

A No-Nonsense Introduction to General Relativity

Sean M. Carroll
Enrico Fermi Institute and Department of Physics,
University of Chicago, Chicago, IL, 60637
carroll@theory.uchicago.edu

©2001

1 Introduction

General relativity (GR) is the most beautiful physical theory ever invented. Nevertheless, it has a reputation of being extremely difficult, primarily for two reasons: tensors are everywhere, and spacetime is curved. These two facts force GR people to use a different language than everyone else, which makes the theory somewhat inaccessible. Nevertheless, it is possible to grasp the basics of the theory, even if you're not Einstein (and who is?).

GR can be summed up in two statements: 1) Spacetime is a curved pseudo-Riemannian manifold with a metric of signature $(-+++)$. 2) The relationship between matter and the curvature of spacetime is contained in the equation

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi GT_{\mu\nu} . \quad (1)$$

However, these statements are incomprehensible unless you sling the lingo. So that's what we shall start doing. Note, however, that this introduction is a very pragmatic affair, intended to give you some immediate feel for the language of GR. It does not substitute for a deep understanding – that takes more work!

Administrative notes: physicists love to set constants to unity, and it's a difficult habit to break once you start. I will *not* set Newton's constant $G = 1$. However, it's ridiculous not to set the speed of light $c = 1$, so I'll do that. For further reference, recommended texts include *A First Course in General Relativity* by Bernard Schutz, at an undergrad level; and graduate texts *General Relativity* by Wald, *Gravitation and Cosmology* by Weinberg, *Gravitation* by Misner, Thorne, and Wheeler, and *Introducing Einstein's Relativity* by D'Inverno. Of course best of all would be to rush to <http://pancake.uchicago.edu/~carroll/notes/>, where you will find about one semester's worth of free GR notes, of which this introduction is essentially an abridgment.

2 Special Relativity

Special relativity (SR) stems from considering the speed of light to be invariant in all reference frames. This naturally leads to a view in which space and time are joined together to form spacetime; the conversion factor from time units to space units is c (which equals 1, right? couldn't be simpler). The coordinates of spacetime may be chosen to be

$$\begin{aligned} x^0 &\equiv ct = t \\ x^1 &\equiv x \\ x^2 &\equiv y \\ x^3 &\equiv z. \end{aligned} \quad (2)$$

These are **Cartesian coordinates**. Note a few things: these indices are *superscripts*, not exponents. The indices go from zero to three; the collection of all four coordinates is denoted x^μ . Spacetime indices are always in Greek; occasionally we will use Latin indices if we mean only the spatial components, e.g. $i = 1, 2, 3$.

The stage on which SR is played out is a specific four dimensional manifold, known as **Minkowski spacetime** (or sometimes “Minkowski space”). The x^μ are coordinates on this manifold. The elements of spacetime are known as **events**; an event is specified by giving its location in both space and time. Vectors in spacetime are always fixed at an event; there is no such thing as a “free vector” that can move from place to place. Since Minkowski space is four dimensional, these are generally known as **four-vectors**, and written in components as V^μ , or abstractly as just V .

We also have the **metric** on Minkowski space, $\eta_{\mu\nu}$. The metric gives us a way of taking the norm of a vector, or the dot product of two vectors. Written as a matrix, the Minkowski metric is

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3)$$

Then the dot product of two vectors is defined to be

$$A \cdot B \equiv \eta_{\mu\nu} A^\mu B^\nu = -A^0 B^0 + A^1 B^1 + A^2 B^2 + A^3 B^3. \quad (4)$$

(We always use the **summation convention**, in which identical upper and lower indices are implicitly summed over all their possible values.) This is especially useful for taking the infinitesimal (distance)² between two points, also known as the **spacetime interval**:

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \quad (5)$$

$$= -dt^2 + dx^2 + dy^2 + dz^2. \quad (6)$$

In fact, an equation of the form (6) is often called “the metric.” The metric contains all of the information about the geometry of the manifold. The Minkowski metric is of course just the spacetime generalization of the ordinary inner product on flat Euclidean space, which we can think of in components as the Kronecker delta, δ_{ij} . We say that the Minkowski metric has **signature** $(-+++)$, sometimes called “Lorentzian,” as opposed to the Euclidian signature with all plus signs. (The overall sign of the metric is a matter of convention, and many texts use $(+---)$.)

Notice that for a particle with fixed spatial coordinates x^i , the interval elapsed as it moves forward in time is negative, $ds^2 = -dt^2 < 0$. This leads us to define the **proper time** τ via

$$d\tau^2 \equiv -ds^2. \quad (7)$$

The proper time elapsed along a trajectory through spacetime will be the actual time measured by an observer on that trajectory. Some other observer, as we know, will measure a different time.

Some verbiage: a vector V^μ with negative norm, $V \cdot V < 0$, is known as **timelike**. If the norm is zero, the vector is **null**, and if it's positive, the vector is **spacelike**. Likewise, trajectories with negative ds^2 (note – *not* proper time!) are called timelike, etc. These concepts lead naturally to the concept of a **spacetime diagram**, with which you are presumably familiar. The set of null trajectories leading into and out of an event constitute a **light cone**, terminology which becomes transparent in the context of a spacetime diagram such as Figure 1.

A path through spacetime is specified by giving the four spacetime coordinates as a function of some parameter, $x^\mu(\lambda)$. A path is characterized as timelike/null/spacelike when its tangent vector $dx^\mu/d\lambda$ is timelike/null/spacelike. For timelike paths the most convenient parameter to use is the proper time τ , which we can compute along an arbitrary timelike path via

$$\tau = \int \sqrt{-ds^2} = \int \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}} d\lambda . \quad (8)$$

The corresponding tangent vector $U^\mu = dx^\mu/d\tau$ is called the **four-velocity**, and is automatically normalized:

$$\eta_{\mu\nu} U^\mu U^\nu = -1 , \quad (9)$$

as you can check.

A related vector is the **momentum four-vector**, defined by

$$p^\mu = mU^\mu , \quad (10)$$

where m is the mass of the particle. The mass is a fixed quantity independent of inertial frame, what you may be used to thinking of as the “rest mass.” The **energy** of a particle is simply p^0 , the timelike component of its momentum vector. In the particle’s rest frame we have $p^0 = m$; recalling that we have set $c = 1$, we find that we have found the famous equation $E = mc^2$. In a moving frame we can find the components of p^μ by performing a Lorentz transformation; for a particle moving with three-velocity $v = dx/dt$ along the x axis we have

$$p^\mu = (\gamma m, v\gamma m, 0, 0) , \quad (11)$$

where $\gamma = 1/\sqrt{1-v^2}$. For small v , this gives $p^0 = m + \frac{1}{2}mv^2$ (what we usually think of as rest energy plus kinetic energy) and $p^1 = mv$ (what we usually think of as Newtonian momentum).

