1
Why (2+1)-dimensional gravity?

The past 25 years have witnessed remarkable growth in our understanding of fundamental physics. The Weinberg–Salam model has successfully unified electromagnetism and the weak interactions, and quantum chromodynamics (QCD) has proven to be an extraordinarily accurate model for the strong interactions. While we do not yet have a viable grand unified theory uniting the strong and electroweak interactions, such a unification no longer seems impossibly distant. At the phenomenological level, the combination of the Weinberg–Salam model and QCD – the Standard Model of elementary particle physics – has been spectacularly successful, explaining experimental results ranging from particle decay rates to high energy scattering cross-sections and even predicting the properties of new elementary particles.

These successes have a common starting point, perturbative quantum field theory. Alone among our theories of fundamental physics, general relativity stands outside this framework. Attempts to reconcile quantum theory and general relativity date back to the 1930s, but despite decades of hard work, no one has yet succeeded in formulating a complete, self-consistent quantum theory of gravity. The task of quantizing general relativity remains one of the outstanding problems of theoretical physics.

The obstacles to quantizing gravity are in part technical. General relativity is a complicated nonlinear theory, and one should expect it to be more difficult than, say, electrodynamics. Moreover, viewed as an ordinary field theory, general relativity has a coupling constant $G^{1/2}$ with dimensions of an inverse mass, and standard power-counting arguments – confirmed by explicit computations – indicate that the theory is nonrenormalizable, that is, that the perturbative quantum theory involves an infinite number of undetermined coupling constants.

But the problem of finding a consistent quantum theory of gravity goes deeper. General relativity is a geometric theory of spacetime, and quantizing gravity means quantizing spacetime itself. In a very basic sense, we do not know what this means. For example:

- Ordinary quantum field theory is local, but the fundamental (diffeomorphism-invariant) physical observables of quantum gravity are necessarily nonlocal;
- Ordinary quantum field theory takes causality as a fundamental postulate, but in quantum gravity the spacetime geometry, and thus the light cones and the causal structure, are themselves subject to quantum fluctuations;
- Time evolution in quantum field theory is determined by a Hamiltonian operator, but for spatially closed universes, the natural candidate for a Hamiltonian in quantum gravity is identically zero when acting on physical states;
- Quantum mechanical probabilities must add up to unity at a fixed time, but in general relativity there is no preferred time-slicing on which to normalize probabilities;
- Scattering theory requires the existence of asymptotic regions in which interactions become negligible and states can be approximated by those of free fields, but the gravitational self-coupling in general relativity never vanishes;
- Perturbative quantum field theory depends on the existence of a smooth, approximately flat spacetime background, but there is no reason to believe that the short-distance limit of quantum gravity even resembles a smooth manifold.

Faced with such problems, it is natural to look for simpler models that share the important conceptual features of general relativity while avoiding some of the computational difficulties. General relativity in 2+1 dimensions – two dimensions of space plus one of time – is one such model. As a generally covariant theory of spacetime geometry, (2+1)-dimensional gravity has the same conceptual foundation as realistic (3+1)-dimensional general relativity, and many of the fundamental issues of quantum gravity carry over to the lower dimensional setting. At the same time, however, the (2+1)-dimensional model is vastly simpler, mathematically and physically, and one can actually write down candidates for a quantum theory. With a few exceptions, (2+1)-dimensional solutions are physically quite different from those in 3+1 dimensions, and the (2+1)-dimensional model is not very helpful for understanding the dynamics of realistic quantum gravity. But for the analysis of conceptual problems – the nature of time, the construction of states and observables, the role of topology and topology change, the relationships among different approaches to quantization – the model has proven highly instructive.

