Chapter 10
Relativity

With his special and general theories of relativity, Einstein revolutionized our understanding of space, time, gravity, and hence motion. Why, then, have we spent so much time with Newton? Newtonian physics is a good approximation when motion is slow compared with the speed of light and gravity is "weak" in a sense to be defined. A lot of astrophysics research is still carried out under these assumptions. That said, discussing our modern conception of gravity and motion opens fascinating topics such as the weirdness of spacetime around black holes and (in Chap. 11) the expanding universe.[1]

10.1 Space and Time: Classical View

All of our discussion of motion so far has relied on an implicit understanding of "space" and "time." Intuitively, we think everyone agrees on what space and time are; we imagine there are universal rods and clocks we can use to define them. Although physicists knew that only *relative* motion matters for inertial observers, they assumed that a universal, absolute reference frame does exist.

To glimpse some consequences of this assumption, let's examine the relation between two inertial reference frames that are moving relative to each other. Consider one frame (x, y, z, t), and a second frame (x', y', z', t') moving relative to the first with a constant speed u in the x-direction. If time and space are the same in both frames, the coordinates must be the same except for a translation in the x-direction:

[1]Parts of this presentation draw from books by Carroll and Ostlie [1] and Schutz [2].

C. Keeton, *Principles of Astrophysics: Using Gravity and Stellar Physics to Explore the Cosmos*, Undergraduate Lecture Notes in Physics, DOI 10.1007/978-1-4614-9236-8_10,

$$
\begin{aligned}
t &= t' \\
x &= x' + ut' \\
y &= y' \\
z &= z'
\end{aligned}
$$

This is referred to as the **Galilean transformation** between reference frames. A direct corollary is

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}x'}{\mathrm{d}t'} + u$$

The velocity with respect to the unprimed frame is the simple sum of the velocity with respect to the primed frame, plus the velocity of the primed frame with respect to the unprimed frame. This certainly makes sense intuitively.

If we take a second derivative, we find

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = \frac{\mathrm{d}^2x'}{\mathrm{d}t'^2}$$

so the accelerations are the same in both frames. Then by Newton's second law the net force must likewise be the same, and the laws of physics are equally valid in either frame.

This all made sense until physicists studied electricity, magnetism, and light in the late nineteenth century. On the theoretical side, James Clerk Maxwell's theory of electrodynamics indicated that light is an electromagnetic wave traveling at speed $c = 3.0 \times 10^8\,\mathrm{m\,s^{-1}}$ *in all inertial reference frames*. On the experimental side, Albert Michelson and Edward Morley tried to measure differences in the speed of light emitted by sources moving at different speeds—*and found that there were no differences*. Physicists were stunned. Some suggested there must be a problem in Maxwell's theory. Others supposed there was some substance known as æther pervading the universe whose properties caused all inertial observers to measure the same speed of light.

10.2 Special Theory of Relativity

Albert Einstein took a different approach: he wondered whether the problem lay in misconceptions about space and time. Instead of assuming absolute space and time, he took an operational view: he described how to use a system of rigid rods and synchronized clocks to construct a coordinate system in any reference frame.[2] To Einstein, space and time could be real only to the extent that they could be measured.

That only served to define coordinates within a given reference frame. To relate different reference frames, Einstein proposed two postulates [4]:

1. The equations of motion of any (mechanical) system are the same in all inertial reference frames.
2. The speed of light is constant and universal.

The first postulate is called the **principle of relativity**, and it predated Einstein. What Einstein did was introduce the second postulate as an extension of the theory of electrodynamics, and show that together the two postulates are inconsistent with the Galilean transformation.

10.2.1 Lorentz Transformation

Einstein worked out what different observers would have to say about space and time in order for them to agree on the speed of light. He found the relations:

$$ct = \gamma\, ct' + \gamma\, \beta\, x' \tag{10.1a}$$

$$x = \gamma\, x' + \gamma\, \beta\, ct' \tag{10.1b}$$

$$y = y' \tag{10.1c}$$

$$z = z' \tag{10.1d}$$

where

$$\beta = \frac{u}{c} \quad \text{and} \quad \gamma = \left(1 - \beta^2\right)^{-1/2} \tag{10.2}$$

Note that we use ct and ct' because working with a quantity that has dimensions of length clarifies the interplay between time and space coordinates. The inverse relations are:

$$ct' = \gamma\, ct - \gamma\, \beta\, x \tag{10.3a}$$

$$x' = \gamma\, x - \gamma\, \beta\, ct \tag{10.3b}$$

$$y' = y \tag{10.3c}$$

$$z' = z \tag{10.3d}$$

[2] Peter Galison [3] notes that, as a clerk in the Swiss Patent Office, Einstein probably saw many patent applications for schemes to synchronize clocks. The spread of the railroad and telegraph had prompted a need for long-distance synchronization.

