A BRIEF INTRODUCTION TO MATHEMATICAL RELATIVITY PART 2: GENERAL RELATIVITY

ARICK SHAO

These notes are the second of a pair of brief articles informally introducing the mathematics behind the theory of relativity. Here, we survey *general relativity*, which extends special relativity (covered in the preceding article) by taking gravity into account. Like special relativity, which offered a radically different model of the universe compared to Newtonian mechanics, general relativity also treats gravitation in a revolutionary way.

In order to keep these notes appropriately short, most of the details—e.g., technical definitions, proofs, and computations—are omitted. Background knowledge in differential geometry would be helpful for better understanding various points, but will not be strictly required due to the informal nature of the discussions. To keep the main thread of discussion nontechnical, some more formal characterizations that rely on background in differential geometry or differential equations are relegated to footnotes.

An important disclaimer is that these notes focus primarily on the mathematical, rather than the physical, aspects of the theory. This is mostly by necessity, since I am a mathematician (and not a physicist), with background in partial differential equations and differential geometry (and not in theoretical physics). Consequently, this article will approach the subject from a mathematical viewpoint, in particular in terms of Lorentzian geometry. Physicists will rightfully have a different perspective on many of these points.

For a detailed mathematical reference, the reader is referred to [9] for a formal development of Lorentzian (as well as Riemannian) geometry and some aspects of relativity. For a physics-oriented text that contains a fair amount of mathematical content, see, for instance, [12].

1. Lorentzian Geometry

In classical Newtonian physics, the gravitational field is modeled as an object existing in space and time, lying on top of the background (Euclidean) geometry. In the context of special relativity, a natural hope would be to construct an analogous object within Minkowski geometry. However, such efforts faced a fundamental difficulty, the *equivalence principle*, which also became a main motivation for Einstein in developing general relativity.

The most common informal explanation of this principle is that there is no way to distinguish between motion in a gravitational field and being in an accelerating reference frame. The thought experiment usually given as an example is the following: one can observe, say, what happens on the surface of the earth due to its gravitational pull, or one can observe what happens in a closed elevator in an accelerating rocket deep in space away from gravitational influences. The principle, then, is that barring additional information (e.g., a window in the elevator to see outside), one cannot distinguish between these two situations.

Another (less colorful) statement of the equivalence principle (see [12], for instance) is that all bodies behave the same way within a gravitational field. Thus, one cannot measure the gravitational field by somehow constructing an observer that is "shielded" from its effects.

From this, Einstein's eventual realization is that gravity perhaps should not be modeled as a field on top of Minkowski geometry, but rather as a part of the intrinsic structure of the spacetime. In other words, his revolutionary idea is that gravity could be represented by the geometry of the spacetime itself, in particular by how the spacetime is "curved".

1.1. **Lorentzian Manifolds.** We can now conveniently take advantage of the fact that in the previous article, we already described special relativity in terms of differential geometry. To be more specific, recall that the background setting in special relativity is Minkowski spacetime, consisting of the 4-dimensional manifold $\mathbb{R}^4$ and the flat metric

$$g_M := -dt^2 + dx^2 + dy^2 + dz^2. \tag{1.1}$$

For general relativity, we want to extend our notion of spacetime, that is, we want to replace Minkowski spacetime $(\mathbb{R}^4, g_M)$ by an abstract manifold $(\mathcal{M}, g)$ "of the same type".

First, we want $\mathcal{M}$ to be another 4-dimensional object. The precise mathematical notion is that of a (smooth) manifold. Less formally, this refers to an object $\mathcal{M}$ that locally looks like $\mathbb{R}^4$ at small enough scales.[1] However, on a global scale, $\mathcal{M}$ can look quite unlike $\mathbb{R}^4$; a simple example (having little to do with relativity) is a four-dimensional sphere.

While $\mathcal{M}$ is the "object", we have not yet determined its "geometry". In special relativity, this was given by the Minkowski metric g_M; here, we define on $\mathcal{M}$ a metric g, i.e. a "scalar product on tangent vectors", which will serve as an appropriate abstraction of g_M:

- Recall that for two tangent vectors v and w based at some $p \in \mathbb{R}^4$, the Minkowski metric g_M defined an scalar product $g_M(v, w)$. Similarly, for two tangent vectors v and w based at some $p \in \mathcal{M}$, the metric g defines a similar scalar product.
- Note also from (1.1) that g_M has the signs $(-, +, +, +)$, so that, informally, $\mathbb{R}^4$ has one "negative" direction and three "positive" directions. We wish to mimic the same notion of signature in g, so that it too has one negative and three positive directions.

These two requirements for g define what is called a *Lorentzian metric* on $\mathcal{M}$.[2] The pair $(\mathcal{M}, g)$ is then called a *Lorentzian manifold*, or in physics terminology, a *spacetime*.[3]

[1]For each $p \in \mathcal{M}$, we can smoothly identify a neighborhood of p in $\mathcal{M}$ with a neighborhood of $\mathbb{R}^4$.

[2]More formally, g is, at each point of $\mathcal{M}$, a bilinear form on the corresponding tangent space which has signature $(-, +, +, +)$ in the linear algebraic sense.

[3]This is not entirely accurate, as spacetimes in physics are also equipped with a *time orientation*, i.e., a distinction between "past" and "future" directions.

Remark. Like in special relativity, one should not favor one set of coordinates on $\mathcal{M}$ over another; this is known as the *principle of covariance* in physics. One consequence is that there is again no canonical notion of time or elapsed time; these are defined only relative to a given observer. Furthermore, if $\mathcal{M}$ itself has exotic structure, then even a relative notion of time may not be globally defined on all of $\mathcal{M}$.

Remark. Readers should note that some physics texts use instead the signature convention $(+, -, -, -)$. While this yields an equivalent theory, many objects will have opposite signs.

For those with some background in different geometry, one should compare the above to the more common notion of a *Riemannian manifold* (Σ, γ). Here, Σ is once again a manifold, while γ is a metric (i.e., a scalar product of tangent vectors) on Σ with purely positive signature. The simplest example of this is Euclidean space,

$$\Sigma = \mathbb{R}^3, \qquad \gamma = dx^2 + dy^2 + dz^2.$$

Note that Lorentzian manifolds are geometric generalizations of Minkowski spacetime in the same manner that Riemannian manifolds generalize Euclidean space.

