** Differentiable structure is added; this means we can do calculus on a manifold. (See, e.g. [6])
- (3) **Smooth Manifold with Connection** We can define parallel translation; note that with a connection we can already define curvature and torsion—these quantities do not require a metric to be defined. (See, e.g. [6, 7])
- (4) **Lorentzian Manifold** We can introduce a metric such that it is compatible with the connection, torsion-free, and has signature $(-, +, +, +)$. Note that the metric is required to define the scalar curvature, $R = g^{ab} R_{ab}$. (See, e.g. [6, 7])

⁸I chose to explain what an action is, albeit very briefly and not very rigorously, because a mathematician reader may not be familiar with this concept.

- (5) **Physics** Physics only arises when we introduce either an action (“physics is where the action is”) or the field equations. (A good mathematically rigorous text of general relativity is [8]). (See also [9])

Understanding the hierarchy of these structures becomes even more important when one wants to contemplate an alternative theory of gravity, since then one will need to know what can actually be modified—modified gravity is not just about modifying the action, the mathematical structures can also be chosen differently! Indeed, despite the fact that general relativity has been a very successful theory, physicists are still not satisfied; with major mysteries such as the dark energy problem still unsolved, many modified theories of gravity have been proposed. One golden feather is not enough; we need more. In this quest, however, it is important to check consistencies of a proposed theory, not just fitting observational data. After all, we need to make sure that the feather is really of genuine gold.

Having said all this about the different layers of mathematical structures, it may be prudent to quote Carl G. Hempel at this point:

[...] to characterize the import of pure geometry, we might use the standard form of a movie-disclaimer: No portrayal of the characteristics of geometrical figures or of the spatial properties of relationships of actual bodies is intended, and any similarities between the primitive concepts and their customary geometrical connotations are purely coincidental.

Despite how we can construct mathematical structures, we should not confuse said constructions with the real physical things.

2.2 Some Subtleties of General Relativity

Subtle is the Lord, but malicious He is not.

—Albert Einstein

General relativity is a rich theory. As Ferreira puts it, “despite being around for almost a century, it continues to yield new results” [10]. In addition, it is also full of subtleties, which are not often appreciated. In this section, we mention some of these remarkable facts, followed by the explanations.

- (1) Energy is, in general, *not* conserved.
- (2) Gravity is not a force, and in fact, can be source-free.
- (3) The Schwarzschild singularity lies in the future, and it is not a location in space.
- (4) The event horizon is not a special place, but this does not imply that there is no way for anyone to realize that he or she has crossed one.
- (5) Whether a freely falling observer detects Hawking radiation from a black hole actually depends on the geometry of the black hole—for asymptotically flat Schwarzschild black holes, the freely falling observer *does* detect particle creation.

Energy conservation is an important physical concept that has been deeply rooted in our thinking ever since high school. It is therefore not surprising that one may take it for granted that it *has* to be true. However, there is actually a deep reason behind energy conservation, which is given by the well-known Noether Theorem [11]: it corresponds to a time translational symmetry. In an expanding universe, for example, dark energy density can remain constant, because the system simply has no time translational symmetry (the universe is bigger now than it was yesterday). In fact, the field equations of general relativity satisfy $\nabla_a T^{ab} = 0$, which is sometimes referred to as a “conservation law.” However, this is different from the usual conservation law in field theory: $\partial_a T^{ab} = 0$. The covariant derivative telling us that the energy–momentum of the material field is exchanging energy in some precise way with the gravitational field. One could of course interpret this to be conservation of energy, provided that one takes into account the energy of the gravitational field itself. Unfortunately, gravitational energy cannot be localized (see, e.g., [12, 13]), and so it is difficult to make this statement precise. It is best to interpret $\nabla_a T^{ab} = 0$ as a *non-conservation* law.

Furthermore, as we have just mentioned, the Noether theorem relates a continuous symmetry with a conservation law. In general relativity, symmetry means isometry. Given a Killing vector,⁹ there corresponds a conservation law of some kind. Specifically, if ξ^a is the component of a Killing vector field, then the conservation law associated with the symmetry generated by ξ is

$$\nabla_a (\xi^b T^c_b) = 0. \quad (2.40)$$

Note that given a spacetime manifold, the existence of a Killing vector is not automatic. Therefore, symmetry is generically a rare thing in general relativity.

The second item on our list concerns the interpretation of gravity as geometry instead of as a force in GR. This is reflected in the fact that a free-falling particle follows a geodesic of the underlying geometry, so they are not accelerated. The point that gravity can be source-free is less appreciated. The easiest way to see this is that the *vacuum* field equation ($T_{ab} = 0$, which in turn implies that $R_{ab} = 0$) can nevertheless have non-trivial solutions in GR. The Schwarzschild solution is such an example. This is of course well known. However, it is in relation to item (3) that misconceptions can arise. Perhaps due to misleading cartoon depictions (and popular science descriptions along the line “you will be falling closer and closer towards the singularity, where all known laws of physics break down”), it is often thought that the singularity inside a Schwarzschild black hole is a location in space, namely, it is at the center of the spherical black hole. This is incorrect. The singularity of the Schwarzschild black hole is spacelike (see Sect. 2.5), which means that it lies *in the future*, and cannot be interpreted as the “center” of the black hole. One can of course define an effective center of mass for a Schwarzschild black hole, but this should not

⁹A vector field X is called a *Killing vector field* with respect to the metric g if the Lie derivative $\mathcal{L}_X g = 0$. Essentially, this is saying that the geometry on the manifold M determined by the metric g does not change if we move along the flow of the Killing vector field.

be mistaken as the statement that the mass is concentrated at the “central singularity.” See [14] for further discussion. The reason that one cannot escape the fate of hitting the singularity is the same as the reason that one cannot escape the next dreadful Monday from arriving—they are in the future of the observer.

For what it is worth, let us also mention that the event horizon is *not* a place in the ordinary sense of the word. The event horizon is an outgoing null surface—it is moving radially outward *at the speed of light* with respect to a local freely falling observer (thus the usual statement that one can only escape a black hole by traveling faster than light). The curvature of spacetime is such that, for observers far away from the black hole, the event horizon looks like a nice, static sphere.

The last item on the list emphasizes the importance of being clear about what one means in physics. It is often claimed that for a sufficiently large black hole, the tidal force at the event horizon can be so small that anyone who crosses the event horizon feels nothing out of the ordinary. This is correct. However, it does *not* mean that there is no way to know where the horizon is. One may be tempted to think that knowing the mass of the hole gives us the location of the horizon immediately via the Schwarzschild radius $r = 2M$. However, r is nothing more than a coordinate, and that does not help one to know physically where the horizon is. Nevertheless, there are other quantities that one can measure, which would betray the presence of the event horizon. One such example is the Karlhede scalar [15], constructed from the contraction of the covariant derivative of the Riemann curvature tensor:

$$\mathcal{I} = \nabla_e R_{abcd} \nabla^e R^{abcd}. \quad (2.41)$$

This geometric quantity changes sign at the event horizon *for some black holes*. For example, for a Schwarzschild black hole, its Karlhede scalar is

$$\mathcal{I}[\text{Sch}] = -\frac{720M^2(r - 2M)}{r^9}. \quad (2.42)$$

(For a Kerr black hole, the Karlhede scalar changes sign at the ergosphere, *not* at the horizon; however, there are other invariants that do detect the event horizon in the Kerr geometry [16, 17]). As a comparison, note that the usual Kretschmann scalar, on the other hand, is just the contraction of the Riemann curvature tensor and does not change sign at the event horizon. For a Schwarzschild black hole, it is

$$R_{abcd} R^{abcd} = \frac{48M^2}{r^6}. \quad (2.43)$$

However, note that the firewall controversy has *nothing* to do with the ability to detect the event horizon (despite the claim in [18]), the “no drama” statement in the paradox concerns the applicability of effective field theory at the event horizon of the black hole (since for large black holes, the curvature there is small, and thus EFT should apply).

As for item (5) regarding detectability of Hawking radiation, we already discussed this in Sect. 1.3.

2.3 Is the Metric Just Another Field?

Either, therefore, the reality which underlies space must form a discrete manifold, or we must seek the ground of its metric relations (measure conditions) outside it, in binding forces which act on it [...] This leads us into the domain of another science, of physics, into which the object of this work does not allow us to go today.

—Bernhard Riemann, “On the Hypotheses which lie at the Bases of Geometry”.

The metric tensor $g = g_{ab}dx^a dx^b$ measures “distance” on a manifold. This is particularly clear in the Riemannian case, as the *length* ℓ of a parametrized curve $\lambda \mapsto x(\lambda)$, joining two points $\lambda = a$ and $\lambda = b$, is just defined by the integral

$$\ell := \int_a^b \left[g_{ij}(x(\lambda)) \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} \right]^{\frac{1}{2}} d\lambda, \quad (2.44)$$

and the *distance* between these two points is just the lower bound of the length of all possible curves that join them.

