General Relativity

In 1687, Newton put forward his law of universal gravitation, according to which every massive object in the universe exerts a force $\left|\vec{F}\right| = Gm_1m_2/r^2$ on every *other* massive object in the universe. Newton showed that this single force could explain phenomena ranging from apples falling from trees, to planets rotating about the Sun.

In 1915, Einstein published his General Theory of Relativity (often just called "General Relativity" or "GR"), a completely different understanding in which gravity is no longer viewed as a force. Einstein said that massive objects like the Earth and Sun alter the geometry of space and time around them. The falling of apples and the orbits of planets result from objects moving in that altered geometry.

We have divided our introduction to general relativity into two sections. You can stop after the first section for a high-level overview without too much math, or do both sections if you want more mathematical detail.

I. The first section is a conceptual overview of the theory.

- We begin with some motivation. Why would Einstein choose to reinterpret gravity geometrically?
- We then look at the mathematical idea of curved spacetime, and the GR relationship between matter and spacetime.
- We then discuss some empirical evidence for the theory. In most cases, GR makes virtually the same predictions as Newtonian gravity. But in some cases the two theories diverge, and such cases led to observational validation of Einstein's version of gravity.

II. The second section steps into the math underlying Einstein's geometric approach.

- We begin with the idea of a "geodesic" which takes the Euclidean idea of a straight line and generalizes that idea to curved spaces.
- We then introduce the central idea of the section, a "metric": a frame-independent idea of distance in a curved spacetime.
- As a real-world example of a curved spacetime, we discuss the metric of an expanding universe.

These sections assume that you are familiar with special relativity. Before reading this introduction to GR, you should study Chapter 1 and Section 2.1 of the book. (The rest of Chapter 2 is not assumed.)

General Relativity I: Overview

This "Overview" section introduces the conceptual underpinnings of general relativity.

Discovery Exercise: The Electric Force and the Gravitational Force

The questions in this Discovery Exercise are based on purely classical mechanics: no relativity or quantum mechanical ideas are required.

You are attempting to predict the motion of a proton and an electron. The two particles are far apart, and exert no significant force on each other. Note that we're not asking you to actually do any calculations or write any formulas below; we are asking questions about how you *would* do some calculations. If the whole thing takes you more than five minutes, you're probably overthinking it.

In your first experiment, the two particles are immersed in a constant electric field.

1. First you calculate the force that the electric field exerts on the proton, and also the force it exerts on the electron. What *property* of each particle do you need to know, in order to calculate these forces?
2. From those forces, you calculate the resulting acceleration of each particle. What property of the proton and electron do you need to know for this step?
3. Briefly explain why the electron will accelerate much more than the proton, based on both steps above.

In your second experiment, the two particles are immersed in a constant *gravitational* field.

4. Once again, you begin by calculating the two forces exerted by this field. What property of the two particles do you need to know in order to calculate these forces?
5. From those forces, you calculate the resulting accelerations. What property of the two particles do you need to know for this step?
6. This exercise has been designed to point out a key difference between the classical electric force and the classical gravitational force. What point do you think we're driving at?

Explanation: General Relativity

We often say that modern physics recognizes four forces: the strong (or nuclear) force, the weak force, the electromagnetic force, and gravity. But Einstein argued that gravity is not a force at all, in the traditional sense of the word. Instead, he described gravity as an interplay between matter and the geometry of spacetime.

We begin with the question: why reframe a classic Newtonian concept in this radical way?

Motivation: Inertial Mass and Gravitational Mass

In the Discovery Exercise above, you performed a classical analysis of the behavior of two different particles under the influence of first an electric field, and then a gravitational field.

In the Active Reading Exercise below you will return to that scenario, this time by drawing spacetime diagrams. As in the main text, don't read past an Active Reading Exercise until you have tried all the parts of it yourself.

Active Reading Exercise: Two Particles with Different Forces

Nothing in this exercise requires special relativity per se. However, you do need to be familiar with "spacetime diagrams," as described in Section 2.1.

A proton moves along the x-axis. At time $t = 0$ the proton is at position $x = 0$ with a positive velocity. We aren't going to put a number to that velocity, so your goal in this exercise is to get the overall shapes right, not the numbers.

1. Draw a spacetime diagram that represents the proton's journey if no forces act upon it.
2. Starting over on a blank graph, draw a different spacetime diagram that represents the proton's journey if it is immersed in a constant electric field pointing in the negative x-direction. (Perhaps the proton is inside a capacitor.)
3. On a third graph, draw a third spacetime diagram that represents the proton's journey if it is immersed in a constant *gravitational* field pointing in the negative x-direction. (Perhaps the proton is on Earth, and the x direction is "up.")
4. Now replace the proton with an electron: identical-but-opposite charge, much smaller mass, same initial position and velocity. On each of the three spacetime diagrams you created for the proton, add a second path for the electron. Our particular interest here is whether, and how, the electron's path differs from the proton's path in each scenario.

You should create three spacetime diagrams, with two curves each, before you continue reading!

Our spacetime diagrams are shown in Figure 1.

- Without forces, both particles move along a straight line in spacetime.
- In an electric field, both particles accelerate. The electron accelerates in the opposite direction from the proton, because it has the opposite charge. More interestingly (for our purpose here), the electron accelerates *more* than the proton. That's because the two particles feel the same magnitude of force, so the particle with the smaller mass experiences greater acceleration.
- In a gravitational field, both particles move identically. The heavier particle feels a much *greater* force, causing the two particles to experience the *same* acceleration. This is how Newton's description of gravity explains Galileo's famous observation that a heavy ball and a light ball fall to the Earth at the same rate.