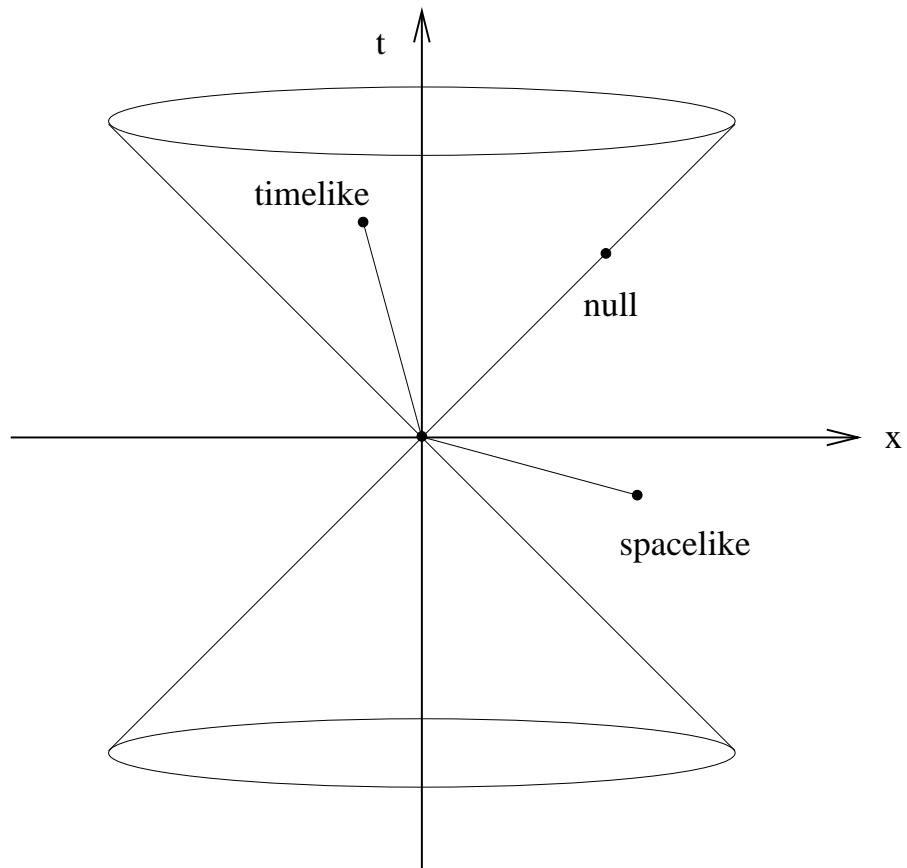


Figure 1: A lightcone, portrayed on a spacetime diagram. Points which are spacelike-, null-, and timelike-separated from the origin are indicated.

3 Tensors

The transition from flat to curved spacetime means that we will eventually be unable to use Cartesian coordinates; in fact, some rather complicated coordinate systems become necessary. Therefore, for our own good, we want to make all of our equations **coordinate invariant** – i.e., if the equation holds in one coordinate system, it will hold in any. It also turns out that many of the quantities that we use in GR will be **tensors**. Tensors may be thought of as objects like vectors, except with possibly more indices, which transform under a change of coordinates $x^\mu \rightarrow x^{\mu'}$ according to the following rule, the **tensor transformation law**:

$$S^{\mu'}{}_{\nu'\rho'} = \frac{\partial x^{\mu'}}{\partial x^\mu} \frac{\partial x^\nu}{\partial x^{\nu'}} \frac{\partial x^\rho}{\partial x^{\rho'}} S^\mu{}_{\nu\rho} . \quad (12)$$

Note that the unprimed indices on the right are dummy indices, which are summed over. The pattern in (12) is pretty easy to remember, if you think of “conservation of indices”: the upper and lower *free* indices (not summed over) on each side of an equation must be the same. This holds true for any equation, not just the tensor transformation law. Remember also that upper indices can only be summed with lower indices; if you have two upper or lower indices that are the same, you goofed. Since there are in general no preferred coordinate systems in GR, it behooves us to cast all of our equations in tensor form, because *if an equation between two tensors holds in one coordinate system, it holds in all coordinate systems*.

Tensors are not very complicated; they’re just generalizations of vectors. (Note that scalars qualify as tensors with no indices, and vectors are tensors with one upper index; a tensor with two indices can be thought of as a matrix.) However, there is an entire language associated with them which you must learn. If a tensor has n upper and m lower indices, it is called a (n, m) tensor. The upper indices are called **contravariant** indices, and the lower ones are **covariant**; but everyone just says “upper” and “lower,” and so should you. Tensors of type (n, m) can be **contracted** to form a tensor of type $(n - 1, m - 1)$ by summing over one upper and one lower index:

$$S^\mu = T^{\mu\lambda}{}_\lambda . \quad (13)$$

The contraction of a two-index tensor is often called the **trace**. (Which makes sense if you think about it.)

If a tensor is the same when we interchange two indices,

$$S_{\dots\alpha\beta\dots} = S_{\dots\beta\alpha\dots} , \quad (14)$$

it is said to be **symmetric** in those two indices; if it changes sign,

$$S_{\dots\alpha\beta\dots} = -S_{\dots\beta\alpha\dots} , \quad (15)$$

we call it **antisymmetric**. A tensor can be symmetric or antisymmetric in many indices at once. We can also take a tensor with no particular symmetry properties in some set of indices

and pick out the symmetric/antisymmetric piece by taking appropriate linear combinations; this procedure of symmetrization or antisymmetrization is denoted by putting parentheses or square brackets around the relevant indices:

$$\begin{aligned} T_{(\mu_1\mu_2\cdots\mu_n)} &= \frac{1}{n!} (T_{\mu_1\mu_2\cdots\mu_n} + \text{sum over permutations of } \mu_1 \cdots \mu_n) \\ T_{[\mu_1\mu_2\cdots\mu_n]} &= \frac{1}{n!} (T_{\mu_1\mu_2\cdots\mu_n} + \text{alternating sum over permutations of } \mu_1 \cdots \mu_n). \end{aligned} \quad (16)$$

By ‘‘alternating sum’’ we mean that permutations which are the result of an odd number of exchanges are given a minus sign, thus:

$$T_{[\mu\nu\rho]\sigma} = \frac{1}{6} (T_{\mu\nu\rho\sigma} - T_{\mu\rho\nu\sigma} + T_{\rho\mu\nu\sigma} - T_{\nu\mu\rho\sigma} + T_{\nu\rho\mu\sigma} - T_{\rho\nu\mu\sigma}) . \quad (17)$$

The most important tensor in GR is the metric $g_{\mu\nu}$, a generalization (to arbitrary coordinates and geometries) of the Minkowski metric $\eta_{\mu\nu}$. Although $\eta_{\mu\nu}$ is just a special case of $g_{\mu\nu}$, we denote it by a different symbol to emphasize the importance of moving from flat to curved space. The metric is a symmetric two-index tensor. An important fact is that it is always possible to find coordinates such that, at one specified point p , the components of the metric are precisely those of the Minkowski metric (3) and the first derivatives of the metric vanish. In other words, the metric will look flat at precisely that point; however, in general the second derivatives of $g_{\mu\nu}$ cannot be made to vanish, a manifestation of curvature.

Even if spacetime is flat, the metric can still have nonvanishing derivatives if the coordinate system is non-Cartesian. For example, in spherical coordinates (on space) we have

$$\begin{aligned} t &= t \\ x &= r \sin \theta \cos \phi \\ y &= r \sin \theta \sin \phi \\ z &= r \cos \theta , \end{aligned} \quad (18)$$

which leads directly to

$$ds^2 = -dt^2 + dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 , \quad (19)$$

or

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2 \theta \end{pmatrix} . \quad (20)$$

Notice that, while we could use the tensor transformation law (12), it is often more straightforward to find new tensor components by simply plugging in our coordinate transformations to the differential expression (*e.g.* $dz = \cos \theta dr - r \sin \theta d\theta$).