Many concepts from Minkowski geometry extend directly to the this more abstract setting. For instance, since spacetimes again contain both "positive" and "negative" directions, one can directly extend the notion of *causal character* from Minkowski geometry. Indeed, for a tangent vector v in $\mathcal{M}$ (which represents a particular direction at a particular point of $\mathcal{M}$):

- v is *timelike* iff $g(v, v) < 0$.
- v is *spacelike* iff either $g(v, v) > 0$ or $v = 0$.
- v is *null*, or *lightlike*, iff $g(v, v) = 0$ and $v \neq 0$.

Like in special relativity, directions with different causal characters have very different physical interpretations, which we briefly summarize below.

1.2. **Causality.** Analogous to the special relativistic setting, points on the spacetime are called *events*, representing a specific particle at a specific time. *Observers* in this universe are modeled by timelike curves, i.e., curves which everywhere point in a timelike direction. These again represent a single particle existing through time in the universe. Similarly, light is presumed to travel along null curves in $\mathcal{M}$.

Recall that in special relativity, straight timelike lines in $\mathbb{R}^4$ represent an object in free fall. On general manifolds, this notion of straight lines in the strict sense does not exist. However, on Lorentzian (and also Riemannian) manifolds, one does have a natural generalization of straight lines known as *geodesics*. These are a special family of curves for which "its direction does not change"; physically, this represents nonaccelerating motion.[4]

[4]Like in Riemannian geometry, the Lorentzian metric g induces a notion of covariant derivative of vector fields. Geodesics then refer to curves for which its tangent vector field is parallel transported.

As a result, in general relativity, observers in free falling motion are represented by timelike geodesics. Similarly, light rays in general relativity are represented by null geodesics.[5]

The above definitions lead to the study of *causality*. According to our model, an event P can be impacted by an event O in the past if and only if O can be connected to P by a timelike or null (if we are taking light rays into account) curve. While in special relativity, this can be entirely understood in terms of null cones and their interiors in $\mathbb{R}^4$, this picture becomes far more complicated in general relativity.

More specifically, observe that the natural generalization of a null cone in Minkowski spacetime is to take the family of null geodesics emanating from a single point p in $\mathcal{M}$. While near p, this will have the same qualitative structure as a Minkowski null cone, further away from p, the story can be radically different. Indeed, these geodesics may twist in other directions or even turn inwards; see Figure 1. At worst, these geodesics may intersect each other or even converge together toward a single point.

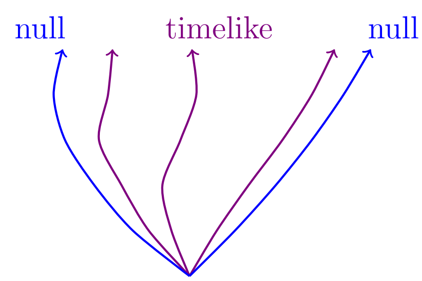

FIGURE 1. Null (blue) and timelike (purple) geodesics emanating from a single point in a spacetime (with spatial directions compacted into one dimension).

1.3. **Curvature and Gravity.** Now that we have established our theory firmly along differential geometric lines, we next discuss how gravity is modeled within the spacetime geometry. One particular aspect of the geometry of $(\mathcal{M}, g)$ is how it is "curved". This information is captured in an object known as the (Riemann) curvature tensor of $(\mathcal{M}, g)$.[6] For the Minkowski spacetime $(\mathbb{R}^4, g_M)$, this curvature vanishes everywhere, hence it is called flat. However, for general Lorentzian manifolds $(\mathcal{M}, g)$, the curvature tensor will be nontrivial.

The main idea, then, is that gravity is represented by this now nonvanishing curvature tensor. A suggestive (but not entirely accurate) way to think about this is to picture a heavy object lying in spacetime as putting a curved dent in the spacetime, like a person would if standing on a soft bed or a trampoline. Then, a ball lying near the object in this dented area would, under the influence of this curvature/gravity, roll toward the object.

[5]One can show that any null curve can be reparametrized so that it becomes a null geodesic.

[6]There are many ways to formally define this curvature. One of the most straightforward is to characterize the curvature as the failure of covariant derivatives to commute with each other; for those familiar with Riemannian geometry, this is a direct analogue of the curvature in Riemannian manifolds.

One should take a moment to appreciate just how revolutionary this notion is. Gravity has, up to this point, been characterized as a force lying on top of a fixed spacetime. Einstein's idea, on the other hand, is to alter the spacetime itself, and to model gravity as its curvature. In particular, gravity is intrinsically built into the geometric structure of the spacetime itself.

2. The Einstein Equations

Through Lorentzian geometry, we have described how the spacetime—the universe—is modeled, as well as how gravity is manifested within the structure of the spacetime. What has not yet been discussed, however, is how matter fields (such as fluids or an electromagnetic fields) fit in this general relativistic model.

For instance, in special relativity, the electric and magnetic fields satisfy *Maxwell's equations*, which can be formulated as a system of vector or tensor partial differential equations on $\mathbb{R}^4$. To port this to general relativity, we again replace the flat background $\mathbb{R}^4$ by its curved counterpart $\mathcal{M}$, and we "geometrize" Maxwell's equations by replacing the usual derivatives on $\mathbb{R}^4$ by Lorentz-geometric equivalents.[7] Similar processes have been used to obtain viable version of other physical theories in the context of general relativity.

However, this still leaves unanswered the question of how matter fields and gravity are related to each other. To address this, Einstein devised another geometric relation, known as the *Einstein field equations*, coupling matter fields to gravity.