The point of this Active Reading Exercise, and of the Discovery Exercise that began this section, is the peculiar double role of mass in Newtonian physics. Mass determines how much

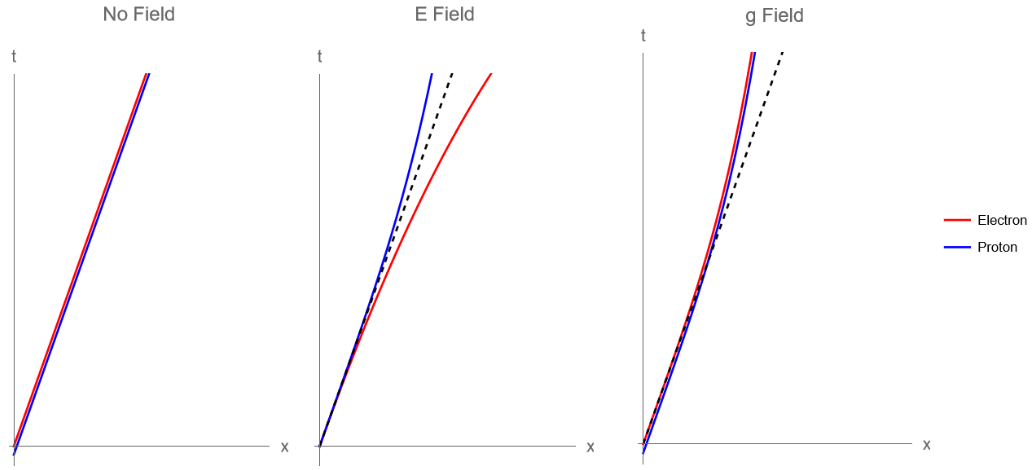


Figure 1 **Worldlines for an electron and a photon in free space (no acceleration), in an electric field (different accelerations), and in a gravitational field (identical accelerations)**

a given object accelerates in response to any given force. But for gravity in particular, mass also determines *how much force* an object exerts and feels.

Physicists sometimes call the property that appears in Newton's second law "inertial mass," and the property in Newton's law of gravity "gravitational mass." So we can restate the previous paragraph more concisely by saying that, empirically, inertial mass equals gravitational mass for all particles.[1] The equivalence of inertial and gravitational mass has no explanation in Newtonian physics: it just looks like a coincidence. But this unexplained coincidence has important consequences.

- In our first scenario above (no forces), the two particles move together because they are both following Newton's *first* law. Starting with the same position and velocity, they move identically from there.
- The second scenario shows that, in the presence of forces, particles generally move differently. That's because the particles can have different properties that cause them to *feel* different amounts of force, and also to *react differently* to forces.
- The third scenario shows that in the particular case of gravity, those two effects—how much force the particles feel, and how they react to those forces—exactly cancel, so all particles accelerate identically in a gravitational field. The classical explanation for this behavior is that inertial and gravitational masses are the same.

The Equivalence Principle

The equivalence of inertial and gravitational mass can be illustrated by a famous thought experiment.

In Section 1.1 we asked you to "imagine yourself on a space ship" which has "no windows and no communication of any kind with the outside world." We saw that you can formulate

1 Technically we can only say inertial mass is *proportional* to gravitational mass, since any constant of proportionality between the two could be absorbed into the constant G.

experiments to discover whether the ship is accelerating or not, but you cannot design any experiment that will discern whether the ship is moving or standing still. Galileo used this thought experiment (with an ocean ship rather than a space ship) to demonstrate that the laws of physics are the same in any inertial reference frame.

Now we return to our space ship to illustrate the point we made above with the three scenarios, and to explain how the equivalence of inertial and gravitational mass led Einstein to GR.

Imagine that you are in this sealed ship, and you find yourself floating weightlessly. You hover in the middle of the room, not moving any closer to the floor, ceiling, or walls. You throw a ball and see it move in a straight line at constant speed, following Newton's first law. You might naturally conclude that your ship is floating freely in deep space.

But your friend Al, floating along next to you, proposes an alternative explanation: the ship is near the Earth, falling. Al says that you are in fact accelerating downward at 9.8 m/s^2. The reason you don't fall towards the floor is that the floor of the ship is also in free fall, accelerating downward at the same rate. So are the ceiling, the walls, the ball, and everything else you can see inside the ship.

You and Al set out to find an experiment that can help you decide if you are actually floating in deep space, or falling towards Earth. But no such experiment works because in both cases you see everything inside the ship seemingly obeying Newton's first law, not accelerating relative to you and the ship (Figure 2).