Just as in Minkowski space, we use the metric to take dot products:

$$A \cdot B \equiv g_{\mu\nu} A^\mu B^\nu . \quad (21)$$

This suggests, as a shortcut notation, the concept of **lowering indices**; from any vector we can construct a $(0, 1)$ tensor defined by contraction with the metric:

$$A_\nu \equiv g_{\mu\nu} A^\mu , \quad (22)$$

so that the dot product becomes $g_{\mu\nu} A^\mu B^\nu = A_\nu B^\nu$. We also define the **inverse metric** $g^{\mu\nu}$ as the matrix inverse of the metric tensor:

$$g^{\mu\nu} g_{\nu\rho} = \delta_\rho^\mu , \quad (23)$$

where δ_ρ^μ is the (spacetime) Kronecker delta. (Convince yourself that this expression really does correspond to matrix multiplication.) Then we have the ability to raise indices:

$$A^\mu = g^{\mu\nu} A_\nu . \quad (24)$$

Note that raising an index on the metric yields the Kronecker delta, so we have

$$g^{\mu\nu} g_{\mu\nu} = \delta_\mu^\mu = 4 . \quad (25)$$

Despite the ubiquity of tensors, it is sometimes useful to consider non-tensorial objects. An important example is the determinant of the metric tensor,

$$g \equiv \det (g_{\mu\nu}) . \quad (26)$$

A straightforward calculation shows that under a coordinate transformation $x^\mu \rightarrow x^{\mu'}$, this doesn't transform by the tensor transformation law (under which it would have to be invariant, since it has no indices), but instead as

$$g \rightarrow \left[\det \left(\frac{\partial x^{\mu'}}{\partial x^\mu} \right) \right]^{-2} g . \quad (27)$$

The factor $\det(\partial x^{\mu'}/\partial x^\mu)$ is the Jacobian of the transformation. Objects with this kind of transformation law (involving powers of the Jacobian) are known as **tensor densities**; the determinant g is sometimes called a “scalar density.” Another example of a density is the volume element $d^4x = dx^0 dx^1 dx^2 dx^3$:

$$d^4x \rightarrow \det \left(\frac{\partial x^{\mu'}}{\partial x^\mu} \right) d^4x . \quad (28)$$

To define an invariant volume element, we can therefore multiply d^4x by the square root of minus g , so that the Jacobian factors cancel out:

$$\sqrt{-g} d^4x \rightarrow \sqrt{-g} d^4x . \quad (29)$$

In Cartesian coordinates, for example, we have $\sqrt{-g} d^4x = dt dx dy dz$, while in polar coordinates this becomes $r^2 \sin \theta dt dr d\theta d\phi$. Thus, integrals of functions over spacetime are of the form $\int f(x^\mu) \sqrt{-g} d^4x$. (“Function,” of course, is the same thing as “scalar.”)

Another object which is unfortunately not a tensor is the partial derivative $\partial/\partial x^\mu$, often abbreviated to ∂_μ . Acting on a scalar, the partial derivative returns a perfectly respectable $(0, 1)$ tensor; using the conventional chain rule we have

$$\partial_\mu \phi \rightarrow \partial_{\mu'} \phi = \frac{\partial x^\mu}{\partial x^{\mu'}} \partial_\mu \phi , \quad (30)$$

in agreement with the tensor transformation law. But on a vector V^μ , given that $V^\mu \rightarrow \frac{\partial x^{\mu'}}{\partial x^\mu} V^\mu$, we get

$$\begin{aligned} \partial_\mu V^\nu \rightarrow \partial_{\mu'} V^{\nu'} &= \left(\frac{\partial x^\mu}{\partial x^{\mu'}} \partial_\mu \right) \left(\frac{\partial x^{\nu'}}{\partial x^\nu} V^\nu \right) \\ &= \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^{\nu'}}{\partial x^\nu} (\partial_\mu V^\nu) + \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial^2 x^{\nu'}}{\partial x^\nu \partial x^\mu} V^\mu . \end{aligned} \quad (31)$$

The first term is what we want to see, but the second term ruins it. So we define a **covariant derivative** to be a partial derivative plus a correction that is linear in the original tensor:

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma_{\mu\lambda}^\nu V^\lambda . \quad (32)$$

Here, the symbol $\Gamma_{\mu\lambda}^\nu$ stands for a collection of numbers, called **connection coefficients**, with an appropriate non-tensorial transformation law chosen to cancel out the non-tensorial term in (31). Thus we need to have

$$\Gamma_{\mu'\lambda'}^{\nu'} = \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^\lambda}{\partial x^{\lambda'}} \frac{\partial x^{\nu'}}{\partial x^\nu} \Gamma_{\mu\lambda}^\nu - \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^\lambda}{\partial x^{\lambda'}} \frac{\partial^2 x^{\nu'}}{\partial x^\nu \partial x^\mu} . \quad (33)$$

Then $\nabla_\mu V^\nu$ is guaranteed to transform like a tensor. The same kind of trick works to define covariant derivatives of tensors with lower indices; we simply introduce a minus sign and change the dummy index which is summed over:

$$\nabla_\mu \omega_\nu = \partial_\mu \omega_\nu - \Gamma_{\mu\nu}^\lambda \omega_\lambda . \quad (34)$$

If there are many indices, for each upper index you introduce a term with a single $+\Gamma$, and for each lower index a term with a single $-\Gamma$:

$$\begin{aligned} \nabla_\sigma T^{\mu_1 \mu_2 \dots \mu_k}_{\nu_1 \nu_2 \dots \nu_l} &= \partial_\sigma T^{\mu_1 \mu_2 \dots \mu_k}_{\nu_1 \nu_2 \dots \nu_l} \\ &+ \Gamma_{\sigma\lambda}^{\mu_1} T^{\lambda \mu_2 \dots \mu_k}_{\nu_1 \nu_2 \dots \nu_l} + \Gamma_{\sigma\lambda}^{\mu_2} T^{\mu_1 \lambda \dots \mu_k}_{\nu_1 \nu_2 \dots \nu_l} + \dots \\ &- \Gamma_{\sigma\nu_1}^\lambda T^{\mu_1 \mu_2 \dots \mu_k}_{\lambda \nu_2 \dots \nu_l} - \Gamma_{\sigma\nu_2}^\lambda T^{\mu_1 \mu_2 \dots \mu_k}_{\nu_1 \lambda \dots \nu_l} - \dots . \end{aligned} \quad (35)$$

This is the general expression for the covariant derivative.