Work on (2+1)-dimensional gravity dates back at least to 1963, when Staruszkiewicz first described the behavior of static solutions with point sources [243]. Work continued intermittently over the next twenty years, but the modern rebirth of the subject can be credited to the seminal work of Deser, Jackiw, 't Hooft, and Witten in the mid-1980s [88, 89, 90, 247, 287, 288]. Over the past decade, (2+1)-dimensional gravity has become an active field of research, drawing insights from general relativity, differential geometry and topology, high energy particle theory, topological field theory, and string theory. The subject is far from being completed, but this book will summarize some of the basic features as they are currently understood.

1.1 General relativity in 2+1 dimensions

The subject of this book is the theory of gravity obtained from the standard Einstein–Hilbert action,

$$I = \frac{1}{16\pi G}\int_M d^3x\,\sqrt{-g}\,(R-2\Lambda) + I_{matter}, \tag{1.1}$$

in three spacetime dimensions. (See appendix C for my conventions for Riemannian geometry.) As in 3+1 dimensions, the resulting Euler–Lagrange equations are the standard Einstein field equations

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = -8\pi G T_{\mu\nu}, \tag{1.2}$$

with a cosmological constant Λ that I will often take to be zero. Just as in ordinary general relativity, the field equations are generally covariant; that is, they are invariant under the action of the group of diffeomorphisms of the spacetime M, which can be viewed as a 'gauge group'.

The fundamental physical difference between general relativity in 2+1 and 3+1 dimensions originates in the fact that the curvature tensor in 2+1 dimensions depends linearly on the Ricci tensor:

$$R_{\mu\nu\rho\sigma} = g_{\mu\rho}R_{\nu\sigma} + g_{\nu\sigma}R_{\mu\rho} - g_{\nu\rho}R_{\mu\sigma} - g_{\mu\sigma}R_{\nu\rho} - \frac{1}{2}(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho})R. \tag{1.3}$$

In particular, this means that every solution of the vacuum Einstein equations with $\Lambda = 0$ is *flat*, and that every solution with a nonvanishing cosmological constant has constant curvature. Physically, a (2+1)-dimensional spacetime has no local degrees of freedom: curvature is concentrated at the location of matter, and there are no gravitational waves. If the spacetime M is topologically trivial, there are, in fact, no gravitational degrees of freedom at all. If M has a nontrivial fundamental group, though, we shall see later that a finite number of global degrees of freedom remain, providing the classical starting point for a quantum theory.

This absence of local degrees of freedom can be verified by a simple counting argument. In n dimensions, the phase space of general relativity is characterized by a spatial metric on a constant-time hypersurface, which has $n(n-1)/2$ components, and its time derivative (or conjugate momentum), which adds another $n(n-1)/2$ degrees of freedom per spacetime point. It is well known, however, that n of the Einstein field equations are constraints on initial conditions rather than dynamical equations, and that n more degrees of freedom can be eliminated by coordinate choices. We are thus left with $n(n-1) - 2n = n(n-3)$ physical degrees of freedom per spacetime point.

If $n = 4$, this gives the four phase space degrees of freedom of ordinary general relativity, two gravitational wave polarizations and their conjugate momenta. If $n = 3$, on the other hand, there are no field degrees of freedom: up to a finite number of possible global degrees of freedom, the geometry is completely determined by the constraints.

Now, a theory of gravity with no propagating degrees of freedom might be expected to have a rather unusual Newtonian limit. This is indeed the case: general relativity in 2+1 dimensions has a Newtonian limit in which there is no force between static point masses. To see this, let us write the metric as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.4}$$

where $\eta_{\mu\nu}$ is the usual flat Minkowski metric and $h_{\mu\nu}$ is a small correction. A gauge can always be chosen in which the n-dimensional field equations take the form

$$\begin{aligned} -\frac{1}{2}\eta^{\mu\nu}\partial_\mu\partial_\nu\bar{h}_{\sigma\tau} + O(h^2) &= 8\pi G T_{\sigma\tau} \\ \eta^{\mu\nu}\partial_\mu\bar{h}_{\nu\sigma} &= 0, \end{aligned} \tag{1.5}$$

where

$$\begin{aligned} \bar{h}_{\sigma\tau} &= h_{\sigma\tau} - \frac{1}{2}\eta_{\sigma\tau}\eta^{\mu\nu}h_{\mu\nu}, \quad \text{i.e.,} \\ h_{\sigma\tau} &= \bar{h}_{\sigma\tau} - \frac{1}{n-2}\eta_{\sigma\tau}\eta^{\mu\nu}\bar{h}_{\mu\nu}. \end{aligned} \tag{1.6}$$