2.1. Statement of the Equations.

The Einstein field equations can be most succinctly expressed as the following tensor equation on $\mathcal{M}$:

$$\operatorname{Ric}_g - \frac{1}{2} \operatorname{Sc}_g \cdot g = T. \tag{2.1}$$

On the left-hand side, Ric_g and Sc_g refer to the *Ricci curvature* and the *scalar curvature* of the spacetime $(\mathcal{M}, g)$, respectively, which are tensor and scalar quantities that are defined from the spacetime (Riemann) curvature. On the right, T is the *stress-energy tensor* associated with the matter field(s), which is a quantity arising from the theory of the matter field.[8]

Remark. To simplify notations, we have chosen units so that the constant in front of T on the right-hand side of (2.1) is exactly 1.

Hence, the full set of equations describing the universe contains:

1. The equations describing the matter fields (arising from the relevant physical theory).
2. The Einstein field equations (coupling the matter to gravity).

---

[7] More specifically, this is the covariant derivative via the Levi-Civita connection induced by g.

[8] For instance, T can be derived from the Lagrangian theory considerations.

For example, if we take as matter the electromagnetic field, then the equations to consider would be Maxwell's equations along with the Einstein equations. We note that the electromagnetic field and the curvature are tightly coupled through these equations:

- By "geometrizing" Maxwell's equations, this system now depends heavily on the metric g, i.e., on the geometric structure of the spacetime.
- Since T depends on the electromagnetic field, the Einstein equations demonstrate that the curvature also depends heavily on the electromagnetic field.

One can physically intuit the Einstein equations in multiple ways. For instance, differential geometric considerations show that the curvature describes how nearby geodesics on a manifold behave, i.e., whether they move closer together or pull further apart. By corresponding this with how tidal forces are modeled in Newtonian theory,[9] one arrives at

$$\mathrm{Ric}_g = T.$$

Note this is not (2.1), and this indeed has a serious defect. In particular, the stress-energy tensor T should be divergence-free, corresponding to the local conservation of energy. The Ricci curvature Ric_g, however, is not divergence-free. The extra term on the left-hand side of (2.1) serves to make the resulting quantity divergence-free, like T.

Finally, if there are no matter fields, that is, if $T = 0$, then (2.1) reduces to

$$\mathrm{Ric}_g = 0. \tag{2.2}$$

This is called the *Einstein-vacuum equations.* As we shall see, even vacuum spacetimes can already exhibit an interesting array of behaviors.

Remark. The Einstein equations can also be derived from variational considerations. For those familiar with this theory, one obtains the Einstein-vacuum equations by looking for critical points of the *Einstein-Hilbert action*,

$$\mathcal{S} = \int_{\mathcal{M}} \mathrm{Sc}_g \,, \tag{2.3}$$

where the quantity being varied is the metric g itself. Furthermore, by adding to (2.3) actions associated to matter fields, one obtains through the same variational considerations the Einstein field equations couple with matter fields.

2.2. **Solving the Einstein Equations.** Now that we have determined what the Einstein field equations are, the next question is to see how these equations can be solved. The general hope that in understanding these equations and their solutions, one can gain insights about the nature of the universe. More ambitiously, this knowledge could then lead to the ability to make predictions about the future or the past of our universe.

[9]A more detailed description of the physical argument can be found in [12, Sect. 4.3].

Before becoming too excited with the possibilities, however, we must first address the question of what it means to solve the Einstein equations. As we will see, this is tightly connected to the study of *partial differential equations* (often abbreviated *PDEs*). To simplify this discussion, let us suppose from now on that there are no matter fields present; in other words, we consider only the question of solving the Einstein-vacuum equations (2.2).

Let us fix some local coordinates on our spacetime $\mathcal{M}$. One can then express the metric g with respect to these coordinates, so that its components in these coordinates become (local) scalar functions. Furthermore, in these same coordinates, the components of Ric_g can be expressed in terms of the components of g, along with their first and second (coordinate) derivatives. Thus, from this viewpoint, the vacuum equations (2.2) can be formulated as a system of 10 second-order partial differential equations for the coordinate components of g.

To be a bit more specific, this system has the form

$$\begin{aligned} 0 &= -\frac{1}{2} \sum_{\alpha,\beta} g^{\alpha\beta} (\partial_\alpha \partial_\beta g_{\mu\nu} - \partial_\beta \partial_\nu g_{\mu\alpha} - \partial_\beta \partial_\mu g_{\nu\alpha} + \partial_\mu \partial_\nu g_{\alpha\beta}) \\ &\quad + \frac{1}{2} \sum_{\alpha,\beta,\gamma,\delta} g_{\mu\nu} g^{\alpha\beta} g^{\gamma\delta} (\partial_\alpha \partial_\beta g_{\gamma\delta} - \partial_\beta \partial_\gamma g_{\alpha\delta}) + F_0(g, \partial g). \end{aligned} \tag{2.4}$$

Here, the final term $F_0(g, \partial g)$ on the right-hand side of (2.4) contains a multitude of lower-order terms not containing second derivatives of g.

Remark. Let us highlight here on an important departure from the usual study of PDEs. While usually one deals with PDEs involving unknowns defined on a fixed background (for instance, Minkowski spacetime $\mathbb{R}^4$), here, for the Einstein equations, *we are solving for the background* $(\mathcal{M}, g)$ *itself*! Because the background itself is now part of the unknowns, one requires some extra care to even formulate this problem sensibly.

From the above system of PDE, one can deduce that 6 of the equations contain second derivatives of g with both derivatives in timelike directions. We refer to these 6 equations as the *evolution equations*. The remaining 4 equations also contain second derivatives of g, but not with both in timelike directions; these are called the *constraint equations*.

Let us first consider the evolution equations. Now, those who have some background with PDEs know that there are many different types of equations:

- *Elliptic equations*, which are similar to the Laplace equation ($\Delta\phi = 0$).
- *Parabolic equations*, which are similar to the heat equation ($\partial_t \phi - \Delta\phi = 0$).
- *Hyperbolic equations*, which are similar to the wave equation ($\partial_t^2 \phi - \Delta\phi = 0$).

Elliptic, parabolic, and hyperbolic equations differ in fundamental ways and hence have vastly different theories for dealing with them. For instance, elliptic equations are generally solved given *boundary conditions*, while parabolic and hyperbolic equations are usually solved with *initial conditions*, or some combination of initial and boundary conditions.