What are these mysterious connection coefficients? Fortunately they have a natural expression in terms of the metric and its derivatives:

$$\Gamma_{\mu\nu}^{\sigma} = \frac{1}{2}g^{\sigma\rho}(\partial_{\mu}g_{\nu\rho} + \partial_{\nu}g_{\rho\mu} - \partial_{\rho}g_{\mu\nu}) . \quad (36)$$

It is left up to you to check that the mess on the right really does have the desired transformation law. You can also verify that the connection coefficients are symmetric in their lower indices, $\Gamma_{\mu\nu}^{\sigma} = \Gamma_{\nu\mu}^{\sigma}$. These coefficients can be nonzero even in flat space, if we have non-Cartesian coordinates. In principle there can be other kinds of connection coefficients, but we won't worry about that here; the particular choice (36) are sometimes called **Christoffel symbols**, and are the ones we always use in GR. With these connection coefficients, we get the nice feature that the covariant derivative of the metric and its inverse are always zero, known as **metric compatibility**:

$$\nabla_{\sigma}g_{\mu\nu} = 0 , \quad \nabla_{\sigma}g^{\mu\nu} = 0 . \quad (37)$$

So, given any metric $g_{\mu\nu}$, we proceed to calculate the connection coefficients so that we can take covariant derivatives. Many of the familiar equations of physics in flat space continue to hold true in curved space once we replace partial derivatives by covariant ones. For example, in special relativity the electric and magnetic vector fields \vec{E} and \vec{B} can be collected into a single two-index antisymmetric tensor $F_{\mu\nu}$:

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix} , \quad (38)$$

and the electric charge density ρ and current \vec{J} into a four-vector J^{μ} :

$$J^{\mu} = (\rho, \vec{J}) . \quad (39)$$

In this notation, Maxwell's equations

$$\begin{aligned} \nabla \times \mathbf{B} - \partial_t \mathbf{E} &= 4\pi \mathbf{J} \\ \nabla \cdot \mathbf{E} &= 4\pi \rho \\ \nabla \times \mathbf{E} + \partial_t \mathbf{B} &= 0 \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \quad (40)$$

shrink into two relations,

$$\begin{aligned} \partial_{\mu}F^{\nu\mu} &= 4\pi J^{\nu} \\ \partial_{[\mu}F_{\nu\lambda]} &= 0 . \end{aligned} \quad (41)$$

These are true in Minkowski space, but the generalization to a curved spacetime is immediate; just replace $\partial_\mu \rightarrow \nabla_\mu$:

$$\begin{aligned}\nabla_\mu F^{\nu\mu} &= 4\pi J^\nu \\ \nabla_{[\mu} F_{\nu\lambda]} &= 0 .\end{aligned}\tag{42}$$

These equations govern the behavior of electromagnetic fields in general relativity.

4 Curvature

We have been loosely throwing around the idea of “curvature” without giving it a careful definition. The first step toward a better understanding begins with the notion of a **manifold**. Basically, a manifold is “a possibly curved space which, in small enough regions (infinitesimal, really), looks like flat space.” You can think of the obvious example: the Earth looks flat because we only see a tiny part of it, even though it’s round. A crucial feature of manifolds is that they have the same dimensionality everywhere; if you glue the end of a string to a plane, the result is not a manifold since it is partly one-dimensional and partly two-dimensional.

The most famous examples of manifolds are n -dimensional flat space \mathbf{R}^n (“ \mathbf{R} ” as in real, as in real numbers), and the n -dimensional sphere S^n . So, \mathbf{R}^1 is the real line, \mathbf{R}^2 is the plane, and so on. Meanwhile S^1 is a circle, S^2 is a sphere, etc. For future reference, the most popular coordinates on S^2 are the usual θ and ϕ angles. In these coordinates, the metric on S^2 (with radius $r = 1$) is

$$ds^2 = d\theta^2 + \sin^2 \theta d\phi^2 .\tag{43}$$

The fact that manifolds may be curved makes life interesting, as you may imagine. However, most of the difficulties encountered in curved spaces are also encountered in flat space if you use non-Cartesian coordinates. The thing about curved space is, you can never use Cartesian coordinates, because they only describe flat spaces. So the machinery we developed for non-Cartesian coordinates will be crucial; in fact, we’ve done most of the work already.

It should come as no surprise that information about the curvature of a manifold is contained in the metric; the question is, how to extract it? You can’t get it easily from the $\Gamma_{\mu\nu}^\rho$, for instance, since they can be zero or nonzero depending on the coordinate system (as we saw for flat space). For reasons we won’t go into, the information about curvature is contained in a four-component tensor known as the **Riemann curvature tensor**. This supremely important object is given in terms of the Christoffel symbols by the formula

$$R^\sigma{}_{\mu\alpha\beta} \equiv \partial_\alpha \Gamma_{\mu\beta}^\sigma - \partial_\beta \Gamma_{\mu\alpha}^\sigma + \Gamma_{\alpha\lambda}^\sigma \Gamma_{\mu\beta}^\lambda - \Gamma_{\beta\lambda}^\sigma \Gamma_{\mu\alpha}^\lambda .\tag{44}$$

(The overall sign of this is a matter of convention, so check carefully when you read anybody else's papers. Note also that the Riemann tensor is constructed from non-tensorial elements — partial derivatives and Christoffel symbols — but they are carefully arranged so that the final result transforms as a tensor, as you can check.) This tensor has one nice property that a measure of curvature should have: *all of the components of $R^\sigma{}_{\mu\alpha\beta}$ vanish if and only if the space is flat.* Operationally, “flat” means that there exists a global coordinate system in which the metric components are everywhere constant.

There are two contractions of the Riemann tensor which are extremely useful: the **Ricci tensor** and the **Ricci scalar**. The Ricci tensor is given by

$$R_{\alpha\beta} = R^\lambda{}_{\alpha\lambda\beta} . \quad (45)$$

Although it may seem as if other independent contractions are possible (using other indices), the symmetries of $R^\sigma{}_{\mu\alpha\beta}$ (discussed below) make this the only independent contraction. The trace of the Ricci tensor yields the Ricci scalar:

$$R = R^\lambda{}_\lambda = g^{\mu\nu} R_{\mu\nu} . \quad (46)$$

This is another useful item.

Although the Riemann tensor has many indices, and therefore many components, using it is vastly simplified by the many symmetries it obeys. In fact, only 20 of the $4^4 = 256$ components of $R^\sigma{}_{\mu\alpha\beta}$ are independent. Here is a list of some of the useful properties obeyed by the Riemann tensor, which are most easily expressed in terms of the tensor with all indices lowered, $R_{\mu\nu\rho\sigma} = g_{\mu\lambda}R^\lambda{}_{\nu\rho\sigma}$:

$$\begin{aligned} R_{\mu\nu\rho\sigma} &= -R_{\mu\nu\sigma\rho} = -R_{\nu\mu\rho\sigma} \\ R_{\mu\nu\rho\sigma} &= R_{\rho\sigma\mu\nu} \\ R_{\mu\nu\rho\sigma} + R_{\mu\rho\sigma\nu} + R_{\mu\sigma\nu\rho} &= 0 \quad . \end{aligned} \quad (47)$$

These imply a symmetry of the Ricci tensor,

$$R_{\mu\nu} = R_{\nu\mu} . \quad (48)$$

In addition to these algebraic identities, the Riemann tensor obeys a differential identity:

$$\nabla_{[\lambda} R_{\mu\nu]\rho\sigma} = 0 . \quad (49)$$

This is sometimes known as the **Bianchi identity**. If we define a new tensor, the **Einstein tensor**, by

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} , \quad (50)$$

then the Bianchi identity implies that the divergence of this tensor vanishes identically:

$$\nabla^\mu G_{\mu\nu} = 0 . \quad (51)$$

This is sometimes called the contracted Bianchi identity.