In modern field theories, however, one is accustomed to the notion that everything there is, is just a quantum field, and particles are nothing but excitations of the field. This has led to the interpretation that the metric is just a spin-2 field, analogous to a spin-0 scalar field or spin-1 vector field. The excitation in the spin-2 field then gives rise to particles that mediate gravitational interaction—the gravitons.¹⁰ For the spin-2 tensor theory to be mathematically consistent, it can only be either free (no matter coupling), or in need of infinitely many correction terms, by including “gravitational energy-momentum” into the field equations (essentially, the field is coupled to its own energy-momentum, a process which iterates indefinitely). Presumably, if one does this correctly, after summing up all the terms, it is possible to recover general relativity (perhaps including higher order terms) in all its glory. However, “gravitational energy-momentum” is not even well defined (they are called “pseudo-tensors” for a reason) and there are many ways to define such objects, so it is not even clear if the result is unique. We shall not explore these issues further, for a nice discussion, see Chap. 3 of [19]. See also the objection raised in [20], but also the counter-arguments in [21].

If one were to take the point of view that the metric is “just another field,” then it is quite natural to consider the possibilities that the associated particle, namely the graviton, may actually be massive (in standard GR, interpreted as a field theory, one could see that graviton is massless). This gives rise to a class of theories that go under

¹⁰In linearized gravity, the metric tensor has components $g_{ab} = \eta_{ab} + h_{ab}$, where η is the background (here, flat) metric, and h is the perturbation (in the solar system, $|h_{ab}| \sim GM_{\odot}/(R_{\odot}c^2) \sim 10^{-6}$). Roughly speaking, quantization of $h_{\mu\nu}$ gives the graviton—it is an excitation of the background metric field.

the name “massive gravity” (for reviews, see [22, 23]). In addition, if gravity is “just another field,” then why cannot there be more than one such field? This gives rise to the bimetric [24], and even multi-metric theories [25] in recent years. However, none of these ideas are anything but natural if we consider GR as a geometric theory. In fact, massive gravity is plagued with many issues, including problems with the laws of black hole thermodynamics [26], as well as superluminality and micro-acausality—closed timelike curves can arise in arbitrarily small regions in spacetime [27–30]. It is not clear if bimetric and multi-metric theories can be free of such pathologies [31].

My humble opinion is that perhaps it is mistaken to view GR as just some dynamical theory that arises out of treating gravity as nothing more than the interaction of gravitons. Perhaps quantization of gravity is hard because we are trying to quantize the wrong thing; perhaps we should take geometry seriously and re-consider the question: what does quantized geometry really mean?¹¹

Indeed, a similar viewpoint that gravity is different from other fields has been mentioned in the literature. For example, H.A. Buchdahl stated that [32]:

I have spoken of ‘energy’, for instance. The energy of what? Of the field perhaps? Field?—We have no field in the sense in which one has a Maxwell field. Whenever I have used the term ‘field’ [in the context of the metric] I have done so as a matter of mere verbal convenience. [...] The classical fields—the electrostatic field, for example—in the first instance had, so to speak, a subjunctive existence. Let P be a particle satisfying all the criteria of being free except in as far as it carries an electric charge. Then if P were placed at some point it would be subject to a force depending on the value of the field intensity there. In particular, when the latter is zero there is no force. [...] the ‘true’ fields subjunctively quantify the extent to which a given particle P is not free, granted that P would be free if any charges it may carry were neutralized [...] this field [the metric], unlike the ‘true’ fields, cannot be absent, cannot be zero.¹²

On page 5 of [33], Hawking also mentioned that, if singularities in general relativity can indeed be smeared out by quantum correction, things would be rather “boring” since gravity would then be “just like any other field,” whereas gravity should be distinctively different since it is not just a player on a spacetime background; it is both a player *and* the evolving stage.

As a side remark, and on similar note, it has often been claimed that geometric quantities like torsion¹³ can be treated just like any other field, and thus there is no need to go beyond the Levi-Civita connection of GR. Such a proposition, while not

¹¹There was one thing that really riled many of the general relativists about string theory: in string theory [...] the geometry of spacetime, the be-all and end-all of general relativity, seemed to disappear. It was all about describing a force [...] – Pedro G. Ferreira. [10].

¹²Indeed, even the “one-metric” theory of massive gravity requires two metrics, but one of which serves as a fixed—nonzero—background for the dynamical metric.

¹³In GR, torsion vanishes identically by construction. The role of torsion in other theories of gravity is theory-dependent: in some theories such as the Einstein–Cartan theory [34–36], torsion couples to spin; however, it is also possible to use only torsion (without curvature), to construct a theory that is, surprisingly, *equivalent* to GR, which is often called “TEGR (teleparallel equivalent of general relativity)” [37]. TEGR reminds us that theories should not be confused with reality—the latter involves observed phenomena, e.g., a falling apple, while the former are attempts to understand said observations.

entirely wrong, can be misleading. In fact, from the point of view of well-posedness of the evolution equations, it is far simpler to work with torsion as what it truly is, namely, a geometric quantity [38].

2.4 Equivalence Principle, Einstein’s elevator, and All that

The Principle of Equivalence performed the essential office of midwife at the birth of general relativity, but [...] I suggest that the midwife be now buried with appropriate honors.

—John L. Synge

Note that we have not mentioned the “equivalence principle” at all in this chapter—general relativity as a matured mathematical theory, based on Lorentzian geometry equipped with the Einstein field equations, *requires no mention of the equivalence principle*.¹⁴ Many physicists like to treat the equivalence principle as an important principle that somehow defines general relativity. This is, at best, misleading.

The equivalence principle says that the gravitational “force” as experienced locally, while standing on a massive body is the same as the “pseudo-force” experienced by an observer in a non-inertial (accelerating) frame of reference. The usual depiction of this statement is using the “Einstein’s elevator.” Consider someone inside an elevator with its cable cut, and thus is in free fall under the gravitational field. The gravitational acceleration on Earth is about 9.8 ms^{-2} . Now, consider another elevator in space, fitted to a rocket engine at the bottom, such that it accelerates at the same rate. An observer inside such an elevator in space—so it is said—would not be able to tell whether he is accelerating in space, or whether he is in free fall on Earth. This allegedly demonstrates that acceleration is somehow equivalent to gravity.

Mathematically, the equivalence principle is merely the statement that one can find a coordinate in the neighborhood of any point p on the spacetime manifold such that the Christoffel symbols vanish at that point. These are called “Riemann normal coordinates centered at p .” The metric, evaluated at p , takes the canonical Minkowski form and furthermore, the first derivative of the metric vanishes. That is to say, higher order derivatives are, in general, *non-vanishing*. Thus, if one has a sufficiently sensitive instrument, one can always detect curvature—curvature cannot be “transformed away” simply because you change a coordinate system! This is also saying that, strictly speaking, there is no such thing as an “inertial frame” in GR. In practice, however, we often talk about “negligible curvature” so that we can talk about an “inertial frame,” but one has to remember that this is an approximate statement.

In terms of Einstein’s elevator, this fact can be seen by noting that, if the elevator is indeed accelerating in space, two dropped balls would fall downward to the floor

(Footnote 13 continued)

It is possible that there can be more than one theory, which are different in terms of their mathematical structures, yet provide equivalent physical predictions.

¹⁴Or for that matter, the “principle of general covariance”—(almost?) *any* theory can be made general covariant. See the debate about this issue in [39].

of the elevator in parallel. However, if the elevator is free falling in the Earth's gravitational field, each of the balls will fall toward the center of the Earth, and therefore, they *cannot* fall down parallel to each other. Of course the effect is very small, but this does not change the fact that *it is there*. The equivalence principle, therefore, is only strictly true for an infinitesimally small elevator—a point—and therefore not an elevator.

While the above technicality about the equivalence principle is nitpicking, there is another folklore which is a serious misunderstanding of the physics, namely, that the equivalence principle implies light bending. This again follows from taking Einstein's elevator too seriously: one imagines that a beam of light is hitting the elevator that is accelerating in space, entering via a small hole on the side. By the time the light ray hits the wall on the opposite wall of the elevator, the elevator would already had accelerated upward a little, and so it would seem that the light beam does not travel in a straight line, but would rather hit the wall at a position somewhat lower than its entry point. By the equivalence principle, it is then claimed that, for an elevator that is freely falling in a gravitational field, light rays would be bent by gravity. This gravitational lensing is evident, for example, by observing the star lights around the Sun during a solar eclipse (one of the classic tests of GR).