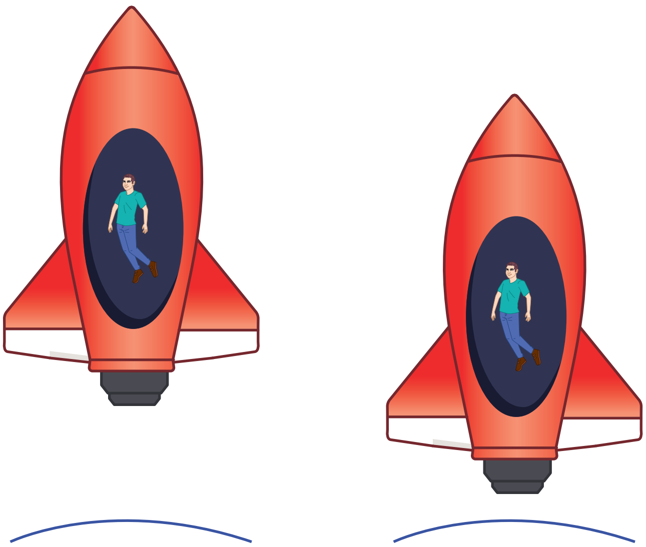


Figure 2 **In a ship in free fall you appear to hover, because you and the ship accelerate downward at the same rate (neglecting air resistance). This is why astronauts in the space station appear weightless. Earth's gravity is fairly strong there and causes them to orbit, but they go around in the same orbit as the walls of the space station, so they hover *relative* to the station.**

(See if you can think of an exception, an experiment that could distinguish floating in space from being in free fall. If you don't use GR there is at least one such experiment, and we'll discuss it later on.)

The fact that these two situations (floating in deep space and freely falling) are indistinguishable is called the "equivalence principle." Make sure you see how this principle *results from* the equivalence of inertial and gravitational mass. If you and the ship were charged objects in an electric field, your situation would certainly not be indistinguishable from floating in space.

Gravitational free fall acts like a zero-acceleration system because you, the ship, and all other objects experience exactly the same acceleration at all times.

Einstein took this thought experiment to its logical conclusion: if these two scenarios are indistinguishable experimentally, then maybe they are physically equivalent. An object in space, with no external forces acting on it, follows Newton's first law and does not accelerate. Maybe an object in free fall, with no external forces other than gravity, is also following Newton's first law and not accelerating!

That explanation sounds nice in a way. Inside the falling ship, it certainly looks like everything you can see is obeying Newton's first law. But to an observer on the ground, it seems clear that the ship and everything in it are accelerating as they fall. To describe this motion as inertial, Einstein had to change how we think about acceleration.

Straight Lines Through Curved Space

In the simplest of the three scenarios in the Active Reading Exercise above, a proton and an electron experienced no forces of any kind. Newton's *first* law says that such particles move in straight lines at constant velocity. You represented this motion by drawing straight lines on your first spacetime diagram (the left side of Figure 1). In the absence of external forces, any object will trace out a straight-line path in a spacetime diagram. Any two such particles with the same initial position and velocity will trace out *identical* straight-line spacetime trajectories. There's nothing unusual or coincidental about that.

GR treats gravity as something other than a force. So the proton and electron in the third scenario in the Active Reading Exercise above traced out identical paths (the right side of Figure 1) because—since gravity doesn't count as a force—there was no force acting on them at all! They followed Newton's first law, and we don't need to invoke any coincidence between inertial and gravitational mass to explain why they stayed together.

The problem with that argument is that their path was curved, not straight, right? But Einstein argues that their path was *effectively* straight, and appeared curved in our drawing because the geometry of spacetime was curved. That claim probably sounds nonsensical, even by the standards of modern physics. To explore what it means, we need to talk a bit about geometry.

Sometime around 300 B.C., Euclid wrote down a systematic set of geometric laws. He rigorously proved many now-familiar rules such as "parallel lines never meet" and "the angles of a triangle add up to 180°." But all proofs have to begin somewhere: Euclid began his system with five simple postulates, and proved everything else from there. The postulates themselves could not be proven, but he attempted to make them so simple and obvious that no reasonable observer could doubt them.

And for thousands of years, pretty much no one did. But in the 19$^{\text{th}}$ century several mathematicians independently discovered that you can start with different sets of postulates and derive rules like "any pair of parallel lines crosses exactly twice," or "the angles of a triangle always add up to less than 180°." These different systems of geometry, starting from different postulates, obviously contradict Euclid's system (and they also contradict each other). However—this is the key mathematical point—you can design such a system that never leads to any contradiction within *itself*. No such self-consistent system can be logically ruled out.

To take a common example, consider the surface of a smooth sphere. This represents a two-dimensional geometry: you're not allowed to move above or below the surface, but you can draw any figure you want on the surface.

This surface-of-a-sphere geometry has a number of odd properties. For one thing, if you start at any point and walk far enough in any direction, you'll wrap around the space and end up back where you started. So unlike a flat plane, the surface of a sphere is a finite space.

Now imagine that you and a friend both start on the equator, a few feet away from each other, and walk North, heading towards the North pole. For a long time you walk side by side, but you gradually get closer to each other until you finally collide at the North pole (Figure 3). This example illustrates the idea that in non-Euclidean geometry, parallel lines can meet. (See if you can convince yourself that Figure 3 also illustrates how, in this geometry, the angles of a triangle add up to more than 180°.)

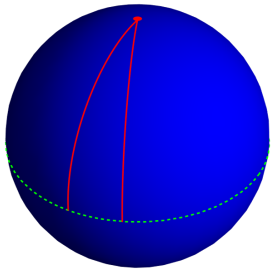


Figure 3 **On the surface of a sphere, two paths that start out parallel and continue without turning left or right will eventually meet.**

You might object that neither of you actually walked a "line"; you moved on curved paths, following the curvature of the sphere. From your point of view, however, you never turned left or right. In a 2D space, there's no other way you *can* turn, so your paths were straight. (We will revisit this issue with a more careful argument in the next section.)

You might then object that the two lines were not "parallel." But remember that you and your friend started at different points, and set off in the same direction—you both headed North—so the lines are parallel.