Basically, there are only two things you have to know about curvature: the Riemann tensor, and geodesics. You now know the Riemann tensor – lets move on to geodesics.

Informally, a **geodesic** is “the shortest distance between two points.” More formally, a geodesic is a curve which extremizes the length functional $\int ds$. That is, imagine a path parameterized by λ , i.e. $x^\mu(\lambda)$. The infinitesimal distance along this curve is given by

$$ds = \sqrt{\left| g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} \right|} d\lambda . \quad (52)$$

So the entire length of the curve is just

$$L = \int ds . \quad (53)$$

To find a geodesic of a given geometry, we would do a calculus of variations manipulation of this object to find an extremum of L . Luckily, stronger souls than ourselves have come before and done this for us. The answer is that $x^\mu(\lambda)$ is a geodesic if it satisfies the famous **geodesic equation**:

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma_{\rho\sigma}^\mu \frac{dx^\rho}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0 . \quad (54)$$

In fact this is only true if λ is an **affine parameter**, that is if it is related to the proper time via

$$\lambda = a\tau + b . \quad (55)$$

In practice, the proper time itself is almost always used as the affine parameter (for timelike geodesics, at least). In that case, the tangent vector is the four-velocity $U^\mu = dx^\mu/d\tau$, and the geodesic equation can be written

$$\frac{d}{d\tau} U^\mu + \Gamma_{\rho\sigma}^\mu U^\rho U^\sigma = 0 . \quad (56)$$

The physical reason why geodesics are so important is simply this: *in general relativity, test bodies move along geodesics*. If the bodies are massless, these geodesics will be null ($ds^2 = 0$), and if they are massive the geodesics will be timelike ($ds^2 < 0$). Note that when we were being formal we kept saying “extremum” rather than “minimum” length. That’s because, for massive test particles, the geodesics on which they move are curves of *maximum* proper time. (In the famous “twin paradox”, two twins take two different paths through flat spacetime, one staying at home [thus on a geodesic], and the other traveling off into

space and back. The stay-at-home twin is older when they reunite, since geodesics maximize proper time.)

This is an appropriate place to talk about the philosophy of GR. In pre-GR days, Newtonian physics said “particles move along straight lines, until forces knock them off.” Gravity was one force among many. Now, in GR, gravity is represented by the curvature of space-time, *not* by a force. From the GR point of view, “particles move along geodesics, until forces knock them off.” Gravity doesn’t count as a force. If you consider the motion of particles under the influence of forces other than gravity, then they won’t move along geodesics – you can still use (54) to describe their motions, but you have to add a force term to the right hand side. In that sense, the geodesic equation is something like the curved-space expression for $\mathbf{F} = m\mathbf{a} = 0$.

5 General Relativity

Moving from math to physics involves the introduction of dynamical equations which relate matter and energy to the curvature of spacetime. In GR, the “equation of motion” for the metric is the famous **Einstein equation**:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi GT_{\mu\nu} . \quad (57)$$

Notice that the left-hand side is the Einstein tensor $G_{\mu\nu}$ from (50). G is Newton’s constant of gravitation (*not* the trace of $G_{\mu\nu}$). $T_{\mu\nu}$ is a symmetric two-index tensor called the **energy-momentum tensor**, or sometimes the stress-energy tensor. It encompasses all we need to know about the energy and momentum of matter fields, which act as a source for gravity. Thus, the left hand side of this equation measures the curvature of spacetime, and the right measures the energy and momentum contained in it. Truly glorious.

The components $T_{\mu\nu}$ of the energy-momentum tensor are “the flux of the μ^{th} component of momentum in the ν^{th} direction.” This definition is perhaps not very useful. More concretely, we can consider a popular form of matter in the context of general relativity: a **perfect fluid**, defined to be a fluid which is isotropic in its rest frame. This means that the fluid has no viscosity or heat flow; as a result, it is specified entirely in terms of the rest-frame energy density ρ and rest-frame pressure p (isotropic, and thus equal in all directions). If use U^μ to stand for the four-velocity of a fluid element, the energy-momentum tensor takes the form

$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + pg_{\mu\nu} . \quad (58)$$

If we raise one index and use the normalization $g^{\mu\nu}U_\mu U_\nu = -1$, we get an even more under-

standable version:

$$T_{\mu}^{\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} . \quad (59)$$

If $T_{\mu\nu}$ encapsulates all we need to know about energy and momentum, it should be able to characterize the appropriate conservation laws. In fact these are formulated by saying that the covariant divergence of $T_{\mu\nu}$ vanishes:

$$\nabla^{\mu}T_{\mu\nu} = 0 . \quad (60)$$

Recall that the contracted Bianchi identity (51) guarantees that the divergence of the Einstein tensor vanishes identically. So Einstein's equation (57) guarantees energy-momentum conservation. Of course, this is a local relation; if we (for example) integrate the energy density ρ over a spacelike hypersurface, the corresponding quantity is not constant with time. In GR there is no global notion of energy conservation; (60) expresses local conservation, and the appearance of the covariant derivative allows this equation to account for the transfer of energy back and forth between matter and the gravitational field.

The exotic appearance of Einstein's equation should not obscure the fact that it is a natural extension of Newtonian gravity. To see this, consider Poisson's equation for the Newtonian potential Φ :

$$\nabla^2\Phi = 4\pi G\rho , \quad (61)$$

where ρ is the matter density. On the left hand side of this we see a second-order differential operator acting on the gravitational potential Φ . This is proportional to the density of matter. Now, GR is a fully relativistic theory, so we would expect that the matter density should be replaced by the full energy-momentum tensor $T_{\mu\nu}$. To correspond to (61), this should be proportional to a 2-index tensor which is a second-order differential operator acting on the gravitational field, i.e. the metric. If you think about the definition of $G_{\mu\nu}$ in terms of $g_{\mu\nu}$, this is exactly what the Einstein tensor is. In fact, $G_{\mu\nu}$ is the only two-index tensor, second order in derivatives of the metric, for which the divergence vanishes.

So the GR equation is of the same essential form as the Newtonian one. We should ask for something more, however: namely, that Newtonian gravity is recovered in the appropriate limit, where the particles are moving slowly (with respect to the speed of light), the gravitational field is weak (can be considered a perturbation of flat space), and the field is also static (unchanging with time). We consider a metric which is almost Minkowski, but with a specific kind of small perturbation:

$$ds^2 = -(1 + 2\Phi)dt^2 + (1 - 2\Phi)d\vec{x}^2 , \quad (62)$$

where Φ is a function of the spatial coordinates x^i . If we plug this into the geodesic equation and solve for the conventional three-velocity (using that the particles are moving slowly), we

obtain

$$\frac{d^2 \vec{x}}{dt^2} = -\nabla \Phi , \quad (63)$$

where ∇ here represents the ordinary spatial divergence (not a covariant derivative). This is just the equation for a particle moving in a Newtonian gravitational potential Φ . Meanwhile, we calculate the 00 component of the left-hand side of Einstein's equation:

$$R_{00} - \frac{1}{2} R g_{00} = 2 \nabla^2 \Phi . \quad (64)$$

The 00 component of the right-hand side (to first order in the small quantities Φ and ρ) is just

$$8\pi G T_{00} = 8\pi G \rho . \quad (65)$$

So the 00 component of Einstein's equation applied to the metric (62) yields

$$\nabla^2 \Phi = 4\pi G \rho , \quad (66)$$

which is precisely the Poisson equation (61). Thus, in this limit GR does reduce to Newtonian gravity.