The relations (10.1) and (10.3) were already known as the **Lorentz transformation** after Hendrik Lorentz. They had been derived in electrodynamics as the transformation that preserves Maxwell's equations in both reference frames.[3] What Einstein offered was a sweeping new interpretation: time and space are no longer separate, absolute quantities. Rather, they are linked in a 4-dimensional structure we now call **spacetime**. Points in spacetime are referred to as **events**. (We will say more about the structure of spacetime beginning in Sect. 10.5.)

We can use the Lorentz transformation to relate velocities measured in the primed and unprimed frames. The differential version of Eq. (10.1) is

$$\begin{aligned} c\,\mathrm{d}t &= \gamma\, c\, \mathrm{d}t' + \gamma\, \beta\, \mathrm{d}x' \\ \mathrm{d}x &= \gamma\, \mathrm{d}x' + \gamma\, \beta\, c\, \mathrm{d}t' \\ \mathrm{d}y &= \mathrm{d}y' \\ \mathrm{d}z &= \mathrm{d}z' \end{aligned}$$

Let's rewrite these relations using the components of velocity measured in the primed frame: $v'_x = \mathrm{d}x'/\mathrm{d}t'$, $v'_y = \mathrm{d}y'/\mathrm{d}t'$, and $v'_z = \mathrm{d}z'/\mathrm{d}t'$. We also use $\beta = u/c$. These substitutions yield

$$\begin{aligned} \mathrm{d}t &= \left(1 + \frac{u v'_x}{c^2}\right) \gamma\, \mathrm{d}t' \\ \mathrm{d}x &= \left(v'_x + u\right) \gamma\, \mathrm{d}t' \\ \mathrm{d}y &= v'_y\, \mathrm{d}t' \\ \mathrm{d}z &= v'_z\, \mathrm{d}t' \end{aligned}$$

Now we can find the velocity components in the unprimed frame:

$$v_x = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{v'_x + u}{1 + u v'_x/c^2} \tag{10.4a}$$

$$v_y = \frac{\mathrm{d}y}{\mathrm{d}t} = \frac{v'_y}{\gamma(1 + u v'_x/c^2)} \tag{10.4b}$$

$$v_z = \frac{\mathrm{d}z}{\mathrm{d}t} = \frac{v'_z}{\gamma(1 + u v'_x/c^2)} \tag{10.4c}$$

Under the Lorentz transformation, velocity in the unprimed frame is no longer a simple sum of the velocity in the primed frame and the velocity of the frame itself. In Problem 10.2 you can see how the modified transformation explains why the speed of light does not depend on the speed of the source.

[3] Maxwell's equations are not invariant under the Galilean transformation.

Non-relativistic Limit

Our experience and intuition are more closely aligned with the Galilean transformation than the Lorentz transformation, but of course they are associated with motion that is slow compared with the speed of light. Let's see if we can confirm our intuition by examining the Lorentz transformation when $u/c \ll 1$. A Taylor series expansion of Eq. (10.1) yields

$$t \approx t' + \frac{ux'}{c^2} + \mathscr{O}\left(\frac{u^2}{c^2}\right) \quad \text{and} \quad x \approx x' + ut' + \mathscr{O}\left(\frac{u^2}{c^2}\right)$$

Since the speed of light is so large, ux'/c^2 is small and to a very good approximation we can write

$$t \approx t' \quad \text{and} \quad x \approx x' + ut$$

A similar analysis applied to Eq. (10.4) yields

$$v_x \approx v'_x + u \qquad v_y \approx v'_y \qquad v_z \approx v'_z$$

Thus, the Lorentz transformation does not actually invalidate the Galilean transformation (which is reassuring since the latter was the basis of all physics prior to the twentieth century). Rather, it clarifies that the Galilean transformation should be used only when motion is slow compared with the speed of light.