Thus, it is important to first determine whether the Einstein evolution equations are elliptic, parabolic, or hyperbolic. The rather unfortunate answer is that *in general,* (2.4) *is none of the above.* This presented a serious conundrum, which (after many contributions by many authors) was finally fully resolved in 1952 by Yvonne Choquet-Bruhat [3].[10] The idea is that one can find a special set of coordinates—called *wave coordinates*—for which the equations see extra cancellations. After these cancellations, the resulting equations for the components of g (called the *reduced Einstein equations*) have the simplified form

$$0 = -\frac{1}{2} \sum_{\alpha,\beta} g^{\alpha\beta} \partial_\alpha \partial_\beta g_{\mu\nu} + F_1(g, \partial g). \tag{2.5}$$

In particular, the evolution equations within (2.5) can be shown to be hyperbolic.

Thus, the appropriate venue for posing and solving the Einstein evolution equations is analogous to that for the wave equation, i.e., that of a *initial value*, or *Cauchy*, *problem.* Roughly speaking, one specifies as initial data the state of the universe at a given "time"—roughly, values of the metric and its "first time derivative"—with the goal being to solve for the spacetime $(\mathcal{M}, g)$ that has this initial data. One can then appeal to classical tools developed for (nonlinear quasilinear) hyperbolic PDE to solve the evolution equations.[11]

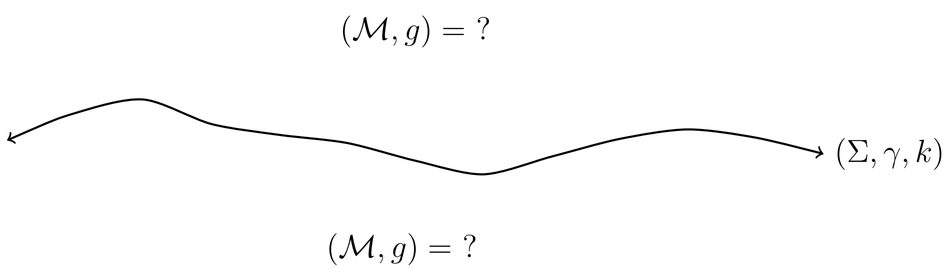

Figure 2. The Cauchy problem for the Einstein equations. The initial data set contains, roughly, the initial (3-dimensional) time slice Σ, its metric γ, and the "first time derivative" k of the metric. The goal is to solve the Einstein-vacuum equations for the spacetime $(\mathcal{M}, g)$ to the future and past of Σ.

The above discussion encapsulates only the evolution equations, leaving the constraint equations yet to be solved. It turns out that these constraint equations can be formulated as (nonlinear elliptic) PDE imposed *on the initial data.* In other words, we cannot take arbitrary initial data for the evolution equations. Rather, we must restrict ourselves only to objects that satisfy the constraint equations.[12]

[10]In addition to establishing one of the seminal results in mathematical relativity, Choquet-Bruhat was also a pioneering woman in mathematics in a period in which such opportunities were rare.

[11]Wave coordinates is by no means the only way to solve the evolution equations. There exist several other viable formulations, for which the equations become hyperbolic or mixed hyperbolic-elliptic.

[12]Solving the constraint equations is a tremendously complicated process in its own right, and the brief discussion given here in no way does this industry justice.

Thus, the full process for solving the Einstein vacuum equations is roughly as follows:

(1) We begin by specifying a 3-dimensional space Σ, which represent the "initial time" on which we will impose the initial data.
(2) We then solve the constraint equations for valid initial data on Σ.
(3) Using this initial data, we solve the evolution equations for the spacetime $(\mathcal{M}, g)$.

Remark. Another important point that should not be overlooked is that the constraint equations are "propagated". In other words, if the constraints hold for the initial data from (2), and one solves the evolution equations as in (3), then one can show that the constraint equations hold everywhere on the resulting spacetime.

Zooming back out to the big picture, in solving the Einstein evolution equations, the basic questions we would like to answer is, in the language of PDEs, that of *well-posedness*:

(1) Given valid initial data, does a solution of the Einstein equations exist?
(2) Is the above solution unique?
(3) Does the solution "depend continuously", in some sense, on the initial data?

Questions (1) or (2) essentially ask whether one can in principle predict the future or past, given the current state of affairs. Question (3) extends this even further and asks whether one can approximately predict the future or past.

The philosophical importance of (3) should not be understated. For example, suppose one wishes to "simulate the universe" by solving the Einstein equations on a supercomputer. In this computational setting, one can of course only approximate both the initial conditions and the equations themselves. As a result, it is important to have reason to believe that the approximate solutions arising from computations actually reflect the real world.

2.3. **Additional Points.** Finally, we address some additional related topics which deviate from the main discussions but are nontheless quite important in relativity.

2.3.1. *Gravitational Waves.* Recall that the Einstein evolution equations were of hyperbolic, or wave-like, nature. Furthermore, it was known by physicists early on that certain linearizations of the Einstein vacuum equations about Minkowski spacetime yielded wave equations. This led physicists, even in the earliest days of general relativity, to predict the presence of "gravitational waves", even if they had no method of observing them. One remarkable triumph in modern experimental physics is the detection of such gravitational waves by the Laser Interferometer Gravitational-Wave Observatory (LIGO) in 2015 [8].

2.3.2. *Non-Vacuum Settings.* The Einstein equations coupled with matter fields can be handled in a manner analogous to the vaccum case. To solve for the metric, one can use the same ideas as before (wave coordinates, etc.). To solve for the matter fields, one must resort to solving equations arising the specific theory behind these fields. One additional point of difficulty is that because the metric and the matter fields are so closely coupled, one would have to solve these equations simultaneously.

2.3.3. *The Cosmological Constant.* One can further generalize the Einstein field equations by adding to (2.1) another term. More specifically, we consider instead[13]

$$\mathrm{Ric}_g - \frac{1}{2}\,\mathrm{Sc}_g \cdot g + \Lambda g = T, \qquad \Lambda \in \mathbb{R}. \tag{2.6}$$

This new parameter Λ is known as the *cosmological constant.*[14]

For nonzero Λ, one can still solve the Einstein equations, but one obtains solutions with vastly different geometric properties. The simplest examples of such solutions are the *de Sitter spacetime* in the case $\Lambda > 0$ and the *Anti-de Sitter spacetime* in the case $\Lambda < 0$; these are the analogues of Minkowski spacetime in the case $\Lambda = 0$.