Although the conclusion that light rays can be bent by gravity is correct, the reasoning via the equivalence principle above is *entirely mistaken*. The fact is that, since the equivalence principle only holds at a point, it is nothing but a *local* statement. This cannot possibly imply the bending of light rays in gravitational field, which is clearly an effect over a finite region of spacetime. In fact, the local light bending implied by equivalence principle is purely kinematical and does not depend on the field equations. This is evident since we have not mentioned anything about how the theory of gravity should be in this thought experiment. One could in fact construct a theory that satisfies the equivalence principle but does *not* predict light bending. For example, Nordström's theory of gravitation¹⁵ [41–43], which is a generalization of the Newtonian Poisson equation of a scalar field ϕ from

$$\nabla^2 \phi = 4\pi G\rho, \quad (2.45)$$

to

$$\phi \square \phi = -4\pi GT, \quad (2.46)$$

where $\square := \partial^a \partial_a$ is the D'Alembert operator and T is the trace of the energy-momentum tensor. This fact is best appreciated by examining the geometrized version of Nordström's theory, also known as the Einstein–Fokker theory [44], in which the scalar curvature of spacetime is related to the trace of the energy–momentum tensor: $R = 24\pi GT/c^4$. Since the energy–momentum tensor for the electromagnetic field is trace-free, it cannot give rise to a curvature effect, and thus no light bending by the gravitational field. The fact that it satisfies the equivalence principle merely follows

¹⁵Readers interested in thought experiments would also enjoy [40].

from the purely geometric statement that *any* (semi)-Riemannian manifold admits a normal coordinate system; see also [45]. For more discussion on the issue of “local versus global” light bending, see [46].

It is not that the equivalence principle is wrong, but if one is not careful, it might lead to the wrong results. It might be best not to bother with the principle at all, and just focus on the mathematics. Of course, different people think in different ways, and not everyone prefers advanced mathematics over simple rods and clocks in relativity, so the philosophy to do away with the equivalence principle in this section is purely a personal, biased, preference.

2.5 Causal Structure and Penrose Diagrams

To see a world in a grain of sand,
And heaven in a wild flower,
Hold infinity in the palms of your hand,
And eternity in an hour.

—William Blake

The existence of a temporal dimension in Lorentzian geometry means that there is a concept of causality. This is what allows one to define a black hole. Mathematically, a black hole is a spacetime region such that whatever is inside the black hole, even light, cannot escape. Let us make this notion more precise.

Given a point $p \in M$, we can define the *causal past* of p , denoted by $J^-(p)$, as the set

$$J^-(p) := \{q \in M \mid \exists \text{ a past-directed causal curve from } p \text{ to } q.\} \quad (2.47)$$

The causal past can be defined for a set S , simply as

$$J^-(S) = \bigcup_{p \in S} J^-(p). \quad (2.48)$$

The causal future of a point and a set can be defined similarly.¹⁶

Light rays travel on null geodesics all the way to “future null infinity,” which is denoted by \mathcal{I}^+ . A *black hole*, BH, is the region from which light cannot escape, so

$$\text{BH} := M \setminus J^-(\mathcal{I}^+). \quad (2.49)$$

The *event horizon*, EH, is its boundary

$$\text{EH} := \partial(M \setminus J^-(\mathcal{I}^+)). \quad (2.50)$$

¹⁶The study of causal structure is an important aspect of Lorentzian geometry, and we refer the readers to [8, 47] for more details.

Note that in this definition, the event horizon is a three-dimensional entity, and it is the “world-tube” of what we usually think of as the event horizon—the two-dimensional surface of a black hole. To obtain the two-dimensional event horizon, one simply take a cross section between EH and a spacelike hypersurface. It turns out that since EH is a null hypersurface, the area of the cross section is independent of the choice of spacelike slices.

In addition to the future null infinity, there is also the “past null infinity,” denoted by \mathcal{I}^- . Massive particles of course do not travel on null geodesics, and their trajectories are always timelike. To this, there corresponds the notion of future and past timelike infinity, denoted by i^+ and i^- , respectively. Spacelike infinity is denoted by i^0 . Since spacetimes are often, though not always, infinite in both temporal and spatial directions, it is difficult to grasp the entire spacetime. The Penrose diagram offers such a mean, by representing the causal structure of the entire spacetime on a finite diagram. Preserving the causal structure means that light cones are still represented by 45° lines.

To be more explicit, suppose that (M, g) is the spacetime of interest. If Ω^2 is a smooth, strictly positive function, then the metric $\tilde{g} = \Omega^2 g$ is said to arise from g due to a *conformal transformation*. The angles on a manifold equipped with a metric are measured using the generalized cosine law: If X and Y are two vector fields, then the angle θ between the vectors at point $p \in M$ is given by

$$\cos \theta = \frac{g(X, Y)}{\sqrt{g(X, X)g(Y, Y)}} \Big|_p. \quad (2.51)$$

Clearly, the angles between two vectors are the same regardless of whether it is measured with respect to the original metric g or the conformally related metric \tilde{g} , since the conformal factor Ω^2 cancels out. In addition, the ratio of the length of any two vectors measured by the two metrics remains unchanged for the same reason. Also, null curves with respect to one metric are also null with respect to the other one. The trick then is to find a suitable Ω such that we can “pull” infinities to some finite ranges. One possible function for such a re-scaling purpose is the arctan function, since it is bounded between $-\pi/2$ and $\pi/2$. However, in practice, usually a few coordinate transformations are required to successfully construct a Penrose diagram.

For an explicit example, we will work out the Penrose diagram for a Schwarzschild black hole. Let us start with the usual metric of the form

$$g[\text{Sch}] = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2). \quad (2.52)$$

This can be transformed into the Kruskal–Szekeres coordinates (U, V, θ, ϕ) via, if $r > 2M$,

$$U = \left(\frac{r}{2M} - 1\right)^{1/2} e^{\frac{r}{4M}} \cosh\left(\frac{t}{4M}\right), \quad V = \left(\frac{r}{2M} - 1\right)^{1/2} e^{\frac{r}{4M}} \sinh\left(\frac{t}{4M}\right). \quad (2.53)$$

and for $r < 2M$,

$$U = \left(1 - \frac{r}{2M}\right)^{1/2} e^{\frac{r}{4M}} \cosh\left(\frac{t}{4M}\right), \quad V = \left(1 - \frac{r}{2M}\right)^{1/2} e^{\frac{r}{4M}} \sinh\left(\frac{t}{4M}\right). \quad (2.54)$$

The metric then takes the form

$$g = \frac{32M^3}{r} e^{-\frac{r}{2M}} (-dV^2 + dU^2) + r^2(d\theta^2 + \sin^2\theta d\phi^2). \quad (2.55)$$

In order to construct the Penrose diagram, we introduce new coordinates u and v such that¹⁷

$$U = \frac{1}{2}(v - u); \quad V = \frac{1}{2}(v + u). \quad (2.56)$$

In order to bring infinities into finite range, we use the aforementioned arctan function, by introducing yet more coordinates (u', v') and (U', V') , as follows:

$$u' := \arctan(u) := V' - U', \quad (2.57)$$

$$v' := \arctan(v) := V' + U'. \quad (2.58)$$

It turns out that light rays move on curves of constant u' and v' , i.e., the 45° lines in the $U'V'$ plane. The ranges for u' and v' are $(-\pi/2, \pi/2)$.

Also, note that, the Kruskal–Szekeres coordinates satisfy

$$\left(\frac{r}{2M} - 1\right) e^{\frac{r}{2M}} = U^2 - V^2 = (U + V)(U - V). \quad (2.59)$$

On the event horizon $r = 2M$, it follows that $U = \pm V$. This corresponds to the “future event horizon” ($U = V$) and “past event horizon” ($U = -V$), respectively. Since $u = V - U$ and $v = V + U$, we can obtain $u = 0 = v$ by making the appropriate substitution. Then $u' = \arctan u = 0$ and similarly $v' = 0$, that is, $V' - U' = 0$ and $V' + U' = 0$. Thus the horizon is represented by the lines $V' = \pm U'$.

At the singularity $r = 0$, and $V > 0$, we see by the defining relations that $-1 = U^2 - V^2 = uv$. Consider the equation $u' + v' = 2V' = \arctan(u) + \arctan(v)$. Then, we have

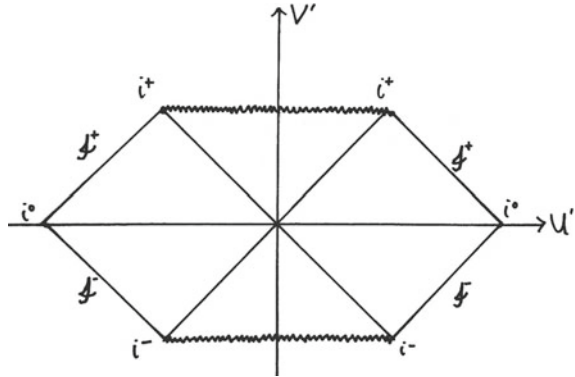
$$\tan(u' + v') = \tan[\arctan(u) + \arctan(v)] = \frac{u + v}{1 - uv} = \frac{1}{2}(u + v) = V. \quad (2.60)$$

Therefore,

$$-\frac{\pi}{2} < u' + v' = \arctan(V) < \frac{\pi}{2}, \quad (2.61)$$

¹⁷Note that both U , V and u , v are dimensionless.