We're presenting a visual model here, not a rigorous geometric system. In the next section we will explore the math behind the model in a little more depth. But this sphere can be used as a model of a rigorous system in which certain postulates lead to the proofs of a self-consistent set of theorems. Because we are describing a two-dimensional geometry, we can use a three-dimensional picture to visualize why (for instance) parallel lines meet, and the angles of a triangle add up to more than 180°. A two-dimensional being would find it impossible to picture such results, just as we ourselves cannot picture curvature in our own three-dimensional universe.

Gravity as Curved Space

We've talked about the equivalence principle, and how Einstein wanted a new explanation of gravity that made this principle less of a coincidence. And we've talked about the idea of curved space. Now we're ready to connect the two ideas. Einstein proposed that any massive object warps the spacetime around itself; it is the geometry of spacetime, not a gravitational force as such, that changes the motion of nearby objects.

The physicist John Wheeler summarized all of general relativity in one sentence.

> "Space-time tells matter how to move; matter tells space-time how to curve."

As an example, consider how a 19^{th} century physicist and Einstein offer different explanations for the Earth's rotation around the Sun.

- 19^{th} century physicist: The Sun creates a gravitational field around itself. This field exerts a force on any massive object, causing the Earth to accelerate continually toward the Sun. Such a central force can cause elliptical motion.

- Einstein: The Sun's mass warps the spacetime around it into a non-Euclidean geometry. The Earth follows a straight line through that geometry, which appears as an elliptical orbit.

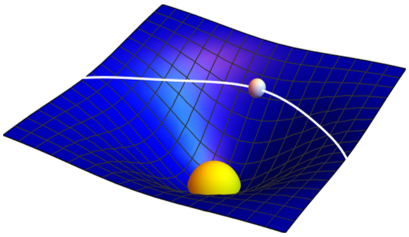


Figure 4 A misleading-but-still-useful image of an orbit in a warped geometry.

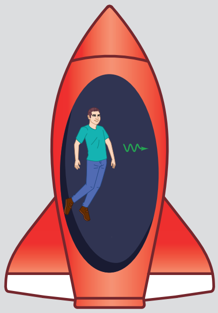
This situation is commonly depicted by drawing space as a two-dimensional deformable surface (like a blanket or a trampoline), with a large depression caused by the Sun, as in Figure 4. This kind of image has limited explanatory power, and can be misleading if taken too literally. One obvious problem is that this model shows a two-dimensional universe being warped into the third dimension due to some mysterious "downward" force on the Sun. A deeper limitation is that actual GR orbits result from the geometry of *spacetime*, not of space alone.

But with all that kept in mind, an image like Figure 4 can be useful. It shows how the planet follows a straight-line path at all times, and how curved geometry can turn that straight-line path into a periodic orbit.

Evidence for General Relativity

One of Einstein's motivations for developing general relativity was the equivalence principle, the seemingly coincidental fact that objects in a room in free fall in a gravitational field behave exactly like objects in a room with no forces on it. But Einstein was also aware of a seeming exception to this rule.

Active Reading Exercise: A Flashlight on a Falling Ship



You and Al are back in your sealed space ship, and you have come up with a new experiment to try. You fire a single photon across the ship and measure the spot where that photon hits the far wall.

Question: Will the photon hit directly across from where you fired it, or slightly below that spot, or slightly above that spot?

Remember that one of our two possible scenarios was your ship floating in space, with no gravity or acceleration. In that case the answer to our question is obvious: the photon will hit directly across from where you fired it.

But where does the photon hit in the other scenario, with the ship in free fall toward the Earth?

1. Answer this question based on a classical description of gravity. *Hint:* Photons are massless.
2. Now answer the same question based on general relativity.

Here are our answers, illustrated in Figure 5.

1. According to Newtonian physics, the photon should be unaffected by gravity because it has no mass. The ship will fall a little, while the light beam moves in a straight line, so you will see the beam hit the far wall *above* where you fired it.
2. But in Einstein's model, the Earth's gravity bends spacetime around it. The light will trace the same "straight line" through that warped spacetime as all other objects, including the ship itself. Therefore, the light will still appear to "fall" toward the Earth, and will perfectly hit the spot on the wall that you aimed at.

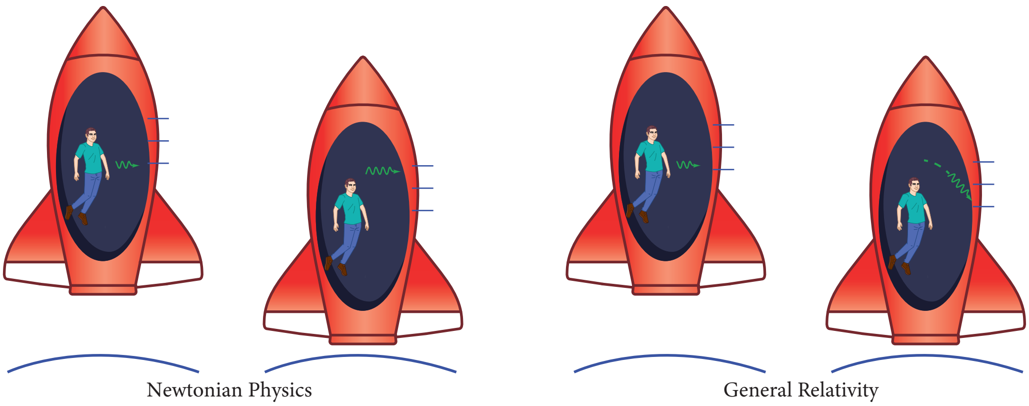


Figure 5 **In Newtonian physics, light (which is massless) is unaffected by gravity. So a light beam in a freely falling room should *appear* to rise. In general relativity, light bends in a gravitational field the same way massive objects do. To an observer in a freely falling room, a light beam thus *appears* straight, striking the opposite wall at the same height as where it was shone from. (We've exaggerated the effect for clarity. Light travels so fast that the ship would only fall a few nanometers in the time the beam crossed the ship.)**