The Newtonian limit is obtained by setting $T_{00} \approx \rho$, where ρ is the mass density; neglecting all other components of the stress–energy tensor; and ignoring time derivatives, which are suppressed by powers of v/c. The only nonzero component of $\bar{h}_{\mu\nu}$ is then

$$\bar{h}_{00} = -4\Phi, \tag{1.7}$$

where Φ is the Newtonian potential,

$$\nabla^2\Phi = 4\pi G\rho. \tag{1.8}$$

In this limit, the geodesic equation

$$\frac{d^2x^\rho}{ds^2} + \Gamma^\rho_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0 \tag{1.9}$$

reduces to

$$\frac{d^2x^i}{dt^2} - \frac{1}{2}\partial_i h_{00} = 0. \tag{1.10}$$

Combining (1.6) and (1.7), we see that

$$\frac{d^2x^i}{dt^2} + \frac{2(n-3)}{n-2}\partial_i\Phi = 0. \tag{1.11}$$

In four dimensions, equation (1.11) gives the standard Newtonian equations of motion, and for $n > 4$ the standard equations may be obtained by rescaling the coupling constant G. In three spacetime dimensions, however, test particles experience no Newtonian force.

This absence of a Newtonian limit does not make the theory trivial: moving particles, for example, can still exhibit nontrivial scattering. In fact, point particle solutions in 2+1 dimensions are good models for parallel cosmic strings in 3+1 dimensions [134]. Cosmic strings are topological solitons that occur in certain gauge theories; it is conjectured that they may have formed during phase transitions in the early universe, where they could have played an important role in the formation of large-scale structure. A straight cosmic string along, say, the z axis is characterized by a stress tensor of the form $T_{00} = -T_{33} = \rho\delta(z)$, and the large tension in the z direction alters the Newtonian limit of ordinary (3+1)-dimensional general relativity. Indeed, the 'effective Newtonian mass density' for an object with pressures $T_{ii} = p_i$ is

$$\rho + \sum_i p_i, \tag{1.12}$$

which vanishes for a cosmic string. The dynamics of a set of such parallel strings may be described in terms of their behavior on the $z = 0$ plane, and for this purpose, (2+1)-dimensional gravity provides a useful model. This is a classical problem, however – at scales at which quantum gravity becomes important, a cosmic string can no longer be represented as a point defect – and I will have little to say about it in the remainder of this book.

1.2 Generalizations

There are several generalizations of (2+1)-dimensional general relativity that restore local degrees of freedom, making the dynamics more like that of realistic (3+1)-dimensional gravity. The quantization of these models is not yet well understood, and they will not be a major topic of this book, but they they warrant a brief introduction.

The first generalization is (2+1)-dimensional dilaton gravity, that is, general relativity coupled to a scalar field φ (the dilaton). In its most general form, the action can be written as [273]

$$I_{DG} = \int_M d^3x\, \sqrt{-g} \left(C[\varphi]R - \frac{\omega[\varphi]}{\varphi} g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi + 2\varphi V[\varphi] \right), \tag{1.13}$$

where C, ω, and V are arbitrary functions of φ. Models of this kind arise naturally in string theory,* with

$$C[\varphi] = \varphi, \quad \omega[\varphi] = -1, \quad V[\varphi] = \Lambda/2, \tag{1.14}$$

while the choice

$$C[\varphi] = \varphi, \quad \omega[\varphi] = \omega_0, \quad V[\varphi] = 0 \tag{1.15}$$

corresponds to three-dimensional Brans–Dicke–Jordan theory. In such models, the scalar field φ becomes a local dynamical degree of freedom, and a judicious choice of couplings can lead to a limit not unlike Newtonian gravity [30]. Many versions of dilaton gravity are known to have black hole solutions (see, for example, [74, 75, 233]), but the quantization of these models has not been studied in any great detail.