Remark. In fact, the geometry is so different that what one means by solving the Einstein equations often changes. For instance, when $\Lambda < 0$, the appropriate problem for the evolution equations is actually a mixed initial boundary value problem.

3. Some Special Spacetimes

The simplest example of a solution of the Einstein-vacuum equations is Minkowski spacetime, $(\mathbb{R}^4, g_M)$, the setting of special relativity. In the context of solving the vacuum equations, one arrives at Minkowski spacetime by taking as initial data Euclidean space

$$(\mathbb{R}^3, \gamma), \qquad \gamma := dx^2 + dy^2 + dz^2,$$

with an additional condition on $\mathbb{R}^3$ that can be interpreted as "$\partial_t \gamma = 0$" (once precisely defined, this can be shown to satisfy the constraint equations).

One can then ask whether there are other *non-flat* solutions of the Einstein-vacuum equations. The answer to this question is a resounding "yes", as one can in fact solve for a wide variety of geometrically interesting spacetimes with nontrivial dynamics. Below, we describe some basic but important examples of such solutions.

3.1. **Schwarzschild Spacetimes.** Mere months after Einstein published his theory of general relativity in 1915, Karl Schwarzschild discovered an explicit family of vacuum spacetimes, now known as *Schwarzschild spacetimes.* One arrives at these spacetimes by assuming that the metric g is spherically symmetric (hence reducing the vacuum equations to 2 dimensions) and then solving the vacuum equations in this simplified setting.

[13]Notice that the left-hand side of (2.6) remains divergence-free.

[14]The cosmological constant was originally proposed by Einstein in order to prevent contracting or expanding universes. Although Einstein eventually went away from this idea, more recent observations in cosmology have suggested that the universe may indeed have a positive cosmological constant.

This *Schwarzschild metric* is most commonly expressed in polar-type coordinates as

$$g_S := -\left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta \cdot d\varphi^2), \tag{3.1}$$

where $m \geq 0$ is a constant representing mass.[15] Observe that the case $m = 0$ is precisely the Minkowski metric on $\mathbb{R}^4$, expressed in polar coordinates. However, when $m > 0$, it is not yet clear what the corresponding manifold $\mathcal{M}_S$ is. In particular, note that the expression for g_S in (3.1) becomes singular when $r = 2m$ and $r = 0$.

The first interpretation of Schwarzschild spacetimes is as describing the vacuum region outside of a fixed, static spherical object with mass m and radius $R > 2m$. In this case, one can think of the underlying manifold as

$$\mathcal{M}_S = \mathbb{R} \times (R, \infty) \times \mathbb{S}^2,$$

where the first "$\mathbb{R}$"-component represents the value of t, and while the interval (R, ∞) represents the value of $r > 2m$. Moreover, for the region $r \leq R$, one can attach to $\mathcal{M}_S$ some non-vacuum portion of spacetime representing the spherical object. In particular, when $R > 2m$, this eliminates the blow-up of (3.1) whenever $r = 2m$ and $r = 0$.

On the other hand, with a fuller geometric understanding, Schwarzschild spacetimes can actually be interpreted as purely vacuum solutions. First, note that while we already mentioned that the formula in (3.1) is well-defined as a vacuum spacetime when $r > 2m$, this also remains true in the region $0 < r < 2m$.[16] What remains to be understood, then, are the "singularities of (3.1) that seem to occur at $r = 2m$ and $r = 0$.

First, we observe that $r = 0$ is a true geometric singularity, since the curvature associated with g_S actually blows up as one approaches $r = 0$. As an immediate consequence, the intrinsic geometry of $(\mathcal{M}_S, g_S)$ breaks down entirely there, and the metric g_S cannot possibly be extended to the boundary $r = 0$.

Even though the expression in (3.1) looks equally bad there at first glance, the horizon $r = 2m$ is a very different story that was not well-understood for a long while after the discovery of the Schwarzschild metric. The resolution is that the apparent "blow-up" of (3.1) at $r = 2m$ only reflects the fact that the polar *coordinates* used in (3.1) degenerate. Although these coordinates fare poorly at $r = 2m$, the *metric* g_S itself actually remains regular there. Thus, from the geometric point of view, there is no singularity at $r = 2m$.

In particular, by applying appropriate changes of coordinates, we can smoothly define g_S on the entire region $r > 0$ as a spherically symmetric solution of the Einstein-vacuum equation. Thus, Schwarzschild spacetimes can be understood entirely as solutions of the vacuum equations, even without the presence of additional spherical object of positive mass.

[15]The expression $r^2(d\theta^2 + \sin^2\theta \cdot d\varphi^2)$ is precisely the metric of a (2-)sphere with radius r.

[16]In this case, $1 - \frac{2m}{r} < 0$, so r now becomes the timelike variable, while t becomes spacelike.

Rather than reflecting the mass of some spherical matter field, the parameter m in this purely vacuum viewpoint can now be interpreted as a "gravitational mass".

We now return to the question of what exactly the manifold $\mathcal{M}_S$ is. More specifically, how far can one extend $\mathcal{M}_S$ so that it remains a spherically symmetric Einstein-vacuum spacetime? While the precise answer can be difficult to visualize in its entirety, it can be represented as follows. By omitting the 2-dimensional spherical symmetry in Schwarzschild spacetime (where nothing interesting is happening), the remaining two dimensions of the maximally extended manifold can be characterized by the picture in Figure 3.[17]

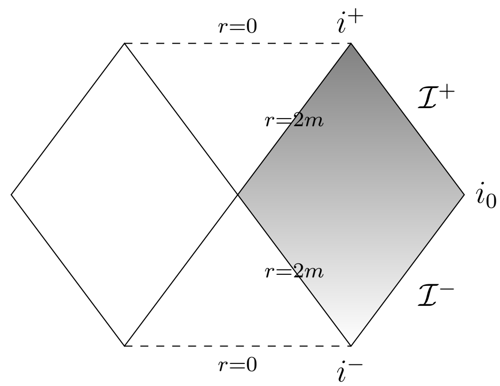

Figure 3. Schwarzschild spacetime, modulo spherical symmetry.