Part of the "moral" of that Active Reading exercise is that Einstein's version of gravity obeys the equivalence principle more strictly than its classical counterpart. In GR, the no-gravity-no-acceleration scenario and the free-fall-near-gravity scenario are indistinguishable, even for massless particles such as photons.

But there is another important takeaway: the Newtonian and relativistic models make measurably different predictions of the path of a light beam under the influence of gravity, and that provides an empirical way to distinguish the two models. During a solar eclipse in 1919, astronomers measured the angular positions of stars very near the Sun. They found that the positions of these stars appeared to be shifted, because their light was bent as it passed by the Sun (Figure 6). This non-Newtonian effect perfectly matched the predictions of general relativity. Many people point to these observations of "gravitational lensing" as the moment when GR was universally accepted.

Gravitational lensing was one of three experimental tests of GR (now often referred to as the "classical tests") proposed by Einstein in 1916. The others don't follow as clearly as lensing does from the ideas of relativity as we've presented them, but they come out when you work through the math.

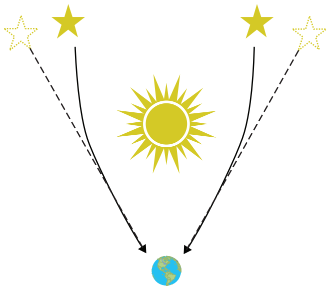


Figure 6 **When light from distant stars gets bent by the Sun's gravity, it appears to us that the starlight is coming from farther away from the Sun than it actually is. This effect can be measured during a solar eclipse, when the sky gets dark enough for us to observe stars near the Sun. During the eclipse, stars near the Sun appear farther away from each other than they do when you observe that same constellation at night.**

- Newton's theories predicted the trajectories of the planets with remarkable precision, but careful observations of Mercury showed a precession that was not commensurate with Newtonian calculations. This problem was recognized in 1859. Einstein showed that his new theory of gravity predicted Mercury's orbit perfectly.
- Just as objects moving relative to us experience a Doppler shift, GR predicts a *gravitational* Doppler shift. As light rises in a gravitational field, it loses energy and thus its wavelength increases. This effect was measured, once again matching Einstein's predictions, in 1954.

This marks the end of our first section (out of two) on general relativity. You now have a fair overview of how GR offers a new formulation of gravity that leads to some but not all of the same predictions as Newton's gravitational force law. You have also seen some of the theoretical and empirical reasons for accepting this new formulation. If you continue to the next section, you'll see some beginnings of the math that is used to describe curved spacetime.

General Relativity II: Metrics

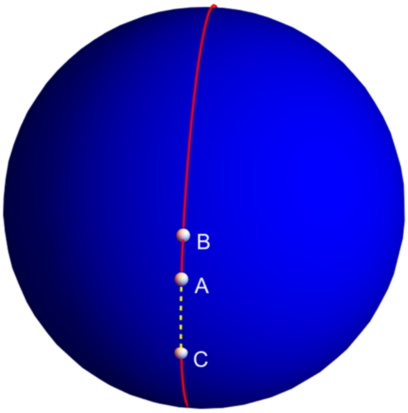


Figure 7 **The red curve is the shortest path between Points A and B, but not the shortest path between A and C. However, we define the entire red curve as a "geodesic," the curved-space generalization of the idea of a straight line.**

In Figure 3 on p. 7, two paths began at different points on the equator of a sphere and converged at the North pole. We asked why these paths qualify as "straight lines," and answered that neither one turns left or right. But how can we precisely define terms like "turn" and "straight" in such a space? One very general answer is to define a straight path as *the shortest distance between two nearby points.*

The word "nearby" in that definition is important. You could keep going past the North pole until you circle around to a point two feet South of where you started (Figure 7). Your path would have taken you almost all the way around the globe: certainly not the shortest distance between your starting and ending points! But we still define that path as a line. If you look at any point on your path and the point you reached 1 mm later, the path you took was the shortest possible path between those points.

A path where each tiny interval is the shortest possible distance between its endpoints is called a "geodesic." In non-Euclidean geometry people don't generally use the phrase "straight line," but stick to the more mathematically precise term "geodesic." In Euclidean geometry, the geodesics are just familiar straight lines.

So what does this have to do with gravity? In general relativity, Newton's first law is recast as:

"In the absence of any non-gravitational forces acting on it, an object will move along a geodesic in spacetime."

Gravity changes the geometry, so those geodesics can now correspond to parabolas, hyperbolas, ellipses, and other more complicated trajectories, as well as straight lines.

Since a geodesic is defined by minimizing distances, we clearly need a mathematical definition of distance. Such a definition, called a "metric," is the property you need to define any given geometry.

Metrics in Space

A "metric" measures the distance between two nearby points.