A second generalization is unique to 2+1 dimensions, where a 'gravitational Chern–Simons term'

$$I_{GCS} = -\frac{1}{32\pi G\mu} \int_M d^3x\, \epsilon^{\lambda\mu\nu} \Gamma^\rho_{\lambda\sigma} \left(\partial_\mu \Gamma^\sigma_{\rho\nu} + \frac{2}{3}\Gamma^\sigma_{\mu\tau}\Gamma^\tau_{\nu\rho} \right) \tag{1.16}$$

can be added to the gravitational action [91, 92]. This rather unusual-looking term appears as a counterterm in the renormalization of quantum field theory in a (2+1)-dimensional gravitational background [265, 262, 132]. The expression (1.16) does not appear to be generally covariant, but it is, at least when the manifold M is closed: it may be checked that an infinitesimal coordinate change merely adds a total derivative to the Lagrangian, leaving the action unchanged.

Variation of the total action $I + I_{GCS}$ yields the equations of motion

$$G^{\mu\nu} + \mu^{-1}C^{\mu\nu} = 0, \tag{1.17}$$

* In string theory, the field φ is usually denoted as $e^{-2\phi}$.

where $C^{\mu\nu}$ is the conformally invariant Cotton tensor,

$$C^{\mu\nu} = \frac{1}{\sqrt{g}}\epsilon^{\mu\rho\sigma}\nabla_\rho(R^\nu_\sigma - \frac{1}{4}\delta^\nu_\sigma R). \tag{1.18}$$

The simple counting argument that gave us the number of degrees of freedom in Einstein gravity no longer holds: for such third-order equations of motion, the spatial metric and its time derivative must both be treated as configuration space variables with associated canonical momenta, and the analysis becomes more elaborate. Instead, as Deser, Jackiw, and Templeton first observed [91], the linearized equations of motion are those of a massive scalar field,

$$(\Box + \mu^2)\phi = 0, \tag{1.19}$$

where

$$\phi = (\delta_{ij} + \hat{\partial}_i\hat{\partial}_j)h^{ij}, \qquad \text{with} \quad \hat{\partial}_i = \partial_i(-\nabla^2)^{-1/2}. \tag{1.20}$$

The existence of such a massive excitation can be confirmed by looking at the effective interaction of static external sources: one finds a Yukawa attraction with an interaction energy

$$E = -\int d^2x\, T_{00}(-\nabla^2 + \mu^2)^{-1}T_{00}, \tag{1.21}$$

as expected for a massive scalar 'graviton'. This model is commonly called topologically massive gravity ('topological', somewhat misleadingly, because the Chern–Simons term (1.16) is important in topology). Topologically massive gravity has been shown to be perturbatively renormalizable [94, 168], and a number of interesting classical solutions are known. Fairly little is known about the quantization of this system, however, although some progress has been made in understanding the canonical structure and the asymptotic states [46, 93, 137].

1.3 A note on units

It is customary in quantum gravity to express masses in terms of the Planck mass and lengths in terms of the the Planck length. In 2+1 dimensions the gravitational constant G has units of an inverse momentum, and the Planck mass (in units with $c = 1$) is

$$M_{Pl} = \frac{1}{G}, \tag{1.22}$$

while the Planck length is

$$L_{Pl} = \hbar G. \tag{1.23}$$

If a cosmological constant is present, $|\Lambda|^{-1/2}$ has units of length. The theory then has a dimensionless length scale,

$$\ell = \frac{1}{16\pi\hbar G|\Lambda|^{1/2}}. \tag{1.24}$$

Roughly speaking, this scale measures the radius of curvature of the universe.