Each point in Figure 3 represents a 2-sphere in $\mathcal{M}_S$. The shaded diamond on the right-hand side refers to the "outer region" $r > 2m$ that we have already discussed.[18] The unshaded diamond on the left-hand side of the diagram represents another mirror copy of this region. Similarly, there are two copies of the "inner region" $0 < r < 2m$, represented by the two triangular sections in Figure 3. For reasons that will become apparent below, these are called the *black hole* (top triangle) and *white hole* (bottom triangle) regions. Finally, the boundary $r = 2m$ dividing the outer and inner regions is known as the *event horizon*.

3.1.1. *The Schwarzschild Black Hole.* We now turn our attention to the upper triangular region in Figure 3 (and, by symmetry, the lower triangular region as well). This region satisfies strikingly different geometric properties compared to Minkowski spacetime.

First, one observes that *light rays and particles that enter this region are then trapped within it*. To be more precise, consider a future-directed null or timelike geodesic ray beginning in the shaded region $r > 2m$ in Figure 3 which at some point crosses the event horizon $r = 2m$ into the upper triangular region $r < 2m$. Then, one can see that after entering this triangular region, the geodesic cannot leave this region at any point in the future.

[17]This is known as the *Penrose diagram* for Schwarzschild spacetime.
[18]The boundary parts in Figure 3, denoted $\mathcal{I}^\pm$, $\iota^\pm$, and ι^0, refer to a formal "infinity" of the spacetime and represents faraway observers. For brevity, we will not discuss this further in these notes.

The second, and even more alarming, observation is that *any such free-falling particle that becomes trapped in this region will reach the singularity $r = 0$ in finite time.* Indeed, any future-directed timelike geodesic beginning in this region will reach $r = 0$ in finite proper time.[19] At this point, this particle will "fail to exist", since $r = 0$ is not a part of the spacetime (recall the scalar curvature blows up as $r \searrow 0$). Thus, any such particle entering this region is not only doomed to remain in this region for the rest of its existence, but will actually reach the singularity and fail to exist in finite time![20]

As a result of these properties, we refer to this upper triangular region as a *black hole*. In particular, since light entering such a black hole cannot escape, one cannot observe the interior of such a black hole directly. The Schwarzschild black hole was the first such example of singular behavior in general relativity.

3.1.2. *Kerr Spacetimes.* Schwarzschild spacetimes are in fact a part of an even larger family of explicit solutions to the Einstein-vacuum equations, discovered in 1963 by Roy Kerr and known as *Kerr spacetimes*. These represent axially symmetric, stationary spacetimes with a rotating black hole (in contrast to the Schwarzschild black hole, which is non-rotating). For brevity, however, we will not discuss these further in greater detail here.

3.2. **Cosmological Spacetimes.** Next, we briefly consider one particular family of *non-vacuum* Einstein equations that served as an early model of cosmology. Indeed, these spacetimes provided scientists with some educated guesses on how our universe began.

To be more specific, we first suppose that our spacetime is *homogeneous* and *isotropic*, that is, the spacetime is independent of both spatial position and direction. The main idea is that these assumptions approximate how the universe would look at large scales. Mathematically, the consequence is that (in the right coordinates), the metric takes the form

$$g = -dt^2 + a(t) \cdot d\Sigma. \tag{3.2}$$

Here, $d\Sigma$ represents the metric of some fixed (Riemannian) 3-manifold of constant curvature representing the level sets of the time function t.

Now, for metrics g of the form (3.2), we consider the Einstein equations, coupled to "dust" matter. Since the dynamics within (3.2) is in the function a, which depends only on t, then these Einstein equations simplify into an ordinary differential equation (ODE) for a. If we fix the geometry of $d\Sigma$, and we set an initial value for a, representing the current state of the universe, we can then solve this ODE in time. These solutions are commonly

[19]Analogous to special relativity, proper time refers to the length of a timelike curve segment with respect to g. Again, this represents the time elapsed as measured by the particle represented by the curve itself.

[20]Philosophically speaking, it is unclear what "failing to exist" means, and it is unclear how singularities are manifested in the real world. However, physicists generally surmise that quantum effects take over within the black hole before one reaches the singularity.

known as *Friedmann–Lemaître–Robertson–Walker*, or *FLRW*, *spacetimes*, named after the (independent groups of) scientists credited for their discovery in the 1920s and 1930s.

Moreover, if we look at these solutions *backwards* in time, we see that $a(t)$ approaches zero after a *finite* amount of time. Thus, if one believes in this cosmological model, then the universe shrinks to nothing and encounters a singularity at some finite time in the past (the "big bang singularity"). In fact, this mathematical model provided the earliest intuitions for the now-popular *Big Bang Theory* in cosmology.

4. Formation of Singularities

In the remaining two sections of this article, we address two general mathematical questions in relativity for which there is much ongoing research. In this section, we delve further into the study of singularities, such as those found within the Schwarzschild black holes.

4.1. **Singularities in Relativity.** From a physical perspective, the Schwarzschild singularity had dire consequences. A freely falling observer that enters the black hole region would, from its own point of view, cease to exist after a finite amount of time, terminating as the curvature of the spacetime blows up. The existence of reasonable observers meeting such a premature end was a serious challenge to the validity of the theory of relativity.

A pressing theoretical question, then, is the following: *Is the Schwarzschild singularity present only because Schwarzschild spacetimes are exceptionally special solutions of the Einstein vacuum equations?* In other words, if one looks at "most" other solutions of the Einstein-vacuum equations, would one not find such singularities?

The more distressing alternative would be that such singularities are "generic". Could there be unavoidably large classes of solutions which exhibit singularities?