Let's start with a purely Euclidean geometry: what is the metric for a plane? Trick question: it depends on what variables you use! For the Cartesian x and y, the metric is $ds^2 = dx^2 + dy^2$ (as you would expect). For the polar ρ and ϕ (also sometimes called r and θ), the metric is $ds^2 = d\rho^2 + \rho^2 d\phi^2$. These are not two different geometries; they are two different descriptions of the same geometry. You can start with $x = \rho\cos\phi$ and $y = \rho\sin\phi$ and do a bit of algebra to convert the polar metric to the Cartesian.

To take a less obvious example, let's return to our surface of a sphere. If you use ϕ for longitude and θ for latitude, the metric is $ds^2 = R^2(\sin^2\theta)d\phi^2 + R^2 d\theta^2$. This metric isn't obvious, but it should make sense if you think about the fact that the same change in longitude (going East or West) moves you a greater distance at some values of θ than at others. At the poles ($\theta = 0$ or $\theta = \pi$), changing longitude has no effect; at the equator ($\theta = \pi/2$), a large change in longitude moves you a large distance.

We mentioned above that the right coordinate transformation can convert the polar $ds^2 = d\rho^2 + \rho^2 d\phi^2$ to the Cartesian $ds^2 = dx^2 + dy^2$. However, no coordinate transformation will turn the spherical geometry metric we wrote above into a planar metric. Spherical geometry is intrinsically non-Euclidean.

Metrics in Spacetime

We mentioned in the previous section that one important limitation of Figure 4 (p. 8) is that it shows a curve through space but not time. Our discussion above suffers from the same limitation, discussing metrics exclusively in space. General relativity requires us to define metrics in *spacetime*.

That's a difficult notion, so let's start with the simplest case possible. We will assume there is no gravity, so spacetime is flat. (In this special case, general relativity reduces to *special* relativity.) We will only consider one dimension of space, so spacetime has one spatial coordinate plus time. And by the way, we'll use relativistic units in which $c = 1$, so space and time have the same units.

What is the metric in our simplified spacetime? That is, what will we use to define a kind of "distance" between two different events, separated by a spatial distance dx and a time interval *dt*? You might initially guess that the metric would be the linear distance between the two events on a spacetime drawing:

$$ds^2 = dx^2 + dt^2 \qquad \text{Not the metric of flat spacetime} \tag{1}$$

The problem with Equation 1 is that it changes when you do a Lorentz transformation. That is, two observers in different inertial reference frames would disagree about the "distance," thus defined, between two events. If an object with no (non-gravitational) forces on it will follow the path that minimizes the metric, all observers must agree on the end result.

So what distance function can we define that is invariant under Lorentz transformations? *Hint:* If you've gone through Chapter 1, you know the answer!

We hope you remembered the spacetime interval. (If not, you may want to review Section 1.4.)

$$ds^2 = -dx^2 + dt^2 \qquad \text{The metric of spacetime without gravity} \tag{2}$$

Special relativity says that the spacetime interval between any two events is the same in any inertial reference frame, which makes ds as defined by Equation 2 a plausible metric. That doesn't prove that this particular ds actually is the correct metric for our flat one-dimensional spacetime, but it is. We're not going to offer any further arguments to convince you of that fact. Instead, we're going to focus on what this metric—and then later, a more complicated metric for a more complicated spacetime—tells us about motion. Our analysis will be entirely based on the rule we discussed above:

Absent any non-gravitational forces, an object will follow a "geodesic": a curve that maximizes the metric for each (sufficiently small) step along its path.

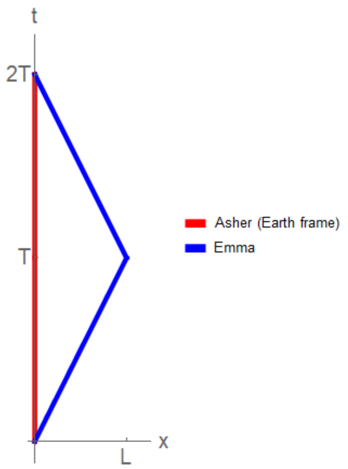


Figure 8 **Asher remains on Earth. Emma journeys outward until time T (as measured by Asher) and then returns home, all at constant speed.**

(Wait, "maximizes"? Because of the negative sign in front of dx^2 in the metric, the geodesics followed by free-falling particles are maxima instead of minima.[2])

To see what that rule can tell us about the behavior of objects, let's bring back our old friends Asher and Emma. As you may recall, Asher stayed on Earth while his twin Emma traveled outward in a rocket and then returned. Figure 8 shows their journeys on a spacetime diagram.

Figure 9 represents one small slice of time along Asher's worldline, somewhere between $t = 0$ and $t = 2T$. Since Asher never does anything, it doesn't matter which slice!

Along the actual path that Asher took, Equation 2 tells us that $ds = dt$. You can integrate this along his full step to conclude that $\Delta s = \Delta t$.

But along any alternate path, each small step involves a ds less than dt. (Again, this follows trivially from Equation 2.) Integrating along such a path will produce a total Δs that is less than Δt.

2 In some texts the spacetime interval, and thus the metric, are defined with the minus sign on time instead of space: $ds^2 = dx^2 - dt^2$. Some equations look different with this convention, but the resulting physics is the same. In that convention the geodesics are paths of minimal distance.

Asher's path therefore provides the largest possible Δs for such each step of his journey, so he is following a geodesic.