Although the full nonlinear Einstein equation (57) looks simple, in applications it is not. If you recall the definition of the Riemann tensor in terms of the Christoffel symbols, and the definition of those in terms of the metric, you realize that Einstein's equation for the metric are complicated indeed! It is also highly nonlinear, and correspondingly very difficult to solve. If we take the trace of (57), we obtain

$$-R = 8\pi G T . \quad (67)$$

Plugging this into (57), we can rewrite Einstein's equations as

$$R_{\mu\nu} = 8\pi G \left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right) . \quad (68)$$

This form is useful when we consider the case when we are in the vacuum – no energy or momentum. In this case $T_{\mu\nu} = 0$ and (68) becomes Einstein's equation in vacuum:

$$R_{\mu\nu} = 0 . \quad (69)$$

This is somewhat easier to solve than the full equation.

One final word on Einstein's equation: it may be derived from a very simple Lagrangian, $\mathcal{L} = \sqrt{-g} R$ (plus appropriate terms for the matter fields). In other words, the action for GR is simply

$$S = \int d^4 x \sqrt{-g} R , \quad (70)$$

an Einstein's equation comes from looking for extrema of this action with respect to variations of the metric $g_{\mu\nu}$. What could be more elegant?

6 Schwarzschild solution

In order to solve Einstein's equation we usually need to make some simplifying assumptions. For example, in many physical situations, we have spherical symmetry. If we want to solve for a metric $g_{\mu\nu}$, this fact is very helpful, because the most general spherically symmetric metric may be written (in spherical coordinates) as

$$ds^2 = -A(r, t)dt^2 + B(r, t)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2), \quad (71)$$

where A and B are positive functions of (r, t) , and you will recognize the metric on the sphere from (43). If we plug this into Einstein's equation, we will get a solution for a spherically symmetric matter distribution. To be even more restrictive, let's consider the equation in vacuum, (69). Then there is a unique solution:

$$ds^2 = -\left(1 - \frac{2Gm}{r}\right) dt^2 + \left(1 - \frac{2Gm}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \quad (72)$$

This is the celebrated **Schwarzschild metric** solution to Einstein's equations. The parameter m , of course, measures the amount of mass inside the radius r under consideration. A remarkable fact is that *the Schwarzschild metric is the unique solution to Einstein's equation in vacuum with a spherically symmetric matter distribution*. This fact, known as **Birkhoff's theorem**, means that the matter can oscillate wildly, as long as it remains spherically symmetric, and the gravitational field outside will remain unchanged.

Philosophy point: the metric components in (72) blow up at $r = 0$ and $r = 2Gm$. Officially, any point at which the metric components become infinite, or exhibit some other pathological behavior, is known as a **singularity**. These beasts come in two types: "coordinate" singularities and "true" singularities. A coordinate singularity is simply a result of choosing bad coordinates; if we change coordinates we can remove the singularity. A true singularity is an actual pathology of the geometry, a point at which the manifold is ill-defined. In the Schwarzschild geometry, the point $r = 0$ is a real singularity, an unavoidable blowing-up. However, the point $r = 2Gm$ is merely a coordinate singularity. We can demonstrate this by making a transformation to what are known as **Kruskal coordinates**, defined by

$$\begin{aligned} u &= \left(\frac{r}{2Gm} - 1\right)^{1/2} e^{r/4Gm} \cosh(t/4Gm) \\ v &= \left(\frac{r}{2Gm} - 1\right)^{1/2} e^{r/4Gm} \sinh(t/4Gm). \end{aligned} \quad (73)$$

In these coordinates, the metric (72) takes the form

$$ds^2 = \frac{32(Gm)^3}{r} e^{-r/2Gm} (-dv^2 + du^2) + r^2(d\theta^2 + \sin^2\theta d\phi^2), \quad (74)$$

where r is considered to be an implicit function of u and v defined by

$$u^2 - v^2 = e^{r/2Gm} \left(\frac{r}{2Gm} - 1 \right) . \quad (75)$$

If we look at (74), we see that nothing blows up at $r = 2Gm$. The mere fact that we could choose coordinates in which this happens assures us that $r = 2Gm$ is a mere coordinate singularity.

The useful thing about the Schwarzschild solution is that it describes both mundane things like the solar system, and more exotic objects like black holes. To get a feel for it, let's look at how particles move in a Schwarzschild geometry. It turns out that we can cast the problem of a particle moving in the plane $\theta = \pi/2$ as a one-dimensional problem for the radial coordinate $r = r(\tau)$. In other words, the distance of a particle from the point $r = 0$ is a solution to the equation

$$\frac{1}{2} \left(\frac{dr}{d\tau} \right)^2 + V(r) = \frac{1}{2} E^2 . \quad (76)$$

This is just the equation of motion for a particle of unit mass and energy E in a one-dimensional potential $V(r)$. This potential, for the Schwarzschild geometry, is given by

$$V(r) = \frac{1}{2} \epsilon - \epsilon \frac{Gm}{r} + \frac{L^2}{2r^2} - \frac{GmL^2}{r^3} . \quad (77)$$

Here, L represents the angular momentum (per unit mass) of the particle, and ϵ is a constant equal to 0 for massless particles and +1 for massive particles. (Note that the proper time τ is zero for massless particles, so we use some other parameter λ in (76), but the equation itself looks the same). So, to find the orbits of particles in a Schwarzschild metric, just solve the motion of a particle in the potential given by (77). Note that the first term in (77) is a constant, the second term is exactly what we expect from Newtonian gravity, and the third term is just the contribution of the particle's angular momentum, which is also present in the Newtonian theory. Only the last term in (77) is a new addition from GR.

There are two important effects of this extra term. First, it acts as a small perturbation on any orbit – this is what leads to the precession of Mercury, for instance. Second, for r very small, the GR potential goes to $-\infty$; this means that a particle that approaches too close to $r = 0$ will fall into the center and never escape! Even though this is in the context of unaccelerated test particles, a similar statement holds true for particles with the ability to accelerate themselves all they like – see below. However, not to worry; for a star such as the Sun, for which the Schwarzschild metric only describes points outside the surface, you would run into the star long before you approached the point where you could not escape.

Nevertheless, we all know of the existence of more exotic objects: **black holes**. A black hole is a body in which all of the mass has collapsed gravitationally past the point of possible escape. This point of no return, given by the surface $r = 2Gm$, is known as

the **event horizon**, and can be thought of as the “surface” of a black hole. Although it is impossible to go into much detail about the host of interesting properties of the event horizon, the basics are not difficult to grasp. From the point of view of an outside observer, a clock falling into a black hole will appear to move more and more slowly as it approaches the event horizon. In fact, the external observer will never see a test particle cross the surface $r = 2Gm$; they will just see the particle get closer and closer, and move more and more slowly.