Fig. 2.2 The Penrose diagram of a maximally extended (asymptotically flat) Schwarzschild black hole. Singularities are represented by wavy lines. The 45° lines that cross at the origin are the event (past and future) horizons



and at the boundary of spacetime $V \rightarrow \infty$, this corresponds to $\arctan V \rightarrow \pi/2$, and

$$V' = \frac{1}{2}(u' + v') = \frac{1}{2} \arctan V \rightarrow \frac{\pi}{4}. \quad (2.62)$$

Similarly, for $r = 0$ and $V < 0$, the singularity maps into the line $V' = -\pi/4$. On the other hand,

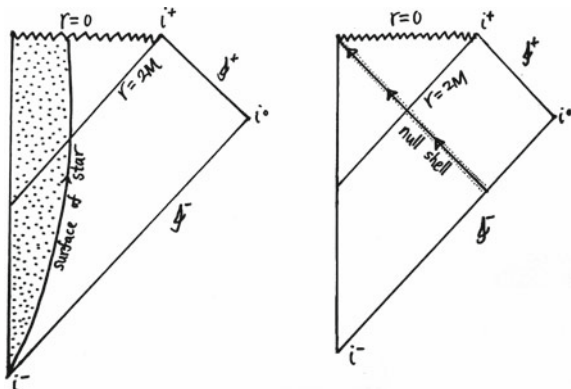
$$\tan(v' - u') = \tan(2U) \implies -\frac{\pi}{2} < v' - u' = 2U' < \frac{\pi}{2}, \quad (2.63)$$

that is, U' is bounded between $-\pi/4$ and $\pi/4$.

We may now proceed to draw the Penrose diagram for the Schwarzschild black hole (Fig. 2.2). Note that a generic point on the diagram is a 2-sphere, that is, angular dimensions have been suppressed. It is now clear from the Penrose diagram that the Schwarzschild singularities are horizontal lines (and so are orthogonal to timelike curves), and are therefore spacelike. It is also clear that any timelike or null curves inside the horizon will hit the singularity. Note also that i^+ and i^- are distinct from $r = 0$ since there are timelike curves (outside the black hole horizon) that never hit the singularity. In fact, there are four regions of spacetime: Region I corresponds to our assumed asymptotically flat universe, region II is a black hole, region III is *another* asymptotically flat universe, connected to Region I via an Einstein–Rosen bridge (the coordinate $(U', V') = (0, 0)$), and Region IV is a *white hole* region. This is the *maximally extended Schwarzschild spacetime*. Note that the Einstein–Rosen bridge is a *non-traversable wormhole*, since only spacelike curves pass through the throat.

One important feature of this Penrose diagram is that it is *time-reflective symmetric*, i.e., invariant under $t \mapsto -t$. Such a black hole is not a very realistic one. In the astrophysical context, black holes are the end stages of gravitational collapse of massive stars. The interior of the star is of course not a vacuum, and hence not described by a Schwarzschild metric. Therefore, there is no reason to expect the

Fig. 2.3 The Penrose diagram of an asymptotically flat Schwarzschild *black hole* formed from stellar collapse—dotted part is the stellar interior (left), and from a collapsing null shell (right). Note that in the case of a collapsing null shell, an observer can already be inside the horizon well before the null shell arrives at his or her location



presence of a wormhole connecting to another universe, not even a non-traversable one. The past of this collapsing star spacetime is also not the same as that of the maximally extended Schwarzschild metric—notably, there was no white hole. The Penrose diagram is therefore quite different.

Note that we can also collapse a “null shell,” an incoming spherically symmetric shell of radiation, into a black hole, provided we have enough energy in the shell. The Penrose diagram in Fig. 2.3 makes it obvious that the event horizon forms before the null shell has even arrived. This means that it is possible that an event horizon is now forming right where you are, without you realizing. This gives another indication that the event horizon is really not a special entity at all.

For completeness, let us also mention the concept of a Cauchy hypersurface. A set is *achronal* if no two points on S can be joined by timelike curves. If S is a set that is achronal, and in addition, every causal curve in M crosses it precisely once, then it is a *Cauchy surface*. A spacetime (M, g) that admits a spacelike hypersurface Σ , which is Cauchy, is said to be *globally hyperbolic*. In general relativity, global hyperbolicity means that one can set up a well-posed Cauchy problem, i.e., given initial conditions on Σ , one can evolve the system forward (or backward) in time to study its evolution. In a sense, this is what “doing physics” means.

If one examines the Penrose diagram of an asymptotically flat Reissner–Nordström black hole carefully (Fig. 2.4), one would notice that the physics in the region behind the inner horizon of the black hole cannot be determined from the initial data outside of said horizon alone, but must be fixed by boundary conditions on the (timelike) singularity. Therefore, this spacetime is not globally hyperbolic, and the inner horizon is called a *Cauchy horizon*. The asymptotically flat Kerr black hole has a similar Cauchy horizon. For more discussion, see [48].

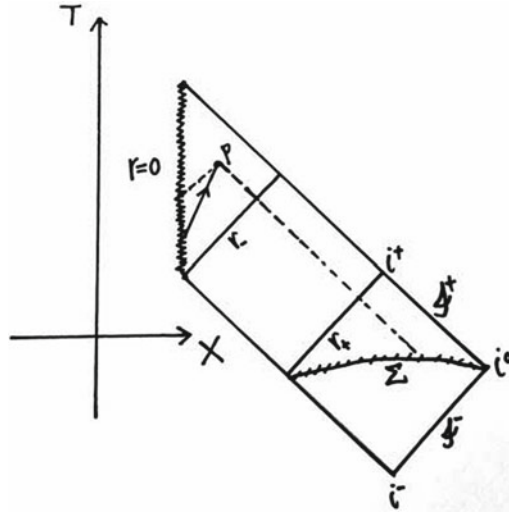


Fig. 2.4 The Penrose diagram of an asymptotically flat Reissner–Nordström *black hole* (not maximally extended). T and X are just labels for some temporal and spatial direction, respectively. There are two horizons: r_+ and r_- . The latter is a Cauchy horizon. Given a Cauchy hypersurface Σ in the exterior spacetime, the initial data cannot determine the physics at any point p behind r_- , since the timelike singularity is in the causal past of p (the boundary of the causal past is denoted by *dashed lines*), and can therefore affect p (here, an *arrow line* emanating from the singularity demonstrates this)

2.6 Anti-de Sitter Spacetime and Holography

There is geometry in the humming of the strings.

–Pythagoras

In this section, we start by reviewing the technique of stereographic projection on a sphere, and then apply the same method to hyperbolic space. We explain how anti-de Sitter spacetime is related to hyperbolic space. Several coordinate systems that are commonly used in the literature are introduced, and the causal structure of AdS spacetime is discussed. After that we introduce the idea of holography, which says that physics with (quantum) gravity—in fact, string theory—in anti-de Sitter spacetime is in some precise sense equivalent to physics of supersymmetric field theory without gravity on the conformal boundary of the same spacetime. This so-called AdS/CFT correspondence will play a central role in this thesis, with a black hole placed in the AdS bulk being dual to some field theory with finite temperature on the boundary that behaves a lot like quantum chromodynamics (the study of quarks and gluons).

2.6.1 Stereographic Projection and Hyperbolic Geometry

You must not attempt this approach to parallels. I know this way to its very end. I have traversed this bottomless night, which extinguished all light and joy of my life. [...] For God's sake, I beseech you, give it up. Fear it no less than sensual passions because it too may take all your time and deprive you of your health, peace of mind and happiness in life.

—Farkas Bolyai, to his son János Bolyai, one of the founders of non-Euclidean geometry.

We first review the method of stereographic projection on a sphere, usually taught in the first course of differential geometry. We will do it on a 3-sphere S^3 , and then generalize the method to hyperbolic 3-space \mathbb{H}^3 .

Let $\{y^a\} = \{y^i, y^4\}$ denote the coordinates in \mathbb{R}^4 . First recall that S^3 is defined by the equation (setting the radius to unity)

$$\sum_{i=1}^3 (y^i)^2 + (y^4)^2 = 1 \quad (2.64)$$

in \mathbb{R}^4 . Our convention for stereographic projection is to project from the north pole to a plane that the south pole rests on (See Fig. 2.5). Let $\{x^i\}$ denote the coordinates on the projection plane. The origin of the sphere is at $(0, 0, 0, 0)$, and the plane has $y^4 = -1$. Elementary geometry shows that

$$x^i = \frac{2y^i}{1 - \sqrt{1 - \sum_{i=1}^3 (y^i)^2}} = \frac{2y^i}{1 - y^4}. \quad (2.65)$$

That is to say, the metric tensor on the ambient space

$$\delta[\mathbb{R}^4] = (dy^1)^2 + (dy^2)^2 + (dy^3)^2 + (dy^4)^2 \quad (2.66)$$

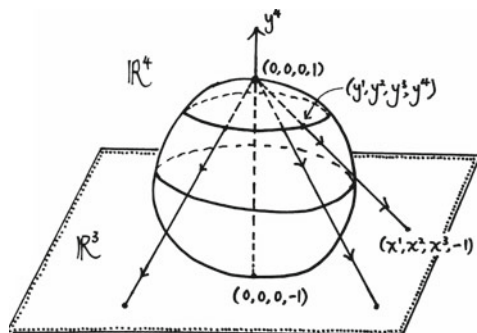


Fig. 2.5 Stereographic projection of a 3-sphere, with one dimension suppressed, from the north pole to a 3-plane that the sphere sits on

restricted to S^3 gives the equation

$$\frac{4}{(1-y^4)^2} \sum_{i=1}^3 (dy^i)^2 = \sum_{i=1}^3 (dx^i)^2. \quad (2.67)$$