Throughout this book, I will use units such that $16\pi G = 1$ and $\hbar = 1$, unless otherwise stated. This choice simplifies a number of equations, particularly those involving canonical momenta. In concrete applications, of course – if we are interested in the thermodynamic characteristics of black holes, for instance – it is important to restore factors of G and $\hbar$.

2

Classical general relativity in 2+1 dimensions

If we wish to quantize (2+1)-dimensional general relativity, it is important to first understand the classical solutions of the Einstein field equations. Indeed, many of the best-understood approaches to quantization start with particular representations of the space of solutions. The next three chapters of this book will therefore focus on classical aspects of (2+1)-dimensional gravity. Our goal is not to study the detailed characteristics of particular solutions, but rather to develop an understanding of the generic properties of the space of solutions.

In this chapter, I will introduce two fundamental approaches to classical general relativity in 2+1 dimensions. The first of these, based on the Arnowitt–Deser–Misner (ADM) decomposition of the metric, is familiar from (3+1)-dimensional gravity [9]; the main new feature is that for certain topologies, we will be able to find the general solution of the constraints. The second approach, which starts from the first-order form of the field equations, is also similar to a (3+1)-dimensional formalism, but the first-order field equations become substantially simpler in 2+1 dimensions.

In both cases, the goal is to set up the field equations in a manner that permits a complete characterization of the classical solutions. The next chapters will describe the resulting spaces of solutions in more detail. I will also derive the algebra of constraints in each formalism – a vital ingredient for quantization – and I will discuss the (2+1)-dimensional analogs of total mass and angular momentum.

2.1 The topological setting

Before plunging into a detailed analysis of the field equations, it is useful to ask a preliminary question: what spacetime topologies can occur in (2+1)-dimensional gravity?

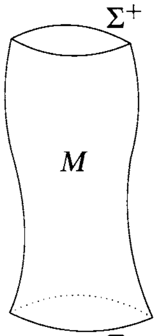

Fig. 2.1. The manifold M has an initial boundary Σ^- and a final boundary Σ^+.

As we shall see below, the interesting cosmological solutions have nontrivial topologies, and to understand their structure, we shall need a number of mathematical tools. For readers unfamiliar with the fundamentals of the topology of manifolds, appendices A and B provide a brief summary of some relevant mathematics. Readers familiar with topology at the level of reference [204] should be able to skip these appendices, although they may serve as useful references for some particular applications.

It is helpful to divide our question into two parts. First, which three-manifolds admit Lorentzian metrics, that is, metrics that have the signature $(-++)$? Second, which of these manifolds admit solutions of the empty space Einstein field equations? Note that in 2+1 dimensions, this second question is more tractable than it might appear. As we saw in chapter 1, the vacuum field equations (with $\Lambda = 0$) require the metric to be flat, so we are really asking which three-manifolds admit flat Lorentzian metrics.

The first of these questions can be answered in full. In appendix B, it is shown that any noncompact three-manifold admits a Lorentzian metric, as does any closed three-manifold. ('Closed' means 'compact and without boundary'.) For compact manifolds with boundary, the problem becomes more interesting. Given a manifold with several boundary components, one can look for a Lorentzian metric for which these components are the past and future spatial boundaries of the universe, as in figure 2.1. Sorkin has shown that a three-manifold M admits a time-orientable Lorentzian metric with spacelike past boundary Σ^- and spacelike future boundary Σ^+ (and no other boundary components) if and only if

$$\chi(\Sigma^-) = \chi(\Sigma^+), \tag{2.1}$$

where $\chi(\Sigma)$ is the Euler number (or Euler–Poincaré characteristic) of Σ [242].

If Σ^- and Σ^+ are both connected, this result prohibits topology change, since the Euler number of a connected surface completely determines its