Contrast this to what you would experience as a test observer actually thrown into a black hole. To you, time always seems to move at the same rate; since you and your wristwatch are in the same inertial frame, you never “feel time moving more slowly.” Therefore, rather than taking an infinite amount of time to reach the event horizon, you zoom right past – doesn’t take very long at all, actually. You then proceed directly to fall to $r = 0$, also in a very short time. *Once you pass $r = 2Gm$, you cannot help but hit $r = 0$* ; it is as inevitable as moving forward in time. The literal truth of this statement can be seen by looking at the metric (72) and noticing that r becomes a *timelike* coordinate for $r < 2Gm$; therefore your voyage to the center of the black hole is literally moving forward in time! What’s worse, we noted above that a geodesic (unaccelerated motion) maximized the proper time – this means that the more you struggle, the sooner you will get there. (Of course, you won’t struggle, because you would have been ripped to shreds by tidal forces. The grisly death of an astrophysicist who enters a black hole is detailed in Misner, Thorne, and Wheeler, pp. 860-862.)

The spacetime diagram of a black hole in Kruskal coordinates (74) is shown in Figure 2. Shown is a slice through the entire spacetime, corresponding to angular coordinates $\theta = \pi/2$ and $\phi = 0$. There are two asymptotic regions, one at $u \rightarrow +\infty$ and the other at $u \rightarrow -\infty$; in both regions the metric looks approximately flat. The event horizon is the surface $r = 2Gm$, or equivalently $u = \pm v$. In this diagram all light cones are at $\pm 45^\circ$. Inside the event horizon, where $r < 2Gm$, all timelike trajectories lead inevitably to the singularity at $r = 0$. It should be stressed that this diagram represents the “maximally extended” Schwarzschild solution — a complete solution to Einstein’s equation in vacuum, but not an especially physically realistic one. In a realistic black hole, formed for instance from the collapse of a massive star, the vacuum equations do not tell the whole story, and there will not be two distinct asymptotic regions, only the one in which the star originally was located. (For that matter, timelike trajectories cannot travel between the two regions, so we could never tell whether another such region did exist.)

In the collapse to a black hole, all the information about the detailed nature of the collapsing object is lost: what it was made of, its shape, etc. The only information which is not wiped out is the amount of mass, angular momentum, and electric charge in the hole. This fact, the **no-hair theorem**, implies that the most general black-hole metric will be a function of these three numbers only. However, real-world black holes will prob-

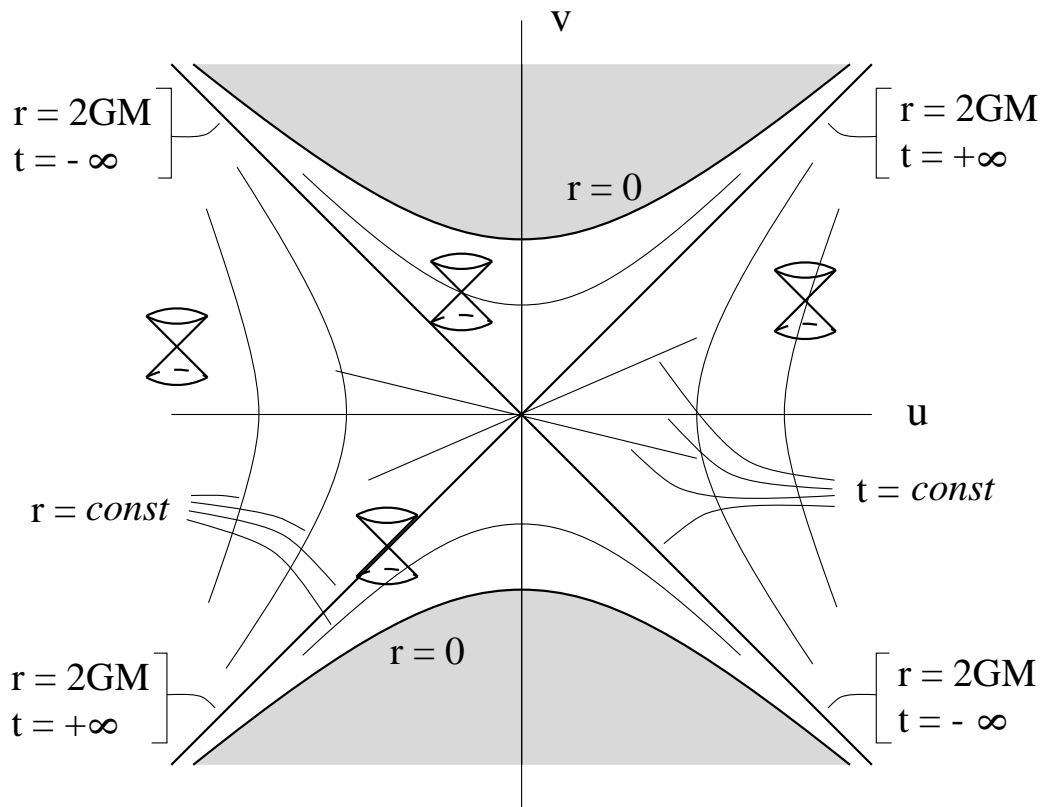


Figure 2: The Kruskal diagram — the Schwarzschild solution in Kruskal coordinates (74), where all light cones are at $\pm 45^\circ$. The surface $r = 2Gm$ is the event horizon; inside the event horizon, all timelike paths hit the singularity at $r = 0$. The right- and left-hand side of the diagram represent distinct asymptotically flat regions of spacetime.

ably be electrically neutral, so we will not present the metric for a charged black hole (the **Reissner-Nordstrom metric**). Of considerable astrophysical interest are spinning black holes, described by the **Kerr metric**:

$$ds^2 = - \left[\frac{\Delta - \omega^2 \sin^2 \theta}{\Sigma} \right] dt^2 - \left[\frac{4\omega m G r \sin^2 \theta}{\Sigma} \right] dt d\phi + \frac{\Sigma}{\Delta} dr^2 + \Sigma d\theta^2 + \left[\frac{(r^2 + \omega^2)^2 - \Delta \omega^2 \sin^2 \theta}{\Sigma} \right] \sin^2 \theta d\phi^2, \quad (78)$$

where

$$\Sigma \equiv r^2 + \omega^2 \cos^2 \theta, \quad \Delta \equiv r^2 + \omega^2 - 2Gmr, \quad (79)$$

and ω is the angular velocity of the body.

Finally, among the many additional possible things to mention, there's the **cosmic censorship conjecture**. Notice how the Schwarzschild singularity at $r = 0$ is hidden, in a sense – you can never get to it without crossing an horizon. It is conjectured that this is *always* true, in any solution to Einstein's equation. However, some numerical work seems to contradict this conjecture, at least in special cases.

7 Cosmology

Just as we were able to make great strides with the Schwarzschild metric on the assumption of spherical symmetry, we can make similar progress in cosmology by assuming that the Universe is homogeneous and isotropic. That is to say, we assume the existence of a “rest frame for the Universe,” which defines a universal time coordinate, and singles out three-dimensional surfaces perpendicular to this time coordinate. (In the real Universe, this rest frame is the one in which galaxies are at rest and the microwave background is isotropic.) “Homogeneous” means that the curvature of any two points at a given time t is the same. “Isotropic” is trickier, but basically means that the universe looks the same in all directions. Thus, the surface of a cylinder is homogeneous (every point is the same) but not isotropic (looking along the long axis of the cylinder is a preferred direction); a cone is isotropic around its vertex, but not homogeneous.