Since

$$\frac{4}{(1-y^4)^2} = \left[\frac{2(1-y^4)}{(1-y^4)^2} \right]^2 = \left[\frac{1 + (y^4)^2 - 2y^4 + 1 - (y^4)^2}{(1-y^4)^2} \right]^2 \quad (2.68)$$

$$= \left[\frac{(1-y^4)^2 + \sum_{i=1}^3 (y^i)^2}{(1-y^4)^2} \right]^2, \quad (2.69)$$

we have

$$\left[1 + \sum_{i=1}^3 \frac{(y^i)^2}{(1-y^4)^2} \right]^2 \sum_{i=1}^3 (dy^i)^2 = \sum_{i=1}^3 (dx^i)^2, \quad (2.70)$$

that is, S^3 can be described by a metric tensor of the form

$$g[S^3] = \frac{1}{\left[1 + \frac{1}{4} \sum_{i=1}^3 (x^i)^2 \right]^2} \sum_{i=1}^3 (dx^i)^2. \quad (2.71)$$

Note that this metric is manifestly homogeneous and isotropic.

Let $y^4 = \cos \rho$, $y^i = \sin \rho \tilde{x}^i$, where

$$\begin{cases} \tilde{x}^1 = \sin \theta \cos \phi, \\ \tilde{x}^2 = \sin \theta \sin \phi, \\ \tilde{x}^3 = \cos \theta. \end{cases} \quad (2.72)$$

We can then write the metric in the form

$$ds^2 = d\rho^2 + \sin^2 \rho (d\theta^2 + \sin^2 \theta d\phi^2). \quad (2.73)$$

Let $r = \sin \rho$, we obtain another form of the metric

$$ds^2 = \frac{1}{1-r^2} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2). \quad (2.74)$$

Readers familiar with cosmology will recognize that this is the form that goes into the spatial part of the closed Friedmann–Lemaître–Robertson–Walker (FLRW) metric.

The stereographic projection for hyperbolic 3-space can be carried out in the exactly same manner, using the defining equation:

$$\sum_{i=1}^3 (y^i)^2 - (y^4)^2 = -1, \quad y^4 > 1, \quad (2.75)$$

as a hypersurface in Minkowski spacetime $\mathbb{R}^{3,1}$. This is the so-called “hyperboloid model” of \mathbb{H}^3 . We may now project the points on the hyperboloid onto the plane $y^4 = -1$, via the origin as the projection point (Fig. 2.6).

Then we would obtain, similarly, a metric for \mathbb{H}^3 :

$$g[\mathbb{H}^3] = \frac{1}{\left[1 - \frac{1}{4} \left(\sum_{i=1}^3 (x^i)^2\right)\right]^2} \sum_{i=1}^3 (dx^i)^2. \quad (2.76)$$

This is the Poincaré ball.

The Poincaré disk \mathbb{H}^2 is easier to imagine, and it is the disk of radius 1 equipped with the metric

$$ds^2 = \frac{4(dx^2 + dy^2)}{(1 - x^2 - y^2)^2}. \quad (2.77)$$

Readers familiar with complex analysis would appreciate this disk in complex coordinates¹⁸

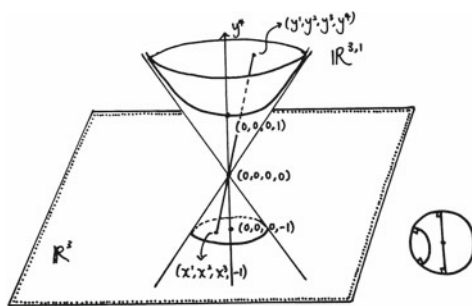


Fig. 2.6 Stereographic projection of a hyperbolic 3-space, with one dimension suppressed, from the hyperboloid model to a Poincaré sphere model (shown as a disk here). The disk at the side shows some of the geodesics on this space—they are *straight lines* that pass through the origin, or arcs that are perpendicular to the boundary at the points they meet the boundary

¹⁸Readers with a complex analysis background are encouraged to read [49].

$$ds^2 = \frac{4 \, dz d\bar{z}}{(1 - |z|^2)^2}. \quad (2.78)$$

The hyperbolic distance from the origin to any point z in the disk is given by

$$d(0, z) = \int \frac{2 \, d|z|}{(1 - |z|^2)} = \ln \left(\frac{1 + |z|}{1 - |z|} \right). \quad (2.79)$$

Note that this expression tends to infinity in the limit $|z| \rightarrow 1$; the boundary of the disk is infinitely far away.

The unit disk in the complex plane can be mapped into the upper-half plane via the *inverse Cayley transform*, which is a Möbius transformation given by

$$f(z) := i \frac{1 + z}{1 - z}. \quad (2.80)$$

Note that the boundary of the disk is mapped into the real line. The (real) metric on the upper-half plane takes the form:

$$ds^2 = \frac{dx^2 + dy^2}{y^2}. \quad (2.81)$$

Note again that distance becomes unbounded as $y \rightarrow 0$, the boundary of the upper-half plane. For \mathbb{H}^3 , we would have an upper-half space model, with metric

$$ds^2 = \frac{dx^2 + dy^2 + dz^2}{z^2}. \quad (2.82)$$

From the Poincaré ball, we can also define $y^4 = \cosh \rho$ and $y^i = \sinh \rho \tilde{x}^i$ analogously, and furthermore $r = \sinh \rho$, to put the metric into the “spatial FLRW form”:

$$ds^2 = \frac{1}{1 + r^2} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2). \quad (2.83)$$

2.6.2 The Geometry of Anti-de Sitter Spacetime

I regret that it has been necessary for me in this lecture to administer such a large dose of four-dimensional geometry. I do not apologize, because I am really not responsible for the fact that nature in its most fundamental aspect is four-dimensional. Things are what they are; and it is useless to disguise the fact that “what things are” is often very difficult for our intellects to follow.

—Alfred N. Whitehead

Anti-de Sitter (AdS) spacetime will play a central role in this thesis, as eventually we would like to study black hole solutions that are asymptotically locally AdS. It is therefore prudent to properly introduce AdS spacetime in some details.¹⁹ For simplicity, we will restrict our discussion to four-dimensional spacetime, denoted by AdS_4 . An AdS spacetime is a solution to the Einstein Field Equations with a *negative* cosmological constant.

The best way to visualize this geometry is to consider it as a hypersurface in a higher dimensional spacetime. In particular, consider $\mathbb{R}^{3,2}$ (note that there are *two* time dimensions!) with canonical metric

$$\delta[\mathbb{R}^{3,2}] = -(dy^4)^2 + \sum_{i=1}^3 (dy^i)^2 - (dy^0)^2. \quad (2.84)$$

AdS spacetime is then the hypersurface defined by the equation

$$-(y^4)^2 + \sum_{i=1}^3 (y^i)^2 - (y^0)^2 = -1. \quad (2.85)$$

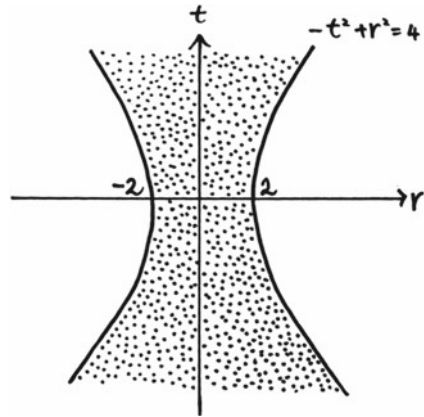
Note that for each constant y^i slice, we have a circle defined by $(y^4)^2 + (y^0)^2 = \text{const.}$, and thus the topology is $S^1 \times \mathbb{R}^3$. We can again perform stereographic projection, but we will not carry this out explicitly. From the previous discussion, we know that metrically \mathbb{R}^3 here is really diffeomorphic to hyperbolic space \mathbb{H}^3 . Due to the temporal *plane* spanned by the two temporal dimensions y^0 and y^4 , there is a *closed timelike curve* (CTC). Physicists are usually very uncomfortable with CTCs, and speak of “passing to the universal covering spacetime $\widetilde{\text{AdS}}$ ” instead, that is to say, one “unwraps” the circle S^1 representing time coordinate into its covering space \mathbb{R} , and by “AdS” one actually secretly means $\widetilde{\text{AdS}}$.