The analysis is more mathematically complicated along Emma's journey. But for any small step during her outbound journey, or during her return trip, the end result is the same: the metric $ds^2 = -dx^2 + dt^2$ is maximized along the straight-line path that she follows, so she is also following a geodesic. (One way to show this is to remember that you can do a Lorentz transformation to a reference frame in which Emma's straight line path is vertical, and use the same arguments we just used for Asher's path. Since the spacetime interval is the same in all inertial reference frames, you can conclude that this straight line portion of Emma's path maximizes ds in all reference frames.)

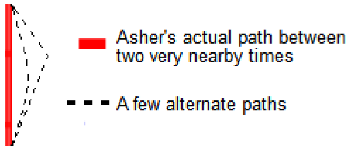


Figure 9 **A few possible ways to get between two events on Asher's world line, separated by no Δx and a small Δt.**

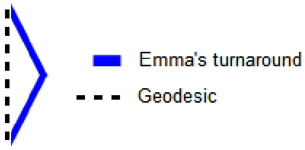


Figure 10 **A very small time interval around Emma's turnaround point.**

Suppose, though, that you draw a small step that includes Emma's turnaround at time T. No matter how small you draw that step, her journey does *not* maximize the metric for that interval. If you integrate ds along Emma's path in Figure 10, you get a smaller total Δs than you would get by connecting the same starting and ending points with a vertical line. For this brief moment, she is not following a geodesic.

What do we learn from all that math? Asher's stationary sojourn, and Emma's outward journey, and her return journey, all describe inertial behavior, freely coasting through space. But Emma's turnaround must result from the application of an external force (her engines). You already knew all that; our point is how all that behavior can be deduced by maximizing the metric in this Euclidean spacetime.

Generalizing, you know that the geodesics for the Euclidean distance $ds^2 = dx^2 + dy^2$ are straight lines. (The shortest distance between two points…) We've now argued that introducing a minus sign doesn't change that fact: the geodesics in Equation 2 correspond to straight lines on a spacetime diagram. That brings us back to Newton's first law: in an inertial reference frame, particles move along straight world lines, meaning they move in straight lines at constant velocity.

The spacetime described by this no-gravity metric, which is the spacetime of special relativity, is called "Minkowski space." To see how gravity can change the geometry of spacetime, let's consider one physically important example of a spacetime other than Minkowski space.

Example: An Expanding Universe

There is a fair amount of overlap between Section 14.4 and the discussion below. We have written the two discussions to be independent, in the sense that either can be read without the other. But reading both will give you a deeper understanding of both general relativity and cosmology.

Astronomical observations suggest that on large scales the universe is approximately homogeneous (the same everywhere) and isotropic (the same in all directions).[3] If we take

3 It may seem like homogeneity implies isotropy, but they are independent conditions. For example, a space filled with a uniform magnetic field could be homogeneous, but the magnetic field would cause objects moving in one direction to behave differently from objects moving in other directions, so the space would not be isotropic.

homogeneity and isotropy as assumptions, that leaves only two properties that distinguish different possible metrics for spacetime. Those two properties are curvature and expansion.

We've already introduced curvature. In general, there are three possibilities for the curvature of space. Zero curvature means at any fixed time the geometry of space is Euclidean. Positive curvature means the space has a metric like the surface of a sphere. (Of course the surface of a sphere would be a 2D space with positive curvature, but the same basic properties hold for a 3D space.) Negative curvature describes a 3D universe that is roughly analogous to the 2D surface of a saddle-shape. We're going to keep this first property simple by continuing to consider a spatially flat (Euclidean) universe, which does seem to describe our own universe very well.

The second property, expansion of the universe, is not a curvature in space—but it is a curvature in spacetime. Expansion can be mathematically described by throwing a new (unitless) function of time, which we will write as $a(t)$, into the spacetime interval in Equation 2. We're also going to move up from one spatial dimension to three, giving us:

$$ds^2 = -(a(t))^2\left(dx^2 + dy^2 + dz^2\right) + dt^2 \qquad \text{The metric of a flat, expanding universe} \qquad (3)$$

The function $a(t)$ is called the "scale factor." It's easy to see that if $a(t) = 1$, Equation 3 reduces to Equation 2. A constant $a(t)$ function produces the flat, non-expanding universe of Minkowski space. In that case we already know that the geodesics are straight lines, so an object will follow a straight-line path at constant speed, just as Newton would have predicted.

For other $a(t)$ functions, the geodesics can get more complicated. But there is one special set of geodesics that is simple no matter what $a(t)$ is. Consider an object at rest, so its world line is a vertical line on a spacetime diagram. In other words, it moves through time but always stays at the same place. Is such a path a geodesic, the path that extremizes the metric?

Yes, it is, and we can see that using the same argument we made in Minkowski space. In the case of Asher and Emma, we saw that Asher's purely vertical worldline was indeed a geodesic: it maximized ds at every step. Now we are throwing in a scale factor, but for the particular case of a vertical worldline, that scale factor is irrelevant. If $dx^2 + dy^2 + dz^2 = 0$ then Equation 3 becomes $ds = dt$. That maximizes the metric because any curved path yields a smaller total Δs, no matter what $a(t)$ is. (We do restrict ourselves to real scale factors.)

We conclude that for any function $a(t)$, one possible behavior for an object experiencing no (non-gravitational) forces is to sit still.

So what effect does $a(t)$ have? To answer that, let's look at two objects at rest at two nearby locations. Because their spatial coordinates aren't changing, $dx^2 + dy^2 + dz^2$ is a non-zero constant. But this constant is *not* the distance between the two objects; the distance is given by the metric.