10.2.2 Loss of Simultaneity

When speeds are not small, however, we must use the full Lorentz transformation. Following Einstein, we can use some *gedanken* (German for "thought") experiments to uncover some consequences of the interplay between space and time. First, consider two lights that are set up to flash at the same time in the primed coordinate system (which we can take to be $t' = 0$). What are the times of the flashes in the unprimed coordinate system? Let the first event be the flash of light #1:

$$\text{event 1:} \quad (t', x') = (0, x'_1) \quad \Rightarrow \quad ct_1 = \gamma\, \beta\, x'_1$$

Let the second event be the flash of light #2:

$$\text{event 2:} \quad (t', x') = (0, x'_2) \quad \Rightarrow \quad ct_2 = \gamma\, \beta\, x'_2$$

The time between the two flashes is $\Delta t' = 0$ in the primed frame, but the time between the two flashes in the unprimed frame is

$$c\, \Delta t = \gamma\, \beta\, (x'_2 - x'_1) \neq 0$$

In other words, events that are simultaneous in one reference frame *are not simultaneous in other reference frames*. This is the first indication that there is something decidedly non-intuitive in the new way of thinking about the universe.

10.2.3 Time Dilation

Now focus on a single flashing light and consider the time between flashes. For simplicity, let's put the light at the origin of the primed frame. If we again consider two flashes as two spacetime events, we have:

$$\begin{aligned} &\text{event 1:} \quad (t', x') = (0, 0) \qquad \Rightarrow \qquad t_1 = 0 \\ &\text{event 2:} \quad (t', x') = (\Delta t', 0) \quad \Rightarrow \qquad t_2 = \gamma \, \Delta t' \end{aligned}$$

In other words, the time between flashes as measured in the unprimed frame is

$$\Delta t = \gamma \, \Delta t' \tag{10.5}$$

Since $\gamma \geq 1$, more time passes between flashes in the unprimed frame than in the primed frame. This effect is known as **time dilation**. If we think of a flashing light as a kind of clock, we can distill this into the maxim "moving clocks run slowly."

If the measurement of time between events depends on the reference frame, how can we single out a frame to focus on when we study physical laws? The most natural quantity is the time interval measured by a clock at rest with respect to the events, which has the advantage of being the *smallest* time interval that any clock will measure. We call this the **proper time**.

Time dilation is a definite and weird prediction of relativity, so it deserves to be tested experimentally. One of the classic tests was performed in 1963, when David Frisch and James Smith [5] studied elementary particles called muons coming from space. Frisch and Smith compared the number of muons detected at the top of Mt. Washington in New Hampshire (1,907 m above sea level) with the number detected at sea level. It takes a certain amount of time Δt for muons to travel the intervening distance, but the measurements indicated that the muons "experienced" a much shorter interval $\Delta t' < \Delta t$. The experiment confirmed predictions of time dilation, as you can see in more detail in Problem 10.3. In 1971, Joseph Hafele and Richard Keating [6, 7] flew atomic clocks on airplanes to do a more controlled test of time dilation. That experiment involved gravity as well as motion, so we will consider it among tests of general relativity (Sect. 10.4.5). Today, relativistic time dilation is built into the Global Positioning System (Sect. 10.4.6).

10.2.4 Doppler Effect

In a final use of the flashing light, let's consider the times when flashes reach an observer who is stationary at the origin of the unprimed frame. This is what the observer (whether a person or an instrument) would actually measure. The first flash occurs at coordinates $(t'_1, x'_1) = (t'_1, 0)$ in the primed frame, which correspond to coordinates $(t_1, x_1) = (\gamma\, t'_1\,,\ \gamma\, u\, t'_1)$ in the unprimed frame. In order for this flash to be observed, it must travel to the observer at the origin. The distance it must travel is $\gamma\, u\, t'_1$, and the time it takes is $\gamma\, u\, t'_1/c$. The time at which the flash is observed is therefore

$$t_{1,\text{obs}} \;=\; \gamma\, t'_1 + \frac{\gamma\, u t'_1}{c} \;=\; \gamma \left(1 + \frac{u}{c}\right) t'_1$$