Before we answer this question, we must first clarify what one means mathematically by "singularity". The point of view of partial differential equations provides the usual answer: one thinks of a singularity forming when a solution of a PDE fails to exist after only a finite amount of time. However, for the Einstein equations, this fails to be satisfactory, since there is no absolute notion of time. For instance, if a solution of the Einstein-vacuum equation blows up at a finite value of a time coordinate t, then one can transform to a different time coordinate t' which goes to infinity at the blow-up.

To capture a reasonable absolute notion of singularity, we must adopt a more geometric view. More specifically, we characterize singularities in the same manner as the Schwarzschild singularity, through the notion of *geodesic incompleteness*. We say that *a singularity forms if there exist timelike or null geodesics that terminate before its parameter goes to infinity.*[21] Note that while we can certainly reparametrize geodesics (e.g., changing the speed of the

[21] In the timelike case, this occurs when freely falling observers terminate after finite proper time.

free fall), whether the geodesic terminates finitely or not remains unchanged. Thus, this characterization of singularities remains intrinsic to the spacetime.

4.2. **Singularity Theorems.** We now return to the question of whether singularities are generic. This question was answered positively through a number of spectacular singularity theorems, beginning with the *Penrose singularity theorem* in 1965 [10].[22]

Very roughly, the Penrose singularity theorem stated the following:

- *The presence of a trapped surface in the spacetime (along with some other generic conditions) guarantees the existence of a singularity.*[23]

Informally, by a *trapped surface*, we mean a surface S in the spacetime on which gravity is so strong that all light rays emanating from S, both ingoing and outgoing, are being pulled closer together. The simplest examples of such trapped surfaces are two-dimensional spheres in Schwarzschild spacetimes within the black hole region, which by the singularity theorem foreshadow the upcoming gravitational collapse.

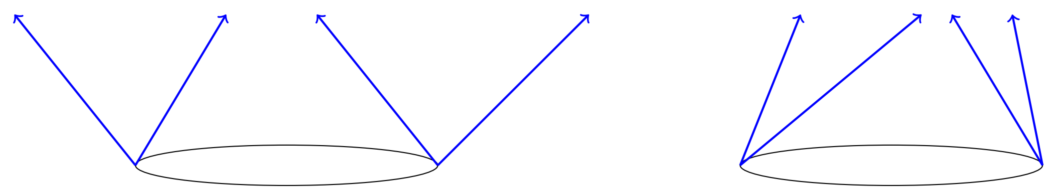

FIGURE 4. Untrapped 2-sphere (left) versus trapped 2-sphere. The blue arrows represent future-pointing null directions emanating from the spheres.

Note in particular that the existence of trapped surfaces is a very generic condition, in that if we slightly perturb a solution of the Einstein equations that contains a trapped surface (such as a Schwarzschild spacetime), then the perturbed spacetime will also contain a trapped surface. As a result, the formation of singularities is a "typical" phenomenon intrinsic to relativity and must be addressed by any serious study of the theory.

On the other hand, while the Penrose singularity theorem gives the existence of a singularity, it gives no information about the nature of this singularity, or how it forms. Obtaining a deeper understanding of the formation of singularities is one of main research directions in mathematical relativity. While there exist many specific examples of singularities (such as the aforementioned Schwarzschild and big bang singularities), the rigorous study of more general singularities is still in its infancy.

4.3. **Dynamical Formation of Singularities.** For the full Schwarzschild spacetimes, any corresponding initial data set for the Einstein-vacuum equations must already contain a trapped surface. In other words, from the PDE point of view, before even solving the

[22]The theorem is named after physicist Sir Roger Penrose.

[23]More specifically, the Penrose theorem establishes null geodesic incompleteness.

equations (i.e., predicting the future), we have already condemned ourselves to a having a singularity form. This suggests the following question: *do there exist "nice" initial data for the Einstein equations containing no trapped surfaces for which the resulting spacetime develops a singularity at some point in the future.*

This was a long-standing open problem which was recently answered affirmatively. In 2009, in a book spanning almost 600 pages, D. Christodoulou [4] constructed a large class of spacetimes initially containing no trapped surfaces but eventually develops one in the future. This result was a breakthrough not only in relativity but also in studying nonlinear PDEs.

4.4. **Cosmic Censorship.** The inevitable possibility of singularities in relativity poses yet another issue. As the laws of physics break down as one approaches this singularity (moreover, quantum effects are expected to take precedence near singularities), if we could observe such a singuliarty directly, then one may argue that we could lose the ability to "predict the future". In other words, general relativity could fail to be deterministic.

This led Sir Roger Penrose in 1969 to conjecture [11] that if a singularity does form, then it must form within a black hole, where no light can escape and the singularity could not be observed. This conjecture, known as *cosmic censorship*, is split into two separate statements:

- *Weak cosmic censorship*: singularities must be hidden within event horizons and hence cannot be observed from the outside.
- *Strong cosmic censorship*: general relativity is deterministic—given appropriate initial data for the universe, we can predict the future and past of the universe.

There are numerous issues associated with establishing cosmic censorship. For example, a less serious but irritating issue is that despite the names "weak" and "strong" cosmic censorship, it is known that neither statement implies the other.

A far more serious issue, though, is that both the weak and strong cosmic censorship statements, informally stated as above, are false. For instance, there are numerous isolated examples of spacetimes with *naked singularities* that are not hidden within black holes. Undeterred, physicists then sought to refine the notion of cosmic censorship:

- *Under "reasonable generic" conditions, singularities must be hidden in black holes.*

In general, these refined cosmic censorship conjectures are widely open problems.[24] An immediate issue arises from the vague terms "reasonable" and "generic", as the precise mathematical formulation of the conjectures remains unclear. Thus, the question is not just to find mathematical proofs for these conjectures, but to also figure out what the mathematical statement we wish to prove is in the first place.

[24]There do exist proofs of cosmic censorship statements in restricted settings. For instance, in the works of Christodoulou in the 1990s on spherically symmetric Einstein-scalar field spacetimes, there do exist spacetimes with naked singularities, but these disappear after a slight perturbation of the spacetime.