As we will soon see, AdS spacetime is really a peculiar one: in addition to CTC, it is not globally hyperbolic. The fact that it attracted so much attention despite these otherwise undesired features (from the point of view of classical general relativity) is due to its importance in supergravity and string theory. See a review by Gibbons [51] for some applications of AdS spacetime. His comment on CTCs in AdS is especially noteworthy:

Many physicists are unhappy with the CTC’s in AdS_{p+2} and seek to assuage their feelings of guilt by claiming to pass to the universal covering spacetime $\widetilde{\text{AdS}}_{p+2}$. In this way they feel that they have exorcised the demon of “acausality”. However therapeutic uttering these words may be, nothing is actually gained in this way. Consider for example the behavior of test particles. Every timelike geodesic on AdS_{p+2} is a closed curve of the same durations equal to $2\pi R$, which Heraclitus would have called the ‘Great Year’.

¹⁹For even more details, see [50].

Fig. 2.7 AdS spacetime (dotted region) as a portion of Minkowski spacetime given by the metric $g = [1 - \frac{1}{4}(-t^2 + r^2)]^{-2}\eta$



AdS spacetime, like de Sitter spacetime and Minkowski spacetime, is maximally symmetric, in the sense that there are—in four dimensions—10 Killing vectors²⁰

$$y_A \frac{\partial}{\partial y^B} - y_B \frac{\partial}{\partial y^A}; \quad A \neq B, \quad A = 0, 1, 2, 3, 4, \quad (2.86)$$

where $y_A = \delta_{AB} y^B$, with δ being the metric 2.84.

An alternative way to represent AdS spacetime is as a portion of Minkowski space. This is shown in Fig. 2.7. In fact, AdS spacetime admits metric of the form

$$g[\text{AdS}] = \frac{1}{[1 - \frac{1}{4}(-t^2 + x^2 + y^2 + z^2)]^2} \eta, \quad (2.87)$$

where η is the standard metric on $\mathbb{R}^{3,1}$. This form of the metric makes it clear that AdS spacetime is conformally related to Minkowski spacetime. In fact, for $t = 0$, one immediately sees that the metric is just that of a hyperbolic space given by the metric 2.76. This is also true for different values of t —geometrically, they are just hyperbolic balls with different radii.

For the discussions involving black holes, it is best to use the static coordinates (which makes comparison to the usual Schwarzschild metric apparent):

$$g[\text{AdS}] = - \left(\frac{r^2}{L^2} + 1 \right) dt^2 + \left(\frac{r^2}{L^2} + 1 \right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2), \quad (2.88)$$

²⁰A maximally symmetric spacetime has, in d -dimensions, a total of $d(d+1)/2$ Killing vectors. See Lemma (9.28) of [8].

where L is called the curvature length scale of AdS spacetime. It is related to the negative cosmological constant²¹ Λ by $\Lambda = -3/L^2$. AdS spacetime has constant scalar curvature²² $R = -12/L^2$. Again, every constant time slice is just a hyperbolic space given by the metric 2.83 (where $L = 1$).

The static coordinates relate to the embedding coordinates (y^a) by

$$\begin{cases} \frac{y^0}{L} = \sqrt{1 + \left(\frac{r}{L}\right)^2} \cos\left(\frac{t}{L}\right), \\ \frac{y^i}{L} = \frac{r}{L} \omega^i, \\ \frac{y^4}{L} = \sqrt{1 + \left(\frac{r}{L}\right)^2} \sin\left(\frac{t}{L}\right), \end{cases} \quad (2.89)$$

where $\omega^1 = \cos \theta$, $\omega^2 = \sin \theta \cos \phi$, and $\omega^3 = \sin \theta \sin \phi$. The static coordinates cover the entire spacetime (except for trivial coordinate singularities), so that it is evident that AdS_4 is globally static. By construction, it has symmetry group $\text{SO}(3, 2)$. This should be contrasted to the static coordinates for *de Sitter* spacetime, which unlike AdS case has positive cosmological constant. In this case, the static coordinates do not cover the entire spacetime due to the presence of the cosmological horizon.

AdS spacetime also admits a coordinate system such that the spatial sections are the upper-half space model of \mathbb{H}^3 , given by the metric (2.82). They are called the *Poincaré* coordinates:

$$ds^2 = \frac{L^2}{z^2} (-dt^2 + dx^2 + dy^2 + dz^2). \quad (2.90)$$

If one defines $r = L^2/z$, then the upper-half space model of \mathbb{H}^3 , upon restoring L , yields

$$ds^2 = \frac{L^2}{r^2} dr^2 + \frac{r^2}{L^2} (dx^2 + dy^2), \quad (2.91)$$

which is just the spatial part of the flat slice parametrization of AdS spacetime:

$$ds^2 = -\frac{r^2}{L^2} dt^2 + \frac{L^2}{r^2} dr^2 + \frac{r^2}{L^2} (dx^2 + dy^2), \quad (2.92)$$

which, in turn, is of course equivalent to the Poincaré patch given by the metric 2.90. In fact, not only are that the spatial parts of AdS_d simply \mathbb{H}^{d-1} , but also the whole

²¹In d -dimensions, $\Lambda = -\frac{(d-1)(d-2)}{2L^2}$.

²²In d -dimensions, the Ricci tensor satisfies $R_{ab} = \frac{2\Lambda}{d-2} g_{ab}$. Thus, the scalar curvature satisfies $R = -\frac{d(d-1)}{L^2}$.

AdS spacetime itself, under Wick rotation (i.e., the complexification $t \rightarrow it$) to “Euclidean-AdS,” becomes \mathbb{H}^d . This provides a means to study physics via powerful tools of complex analysis. As an example, in AdS_3 , one can topologically identify points to construct a black hole solution, known as a BTZ black hole [52]. The Euclidean version of BTZ spacetime turns out to be \mathbb{H}^3/Γ , where $\Gamma \subset \text{PSL}(2, \mathbb{C})$ is a *Schottky group* [53]. The discovery of the BTZ solution was itself a surprise since without a negative cosmological constant, and Einstein gravity is trivial in three dimensions.

2.6.3 Holography: The AdS/CFT Correspondence

String theory at its finest is, or should be, a new branch of geometry. ...I, myself, believe rather strongly that the proper setting for string theory will prove to be a suitable elaboration of the geometrical ideas upon which Einstein based general relativity.

—Edward Witten

Consider first the Minkowski spacetime with canonical metric

$$\eta = -dt^2 + dx^2 + dy^2 + dz^2. \quad (2.93)$$

In terms of null coordinates $u := t - r$, $v := t + r$, we have

$$ds^2 = -du dv + \frac{1}{4}(u - v)^2(d\theta^2 + \sin^2 \theta d\phi^2). \quad (2.94)$$

If we introduce new coordinates (U, V) and (T, R) via²³

$$U = \arctan(u) = \frac{1}{2}(T - R); \quad V = \arctan(v) = \frac{1}{2}(T + R), \quad (2.95)$$

with $U, V \in (-\pi/2, \pi/2)$, then we obtain metric of the form

$$ds^2 = \frac{1}{4 \cos^2 U \cos^2 V} [-4dU dV + \sin^2(V - U)(d\theta^2 + \sin^2 \theta d\phi^2)], \quad (2.96)$$

or equivalently,

$$ds^2 = \Omega^{-2}(T, R) [-dT^2 + dR^2 + \sin^2 R(d\theta^2 + \sin^2 \theta d\phi^2)], \quad (2.97)$$

where $\Omega = 2 \cos U \cos V = \cos T + \cos R$. Therefore, the Minkowski metric is conformally related to the metric

$$ds^2 = -dT^2 + dR^2 + \sin^2 R(d\theta^2 + \sin^2 \theta d\phi^2), \quad (2.98)$$

²³Note that in writing $U = \arctan(u)$, it is really $\tan U = u/1$, where 1 is $u = 1$ in the corresponding unit of length.

where $0 \leq R < \pi$, and $-\pi < T < \pi$. The spatial part of this metric is a 3-sphere, so its topology is $\mathbb{R} \times S^3$. We refer to this as an *Einstein Static Universe*. Suppressing one spatial dimension, we can represent this spacetime as an infinitely long solid cylinder, and hence it is also called an *Einstein Cylinder*. Then, the above argument implies that we can conformally map Minkowski spacetime into a part of the Einstein cylinder (Fig. 2.8).

Since AdS spacetime is conformally related to Minkowski spacetime, it can also be conformally mapped into the Einstein cylinder. To see this explicitly, re-write the static metric (2.88) into the form

$$ds^2 = \left[1 + \left(\frac{r}{L} \right)^2 \right] \left[-dt^2 + \left(1 + \left(\frac{r}{L} \right)^2 \right)^{-2} dr^2 + r^2 \left(1 + \left(\frac{r}{L} \right)^2 \right)^{-1} (d\theta^2 + \sin^2 \theta d\phi^2) \right], \quad (2.99)$$

and then define

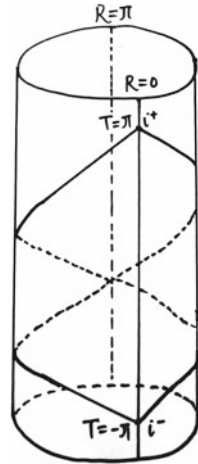
$$\omega := \int \frac{dr}{1 + \frac{r^2}{L^2}} = L \arctan \left(\frac{r}{L} \right). \quad (2.100)$$

This transforms the metric into the form

$$ds^2 = \sec^2 \left(\frac{\omega}{L} \right) \left[-dt^2 + d\omega^2 + L^2 \sin^2 \frac{\omega}{L} (d\theta^2 + \sin^2 \theta d\phi^2) \right]. \quad (2.101)$$

so the metric manifestly conformally maps into the Einstein cylinder. Note that the spatial infinity $r = \infty$ is mapped into $\omega = \pi L/2$. We can therefore visualize AdS spacetime with a cylinder, each fixed time slice of which gives a hyperbolic space.