At any fixed time ($dt = 0$), Equation 3 tells us that the metric is proportional to $a(t)$. If $a(t)$ is an increasing function of time, that means that *when two objects sit perfectly still over a span of time, the distance between those two objects increases.* That's exactly what we see when we look at distant galaxies. An increasing $a(t)$ function describes an expanding universe, such as the one we live in.

As a more complicated example, consider the paths of light beams in this spacetime. We'll drop back down to a one-dimensional universe for the following Active Reading Exercise.

Active Reading Exercise: Light Beams in an Expanding Universe

Consider a spacetime with the metric $ds^2 = -(a(t))^2 dx^2 + dt^2$. Using the fact that a light beam always moves along a path with $ds = 0$, calculate the speed dx/dt of a light beam in this spacetime as a function of time. Using your result, sketch a spacetime diagram with a light cone whose vertex is at the origin. *Hint*: It will not look like a cone.

Setting $ds = 0$ immediately gives $dx/dt = 1/a(t)$. Because a is increasing, dx/dt is decreasing, as shown in our spacetime diagram at the link below. www.cambridge.org/felder-modernphysics/activereadingsolutions

A light beam slows down over time! That can't be right, can it?

Well, it is and it isn't.

Remember that above we wrote two different metrics for a flat plane. These two metrics described the same geometry—the same scenario in the same universe—but expressed in different coordinate systems (Cartesian and polar). The point was that, for any given spacetime, the form of the metric depends on the coordinates you use.

We've been writing the metric for an expanding universe in "comoving coordinates." That means that each galaxy remains at the same spatial coordinates, and the distance between any two fixed values of x grows with time. In those coordinates the speed of light dx/dt seems to slow down because each coordinate distance Δx represents an ever-larger physical distance over time.

You can instead define "physical coordinates" that don't grow apart over time: $x' = ax$ (Figure 11). But it turns out the paths of light beams are even more complicated in physical coordinates than in comoving coordinates. In physical coordinates the speed of a light beam depends on its position x and on the direction of its motion, and light beams can either slow down or speed up over time.

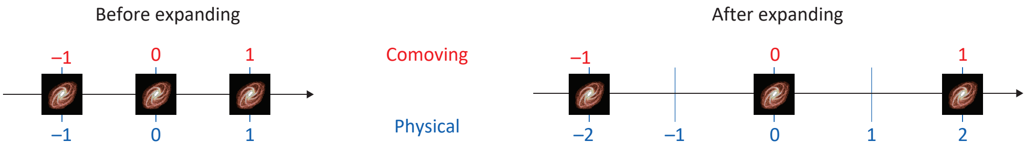


Figure 11 **In an expanding universe comoving coordinates move apart with the galaxies. Physical coordinates stay at fixed distances.**

The big takeaway from Equation 3 is that—no matter what coordinates you use to describe it—a constant $a(t)$ function describes a non-expanding universe, and an increasing $a(t)$ function describes an expanding universe in which actual distances between objects grow over time.

Our universe is expanding, so the $a(t)$ function that describes our universe is increasing. But how fast is it increasing? The math of GR (which we are not teaching you here) says it depends on what kind of energy fills the universe.

- For an expanding universe filled with ordinary matter, a is proportional to $t^{2/3}$.
- For an expanding universe filled with electromagnetic radiation, a is proportional to $t^{1/2}$.

Our universe is not expanding in either of those ways. Rather, we seem to be transitioning towards exponential growth, meaning that a is proportional to e^{Ht} with a constant H. That suggests that our universe is dominated by some form of energy other than matter or radiation. We don't know what that form of energy is, so we call it "dark energy."

Stepping Back: Reference Frames and Coordinates

In special relativity we sometimes loosely describe a "reference frame" as the point of view of an observer, but more precisely a reference frame is a set of coordinates used to label all points in spacetime. If the metric of spacetime in those coordinates is of the form $ds^2 = -(dx^2 + dy^2 + dz^2) + dt^2$, then the reference frame is inertial. If the metric is anything else then the reference frame is non-inertial. In special relativity, you can always choose a set of coordinates that represents an inertial frame.

In the presence of gravity, by contrast, the metric cannot be transformed into an inertial form by any coordinate transformation. Those non-inertial frames can have strange effects like light beams slowing down over time, or moving at different speeds in different locations.

There is a theorem in general relativity, however, that says that even in the presence of gravity you can always find a set of coordinates that makes the metric look inertial *in a small neighborhood around any point*. What that means in practice is that all inertial observers (observers experiencing no non-gravitational forces) will see the universe obeying the laws of special relativity in their local neighborhoods. For example, in our expanding universe there are distant galaxies moving away from us much faster than 3×10^8 m/s, speeds forbidden in special relativity. But you will never see an object move right past you going faster than c. Even in general relativity, you can't outrace a light beam.

Conclusion: What We Are and Aren't Telling You

If you someday take a full course in general relativity, you will find that much of the theory boils down to two fundamental equations, each of which expresses half of Wheeler's one-sentence summary of GR.

- *Matter tells spacetime how to curve*: The "Einstein equation" relates the curvature of spacetime, as expressed by the metric function, to the properties of matter and energy that fill that spacetime.
- *Spacetime tells matter how to move*: The "geodesic equation" tells you how to find geodesics of spacetime from the metric. In the absence of non-gravitational forces, all objects will follow these geodesics.

Both of those equations depend on mathematics beyond the scope of this book. But if you begin a GR course with the general outline from this section, you won't lose sight of what all that math is telling you.