5. Dynamics and Asymptotics

In the remaining section on topics of ongoing investigation, we return to the perpective of mathematical relativity from partial differential equations. When studying an evolutionary PDE, such as the heat, wave, or Schrödinger equations, the ultimate goal is to fully understand the dynamics of solutions. Given any initial data representing the state at a given time, we wish to determine how the solution behaves at later (or in some cases, earlier) times, that is, we wish to "predict the future".

The Einstein-vacuum equations also fit within this frame work. From our previous discussion on well-posedness, "predicting the future" is in principle possible, in the sense that appropriate initial data results in a unique solution of the Einstein equations.[25] However, such existence and uniqueness results give no indication of how these solutions behave, which is a much more difficult question that is actively researched today. Thus, well-posedness results, while foundational, only represent the beginning of a long, ongoing story.

5.1. **The Final State Conjecture.**

If one were to impose general, perhaps badly behaved, initial data, then the corresponding spacetime solving the Einstein equations could behave poorly as well, at least for short times. However, one may ask if one sees something different over long times. Do the Einstein equations contain some special structure that allow generic solutions to resolve in some universal way for long times? Physically speaking, what do the PDEs tell us about the eventual "fate of the universe"?

One can gain some inspiration from simpler nonlinear PDEs. For instance, for various model nonlinear wave equations, the expected result is the so-called *soliton resolution conjecture*: roughly, a generic solution is expected to, asymptotically in time, resolve to a sum of *solitons* (i.e., solutions which are time-independent modulo symmetries) plus a "radiating part" that decays like a linear wave. This is currently a very active area of research in nonlinear PDEs, in which significant recent progress has been made.

One can also formulate a corresponding statement for the Einstein-vacuum equations. Indeed, the *final state conjecture* states that a generic solution of the Einstein-vacuum equations arising from asymptotically flat, or Euclidean, initial data should, asymptotically in time, resolve into a superposition of Kerr black holes (the "solitons") and a "radiative" part.

Given the complexity of the Einstein equations, the final state conjecture is currently an exceedingly difficult problem and is well out of reach. However, research efforts in mathematical relativity have already begun inching slowly toward this goal by first attacking special cases of the final state conjecture.

5.2. **Stability.**

For instance, rather than considering generic solutions of the vacuum equations, we consider subclasses of solutions that are "close to" explicit solutions, such as

[25] Excepting global issues that would arise from a potential failure of cosmic censorship.

Minkowski, Schwarzschild, and Kerr spacetimes. More specifically, if we begin with initial data that is sufficiently near data for one of these spacetimes, then is the resulting spacetime close to the explicit solution for all times? This is the question of *global stability*.[26]

The crowning achievement in this direction is the work of Christodoulou and Klainerman, who in 1993 proved, in a 526-page book, the global stability of Minkowski spacetime [5]. They proved that for initial data "sufficiently close" to Euclidean space, the solution indeed remains close to Minkowski spacetime for all time. Furthermore, asymptotically in time, the solution in fact decays to Minkowski spacetime.

On the other hand, the global stability of the larger family of Kerr spacetimes remains an open problem at this time, though significant recent progress has been made (see [7]). A fundamental difficulty of this more general problem is that if one begins with data close to one Kerr spacetime, then the resulting solution could asymptotically decay to a *different* Kerr spacetime, with different mass and angular momentum.

5.3. **Rigidity.** If we view the final state conjecture from the perspective of soliton resolution, then the Kerr spacetimes correspond to the "solitons" in the theory of nonlinear wave equations. In order to see whether this analogy is viable, though, we must first determine whether these Kerr black holes are essentially the *only* stationary solutions of the Einstein-vacuum equations.[27] This is the question of *black hole rigidity*.

Roughly, the conjecture is that the only asymptotically flat, stationary, Einstein-vacuum black hole spacetimes are the Kerr spacetimes. There have been many partial results in this direction (see, e.g., [1, 2, 6]), but as of now, a full proof of this conjecture remains open.

6. Conclusion

Unfortunately, these short notes can only barely scratch the surface of this fascinating area of research. Mathematical relativity is a young and fast-growing field, with many compelling questions that are of interest in both mathematics and theoretical physics. Hopefully, the interested reader would, upon reading these notes, be inspired to learn more about relativity and to further explore the research being conducted in this area.

References

1. S. Alexakis, A. Ionescu, and S. Klainerman, *Uniqueness of smooth stationary black holes in vacuum: small perturbations of the Kerr spaces*, Commun. Math. Phys. **299** (2010), no. 1, 89–127.
2. ———, *Rigidity of stationary black holes with small angular momentum on the horizon*, Duke Math. J. **163** (2014), no. 14, 2603–2615.

[26] This is in contrast to the local, or short-time, stability mentioned earlier in the discussion of continuous dependence of the solution on the initial data.

[27] One can think of "stationary" as meaning "time-independent", although in the relativistic setting, we have to be careful to define this in a geometric, coordinate-independent manner.

3. Y. Choquét-Bruhat, *Théorème d'existence pour certains systèmes d'équations aux dérivées partielles nonlinéaires*, Acta Math. **88** (1952), 141–225.
4. D. Christodoulou, *The formation of black holes in general relativity*, EMS Publishing House, 2009.
5. D. Christodoulou and S. Klainerman, *Global nonlinear stability of the Minkowski space*, Princeton University Press, 1993.
6. P. Chruściel and J. Costa, *On uniqueness of stationary vacuum black holes*, Astérisque **321** (2008), 195–265.
7. M. Dafermos, G. Holzegel, and I. Rodnianski, *The linear stability of the Schwarzschild solution to gravitational perturbations*, arXiv:1601.06467, 2016.
8. B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Observation of gravitational waves from a binary black hole merger*, Phys. Rev. Lett. **116** (2016), no. 6, 061102.
9. B. O'Neill, *Semi-Riemannian geometry with applications to relativity*, Academic Press, 1983.
10. R. Penrose, *Gravitational collapse and space-time singularities*, Phys. Rev. Lett. **14** (1965), no. 3, 57–59.
11. ———, *Gravitational collapse: The role of general relativity*, Riv. Nuovo Cim. **1** (1969), 252–276.
12. R. Wald, *General relativity*, The University of Chicago Press, 1984.