Fig. 2.8 Minkowski spacetime conformally mapped into the Einstein Static Universe



To do this, it is convenient to further define dimensionless coordinates $\eta := t/L$ and $\chi := \omega/L$, so that the metric (2.101) becomes

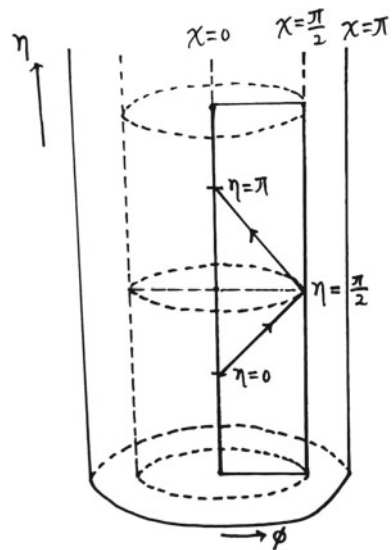
$$ds^2 = \frac{L^2}{\cos^2 \chi} [-d\eta^2 + d\chi^2 + \sin^2(\chi)(d\theta^2 + \sin^2 \theta d\phi^2)]. \quad (2.102)$$

Note that $0 \leq \chi < \pi/2$ (Fig. 2.9).

The Penrose diagram for AdS spacetime can be obtained from the Einstein cylinder by fixing $\phi = \text{const.}$ (Fig. 2.10). It has a peculiar feature that not all infinities can be given a finite range, so the diagram is not compact. Also, we note that in AdS spacetime, a light ray can travel from the (arbitrary) “center”²⁴ to null infinity, and be reflected back in a *finite* proper time of an observer sitting at the center (it takes an infinite affine time for the light ray). This implies that physics depends on the boundary conditions at null infinity—the spacetime is *not* globally hyperbolic. No timelike geodesic can reach the conformal boundary.

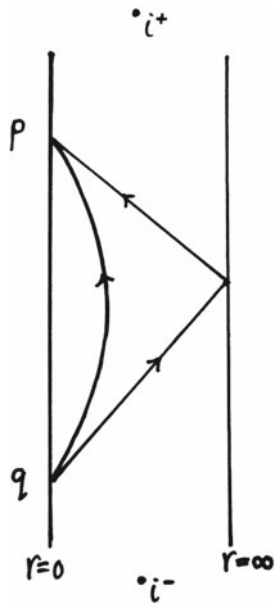
Recall that AdS spacetime has maximal symmetry. For AdS₅, the number of Killing vector fields is 15. In physics, this corresponds to 15 “symmetry transformations.” In Maldacena’s conjecture in string theory [55], the so-called Type IIB string theory in AdS₅ \times S⁵ is dual to a supersymmetric Yang–Mills (SYM) theory that lives on the boundary of AdS₅, which has 10 symmetries (six Lorentz transformations and four spacetime translations; these are all together called “Poincaré symmetry”). Supersymmetry is a symmetry that pairs integer-spin particles with half-integer-spin partners and vice versa (see Box 3.1 for more details). Being “dual” means that

Fig. 2.9 AdS spacetime conformally mapped into the Einstein static universe, shown here as a (solid) cylinder of infinite height. AdS maps into a smaller cylinder with radius $\chi = \pi/2$ instead of $\chi = \pi$. The Penrose diagram for AdS spacetime is obtained from the Einstein cylinder by fixing $\phi = \text{const.}$. For more discussion, see [54]



²⁴This “center” is arbitrary in the same sense that all points in *de Sitter* spacetime are a “center” from which everything else moves away from—there is no real center in such an expanding universe.

Fig. 2.10 The Penrose diagram of AdS spacetime. It is not globally hyperbolic since light rays can bounce off the conformal boundary (which is spatially infinitely far away) and be reflected off into the bulk in a finite proper time of an observer at the (arbitrary) “origin” $r = 0$. Physics at p is determined by both the initial conditions at q and the boundary conditions at $r = \infty$. No timelike geodesic from q can reach the boundary—the negative cosmological constant provides an attractive “force” that pulls massive particles back to the origin at p



there is an equivalence between them—a quantity computed on one side has a corresponding interpretation on the other. This is useful since difficult calculations on the boundary may become easier in the AdS bulk, and vice versa.

Due to the symmetry constraints (15 in the bulk, but only 10 on the boundary), not all field theories can be dual to AdS₅ (the S^5 is compactified and often not mentioned explicitly). The additional symmetries imposed on the field theory are *conformal symmetries*. These are symmetries under the conformal transformations of one *dilatation* and four coordinate *inversions*. Any quantum field theory that is invariant under the conformal transformations is called a conformal field theory (CFT). Furthermore, any theory that is invariant under dilatation is said to be *scale invariant*. Maldacena’s conjecture is then a correspondence between physics in AdS spacetime, with gravity, with physics on its boundary, and *without* gravity. This is known as the AdS/CFT correspondence.

SYM theory has a $SU(N)$ gauge symmetry, where N is the number of “colors” (like that of the quarks, see Box 5.1) in the theory. For the correspondence to be useful, we require that N must be sufficiently large. The reason for this is the ‘t Hooft coupling, $\lambda = g_{YM}^2 N$, where g_{YM} is the Yang–Mills coupling, which determines the interaction strength of the field theory. The local strength of gravity in the AdS bulk is determined by the curvature with corresponding length scale L ; the smaller value of L corresponds to the greater curvature. Note that Maldacena’s conjecture is in the context of string theory, so there are also strings living in the AdS bulk. Therefore, there is a length scale ℓ_s related to the string, which is inversely proportional to the string tension. The string coupling is $g_s \sim g_{YM}^2$. The ‘t Hooft coupling satisfies $\lambda \propto (L/\ell_s)^4$. Thus, if $\ell_s \ll L$ and N is sufficiently large, the strings are weakly

coupled in the bulk, but the field theory on the boundary is strongly coupled. Such a situation makes AdS/CFT correspondence useful—the field theory is too difficult for usual field theory techniques to compute, but the gravity in AdS is essentially classical—that is, we only need to know GR. The opposite regime in which the field theory is weakly coupled, however, means that one would require a full non-perturbative string theory calculation in the bulk. There are of course regimes in which calculations on both sides are difficult.

In the string theory picture, there are actually N D3-branes in the bulk. Their presence induces curvature in the geometry. Near the branes, the geometry takes the $\text{AdS}_5 \times S^5$ form in the low energy limit. We will not go into any more details in this work, and interested readers may refer to e.g. [51]. For some more details at the non-technical level, see [56, 57].

Of course, in physically interesting systems, for example, field theory with finite temperature, scale invariance is broken by the length scale set by the temperature. Thus, many so-called “AdS/CFT correspondence” applications are really neither (pure) AdS nor CFT. A better name would be gauge/gravity duality, or simply, holography.

References

1. Choquet-Bruhat, Y., DeWitt-Morette, C.: Analysis, manifolds and physics: part II, North Holland, 1st edn. (2000)
2. Pais, A.: Subtle is the Lord: the science and the life of Albert Einstein, p. 216. Oxford University Press, New York (1982)
3. Spivak, M.: A comprehensive introduction to differential geometry, vol.2, 3rd edn. Publish or Perish (1999)
4. Curiel, E.: A primer on energy conditions. [arXiv:1405.0403](https://arxiv.org/abs/1405.0403) [physics.hist-ph]
5. Lee, J.M.: Introduction to topological manifolds, graduate texts in mathematics, 2nd edn. Springer (2010)
6. Lee, J.M.: Introduction to smooth manifolds, graduate texts in mathematics, 2nd edn. Springer (2012)
7. Lee, J.M.: Riemannian manifolds: an introduction to curvature, graduate texts in mathematics, Springer (1997)
8. O’Neill, B.: Semi-Riemannian geometry with applications to relativity, pure and applied mathematics, 1st edn., Academic Press (1983)
9. Straumann, N.: General relativity (graduate texts in physics), 2nd ed., Springer (2013)
10. Ferreira, P.G.: The perfect theory: a century of geniuses and the battle over general relativity, Mariner Books (2014)
11. Noether, E.: Invariante variationsprobleme. In: Nachr. D. König. Gesellsch. D. Wiss. Zu Göttingen, Math-phys. Klasse, 235 (1918)
12. Curiel, E.: On tensorial concomitants and the non-existence of a gravitational stress-energy tensor. [arXiv:0908.3322](https://arxiv.org/abs/0908.3322) [gr-qc]
13. Chen, C.-M., Nester, J.M., Tung, R.-S.: Gravitational energy for GR and Poincaré gauge theories: a covariant Hamiltonian approach. Int. J. Mod. Phys. D **24**, 1530026 (2015). [arXiv:1507.07300](https://arxiv.org/abs/1507.07300) [gr-qc]
14. McInnes, B.: No forces, no sources: the ideology of general relativity, an entry for essays on gravitation, Gravity Research Foundation (1987). <http://www.gravityresearchfoundation.org/pdf/awarded/1987/mcinnnes.pdf>. Accessed 1 Jul 2015