These assumptions narrow down the choice of metrics to precisely three forms, all given by the **Robertson-Walker (RW) metric**:

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right], \quad (80)$$

where the constant k can be -1 , 0 , or $+1$. The function $a(t)$ is known as the **scale factor** and tells us the relative sizes of the spatial surfaces. The above coordinates are called **comoving coordinates**, since a point which is at rest in the preferred frame of the universe

will have $r, \theta, \phi = \text{constant}$. The $k = -1$ case is known as an **open** universe, in which the preferred three-surfaces are “three-hyperboloids” (saddles); $k = 0$ is a **flat** universe, in which the preferred three-surfaces are flat space; and $k = +1$ is a **closed** universe, in which the preferred three-surfaces are three-spheres. *Note that the terms “open,” “closed,” and “flat” refer to the spatial geometry of three-surfaces, not to whether the universe will eventually recollapse.* The volume of a closed universe is finite, while open and flat universes have infinite volume (or at least they can; there are also versions with finite volume, obtained from the infinite ones by performing discrete identifications).

There are other coordinate systems in which (8.1) is sometimes written. In particular, if we set $r = (\sin \psi, \psi, \sinh \psi)$ for $k = (+1, 0, -1)$ respectively, we obtain

$$ds^2 = -dt^2 + a^2(t) \left\{ \begin{array}{l} d\psi^2 + \sin^2 \psi (d\theta^2 + \sin^2 \theta d\phi^2) \\ d\psi^2 + \psi^2 (d\theta^2 + \sin^2 \theta d\phi^2) \\ d\psi^2 + \sinh^2 \psi (d\theta^2 + \sin^2 \theta d\phi^2) \end{array} \right\} \begin{array}{l} (k = +1) \\ (k = 0) \\ (k = -1) \end{array} \quad (81)$$

Further, the flat ($k = 0$) universe also may be written in almost-Cartesian coordinates:

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t)(dx^2 + dy^2 + dz^2) \\ &= -a^2(\eta)(-d\eta^2 + dx^2 + dy^2 + dz^2). \end{aligned} \quad (82)$$

In this last expression, η is known as the **conformal time** and is defined by

$$\eta \equiv \int \frac{dt}{a(t)}. \quad (83)$$

The coordinates (η, x, y, z) are often called “conformal coordinates.”

Since the RW metric is the only possible homogeneous and isotropic metric, all we have to do is solve for the scale factor $a(t)$ by using Einstein’s equation. If we use the vacuum equation (69), however, we find that the only solution is just Minkowski space. Therefore we have to introduce some energy and momentum to find anything interesting. Of course we shall choose a perfect fluid specified by energy density ρ and pressure p . In this case, Einstein’s equation becomes two differential equations for $a(t)$, known as the **Friedmann equations**:

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho - \frac{k}{a^2} \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}(\rho + 3p). \end{aligned} \quad (84)$$

Since the Friedmann equations govern the evolution of RW metrics, one often speaks of Friedman-Robertson-Walker (FRW) cosmology.

The expansion rate of the universe is measured by the **Hubble parameter**:

$$H \equiv \frac{\dot{a}}{a}, \quad (85)$$

and the change of this quantity with time is parameterized by the **deceleration parameter**:

$$q \equiv -\frac{\ddot{a}a}{\dot{a}^2} = -\left(1 + \frac{\dot{H}}{H^2}\right). \quad (86)$$

The Friedmann equations can be solved once we choose an equation of state, but the solutions can get messy. It is easy, however, to write down the solutions for the $k = 0$ universes. If the equation of state is $p = 0$, the universe is **matter dominated**, and

$$a(t) \propto t^{2/3}. \quad (87)$$

In a matter dominated universe, the energy density decreases as the volume increases, so

$$\rho_{matter} \propto a^{-3}. \quad (88)$$

If $p = \frac{1}{3}\rho$, the universe is **radiation dominated**, and

$$a(t) \propto t^{1/2}. \quad (89)$$

In a radiation dominated universe, the number of photons decreases as the volume increases, and the energy of each photon redshifts and amount proportional to $a(t)$, so

$$\rho_{rad} \propto a^{-4}. \quad (90)$$

If $p = -\rho$, the universe is **vacuum dominated**, and

$$a(t) \propto e^{Ht}. \quad (91)$$

The vacuum dominated universe is also known as **de Sitter space**. In de Sitter space, the energy density is *constant*, as is the Hubble parameter, and they are related by

$$H = \sqrt{\frac{8\pi G\rho_{vac}}{3}} = constant. \quad (92)$$

Note that as $a \rightarrow 0$, ρ_{rad} grows the fastest; therefore, if we go back far enough in the history of the universe we should come to a radiation dominated phase. Similarly, ρ_{vac} stays constant as the universe expands; therefore, if ρ_{vac} is not zero, and the universe lasts long enough, we will eventually reach a vacuum-dominated phase.

Given that our Universe is presently expanding, we may ask whether it will continue to do so forever, or eventually begin to recontract. For energy sources with $p/\rho \geq 0$ (including both matter and radiation dominated universes), closed ($k = +1$) universes will eventually recontract, while open and flat universes will expand forever. When we let $p/\rho < 0$ things get messier; just keep in mind that spatially closed/open does not necessarily correspond to temporally finite/infinite.

The question of whether the Universe is open or closed can be answered observationally. In a flat universe, the density is equal to the **critical density**, given by

$$\rho_{crit} = \frac{3H^2}{8\pi G} . \quad (93)$$

Note that this changes with time; in the present Universe it's about 5×10^{-30} grams per cubic centimeter. The universe will be open if the density is less than this critical value, closed if it is greater. Therefore, it is useful to define a **density parameter** via

$$\Omega \equiv \frac{\rho}{\rho_{crit}} = \frac{8\pi G\rho}{3H^2} = 1 + \frac{k}{\dot{a}^2} , \quad (94)$$

a quantity which will generally change with time unless it equals unity. An open universe has $\Omega < 1$, a closed universe has $\Omega > 1$.

We mentioned in passing the **redshift** of photons in an expanding universe. In terms of the wavelength λ_1 of a photon emitted at time t_1 , the wavelength λ_0 observed at a time t_0 is given by

$$\frac{\lambda_0}{\lambda_1} = \frac{a(t_0)}{a(t_1)} . \quad (95)$$

We therefore define the redshift z to be the fractional increase in wavelength

$$z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{a(t_0)}{a(t_1)} - 1 . \quad (96)$$

Keep in mind that this only measures the net expansion of the universe between times t_1 and t_0 , not the relative speed of the emitting and observing objects, especially since the latter is not well-defined in GR. Nevertheless, it is common to speak as if the redshift is due to a Doppler shift induced by a relative velocity between the bodies; although nonsensical from a strict standpoint, it is an acceptable bit of sloppiness for small values of z . Then the Hubble constant relates the redshift to the distance s (measured along a spacelike hypersurface) between the observer and emitter:

$$z = H(t_0)s . \quad (97)$$

This, of course, is the linear relationship discovered by Hubble.