15. Karlhede, A., Lindström, U., Amán, J.E.: A note on a local effect at the Schwarzschild sphere, *Gen. Rel. Grav.* **14**, 569 (1982)
16. Abdelqader, M., Lake, K.: Invariant characterization of the Kerr spacetime: locating the horizon and measuring the mass and spin of rotating black holes using curvature invariants. *Phys. Rev. D* **91**, 084017 (2015). [arXiv:1412.8757](#) [gr-qc]
17. Page, D.N., Shoom, A.A.: Local invariants vanishing on stationary horizons: a diagnostic for locating black holes. *Phys. Rev. Lett.* **114**, 141102 (2015). [arXiv:1501.03510](#) [gr-qc]
18. Moffat, J.W., Toth, V.T.: Karlhede's invariant and the black hole firewall proposal. [arXiv:1404.1845](#) [gr-qc]
19. Ortín, T.: *Gravity and strings*, cambridge University Press (2007)
20. Padmanabhan, T.: From gravitons to gravity: myths and reality. *Int. J. Mod. Phys. D* **17**, 367 (2008). [arXiv:gr-qc/0409089](#)
21. Deser, S.: Gravity from self-interaction redux. *Gen. Rel. Grav.* **42**, 641 (2010). [arXiv:0910.2975](#) [gr-qc]
22. Hinterbichler, Kurt: Theoretical Aspects of Massive Gravity. *Rev. Mod. Phys.* **84**, 671 (2012). [arXiv:1105.3735](#) [hep-th]
23. de Rham, C.: Massive gravity. *Living Rev. Relativity* **17**, 7 (2014). [arXiv:1401.4173](#) [hep-th]
24. Hassan, S.F., Rosen, R.A.: Bimetric gravity from ghost-free massive gravity, *JHEP* **02** 126 (2012). [arXiv:1109.3515](#)
25. Khosravi, N., Rahmanpour, N., Sepangi, H.R., Shahidi, S.: Multi-metric gravity via massive gravity. *Phys. Rev. D* **85**, 024049 (2012). [arXiv:1111.5346](#)
26. Capela, F., Tinyakov, P.G.: Black hole thermodynamics and massive gravity. *JHEP* **1104**, 042 (2011). [arXiv:1102.0479](#) [gr-qc]
27. Deser, S., Izumi, K., Ong, Y.C., Waldron, A.: Massive gravity acausality redux. *Phys. Lett. B* **726**, 544 (2013). [arXiv:1306.5457](#) [hep-th]
28. Deser, S., Izumi, K., Ong, Y.C., Waldron, A.: Superluminal propagation and acausality of nonlinear massive gravity. In: *Proceedings of the Conference in Honour of the 90th Birthday of Freeman Dyson*, World Scientific, 430-435 (2014). [arXiv:1312.1115](#) [hep-th]
29. Deser, S., Izumi, K., Ong, Y.C., Waldron, A.: Problems of massive gravities. *Mod. Phys. Lett. A* **30**, 1540006 (2015). [arXiv:1410.2289](#) [hep-th]
30. Motloch, P., Hu, W., Joyce, A., Motohashi, H.: Self-accelerating massive gravity: superluminality, Cauchy surfaces and strong coupling. *Phys. Rev. D* **92**, 044024 (2015). [arXiv:1505.03518](#) [hep-th]
31. Deser, S., Sandora, M., Waldron, A.: No consistent bimetric gravity?, *Phys. Rev. D* **88**, 081501(R) (2013). [arXiv:1306.0647](#)
32. Buchdahl, H.A.: *Seventeen simple lectures on general relativity theory*, Wiley (1981)
33. Hawking, S.W., Penrose, R.: *The nature of space and time*, Princeton University Press (1996)
34. Cartan, E.: Sur une généralisation de la notion de courbure de Riemann et les espaces à torsion', *C. R. Acad. Sci. (Paris)* **174**, 593 (1922)
35. Trautman, A.: Einstein-Cartan theory, *encyclopedia of mathematical physics*, edited by J.-P. Francoise, G.L. Naber, S.T. Tsou, Elsevier, **2**, 189–195 (2006). [arXiv:gr-qc/0606062](#)
36. Hehl, F.W., von der Heyde, P., Kerlick, G.D., Nester, J.M.: General relativity with spin and torsion: foundations and prospects. *Rev. Mod. Phys.* **48**, 393 (1976)
37. Einstein, A.: In *Sitzungsberichte der Preussischen Akademie der Wissenschaften. Berichte Preuss. Akad. Wiss. Berlin*, 217–224 (1928)
38. Nester, J.M., Wang, C.-H.: Can torsion be treated as just another tensor field?. *Int. J. Mod. Phys. Conf. Ser.* **07**, 158 (2012)
39. Norton, J.D.: General covariance and the foundations of general relativity: eight decades of dispute. *Rep. Prog. Phys.* **56**, 791 (1993)
40. Norton, J.: Einstein and Nordström: some lesser known thought experiments in gravitation. In: *The Attraction of Gravitation: New Studies in History of General Relativity*, Birkhäuser, Boston (1993)
41. Nordström, G.: Relativitätsprinzip und gravitation. *Phys. Zeitschr.* **13**, 1126 (1912)

42. Nordström, G.: Zur theorie der gravitation vom standpunkt des relativitätsprinzips, *Anna. Phys.* **42**, 533 (1913)
43. Wellner, M., Sandri, G., Scalar gravitation, *Amer. Jour. Phys.* **32**, 36 (1964), Erratum, *ibid* **32**, 504 (1964)
44. Einstein, A., Fokker, A.D.: Die Nordstromische gravitationstheorie von standpunkte des absoluten differentialkalkuls. *Anna. Phys.* **44**, 321 (1914)
45. Deruelle, N.: Nordstrom's scalar theory of gravity and the equivalence principle. *Gen. Rel. Grav.* **43**, 3337 (2011). [arXiv:1104.4608](https://arxiv.org/abs/1104.4608) [gr-qc]
46. Ehlers, J., Rindler, W.: Local and global light bending in Einstein's and other gravitational theories. *Gen. Rel. Grav.* **29**, 519 (1997)
47. Beem, J.K., Ehrlich, P., Easley, K.: *Global Lorentzian geometry*, Second Edition (1996) (Chapman and Hall/CRC Pure and Applied Mathematics)
48. Carroll, S.M.: *Spacetime and geometry: an introduction to general relativity*. Addison Wesley, San Francisco (2004)
49. Krantz, S.G.: *Complex analysis: the geometric viewpoint, the mathematical association of america*, Second Edition (2004) (Carus Mathematical Monographs)
50. Bengtsson, I.: Anti-de Sitter space, <http://www.fysik.su.se/~ingemar/Kurs.pdf>. Accessed 1 Jul 2015
51. Gibbons, G.: Anti-de Sitter spacetime and its uses. In: *Proceedings for the 2nd Samos Meeting on Cosmology, Geometry and Relativity Conference*, S. Cotsakis, G. W. Gibbons (eds.), *Lect. Notes Phys.* **537**, 102-142 (2000). [arXiv:1110.1206](https://arxiv.org/abs/1110.1206) [hep-th]
52. Bañados, M., Teitelboim, C., Zanelli, J.: The black hole in three-dimensional space-time, *Phys. Rev. Lett.* **69**, 1849 (1992)
53. Krasnov, K.: Holography and Riemann surfaces. *Adv. Theor. Math. Phys.* **4**, 929 (2000). [arXiv:hep-th/0005106](https://arxiv.org/abs/hep-th/0005106)
54. Griffiths, J.B., Podolský, J.: *Exact space-times in Einstein's general relativity*, Cambridge Monographs on Mathematical Physics (1st edn.), Cambridge University Press (2012)
55. Maldacena, J.: The large N limit of superconformal field theories and supergravity. *Adv. Theor. Math. Phys.* **2**, 231 (1998). [arXiv:hep-th/9711200](https://arxiv.org/abs/hep-th/9711200)
56. Klebanov, I.R., Maldacena, J.M.: Solving quantum field theories via curved spacetimes. *Phys. Today* **62**, 28 (2009)
57. Johnson, C.V., Steinberg, P.: What black holes teach about strongly coupled particles, *Phys. Today* **63**, 29 (2010)

Evolution of Black Holes in Anti-de Sitter Spacetime and
the Firewall Controversy

Ong, Y.C.

2016, XXIV, 222 p. 48 illus., 6 illus. in color., Hardcover

ISBN: 978-3-662-48269-8