By similar reasoning, we find the time at which the second flash is observed to be

$$t_{2,\text{obs}} = \gamma \left(1 + \frac{u}{c}\right) t'_2$$

Thus, the time that elapses between observations of the two flashes is

$$\Delta t \;=\; t_{2,\text{obs}} - t_{1,\text{obs}} \;=\; \gamma \left(1 + \frac{u}{c}\right) \Delta t' \;=\; \left(\frac{1 + u/c}{1 - u/c}\right)^{1/2} \Delta t'$$

where $\Delta t' = t'_2 - t'_1$ is the time interval between flashes in the frame in which they are emitted, and in the last step we substitute for γ using Eq. (10.2).

Now if we replace the flashes with peaks of a wave, the time between peaks is the inverse of the frequency of the wave: so $\nu_{\text{obs}} = 1/\Delta t$ in the frame of observations, and $\nu_{\text{em}} = 1/\Delta t'$ in the frame of emission. Then we have

$$\frac{\nu_{\text{obs}}}{\nu_{\text{em}}} = \left(\frac{1 - u/c}{1 + u/c}\right)^{1/2} \tag{10.6}$$

Equivalently, in terms of wavelength we can use $\lambda \propto \nu^{-1}$ to write

$$\frac{\lambda_{\text{obs}}}{\lambda_{\text{em}}} = \left(\frac{1 + u/c}{1 - u/c}\right)^{1/2} \tag{10.7}$$

This is the **relativistic Doppler effect**. It says that if a light source is moving away from the observer ($u > 0$), the observed frequency of light is lower than the emitted frequency; this corresponds to a longer wavelength and hence a redder color, so we call this a **redshift**. Conversely, if a light source is moving toward the observer ($u < 0$), the Doppler effect produces a **blueshift**.

In the non-relativistic limit, we can use a Taylor series expansion to write

$$\frac{\lambda_{\rm obs}}{\lambda_{\rm em}} \approx 1 + \frac{u}{c} + \mathscr{O}\left(\frac{u^2}{c^2}\right)$$

We often express the shift in terms of the change in wavelength, $\Delta\lambda = \lambda_{\rm obs} - \lambda_{\rm em}$:

$$\frac{\Delta\lambda}{\lambda_{\rm em}} \approx \frac{u}{c} \tag{10.8}$$

This is the Doppler shift of light when the source of light is moving at non-relativistic speeds. It is what we use, for example, to measure the motions of stars and discover that they are orbited by planets (Sect. 4.3.1).

10.2.5 Length Contraction

Let's change gedanken tools and consider a ruler oriented along the x-axis that is at rest in the primed frame. Place one end of the ruler at $x'_1 = 0$ and the other end at $x'_2 = L'$, so the ruler's length in the primed frame is L'. What is the length of the ruler in the unprimed frame? It may seem backward at first, but let's use the Lorentz transformation $x'_1 = \gamma\, x_1 - \gamma\, u\, t_1$ (and likewise for the other end). Then we have

$$x'_2 - x'_1 \;=\; (\gamma\, x_2 - \gamma\, u\, t_2) - (\gamma\, x_1 - \gamma\, u\, t_1) \;=\; \gamma\,(x_2 - x_1) - \gamma\, u\,(t_2 - t_1)$$

It is important to measure the ends of the ruler *at the same time* in the unprimed frame. Then we can put $t_2 - t_1 = 0$ and obtain

$$x'_2 - x'_1 = \gamma\,(x_2 - x_1) \quad \Rightarrow \quad L' = \gamma\, L \quad \Rightarrow \quad L = \frac{L'}{\gamma}$$

In other words, the moving ruler appears to have a length $L = L'/\gamma$ that is *shorter* than its length at rest. This is known as **length contraction**: moving objects appear shorter in the direction of motion. As with proper time, if we want to single out a particular length then we usually use the **proper length** measured when the object is at rest.

10.3 General Theory of Relativity

In order to deal with gravity, Einstein had to generalize his theory from inertial to accelerated reference frames. This led him to sophisticated mathematical structures including non-Euclidean geometry, manifolds, tensors, and more. We will glimpse some of the mathematical framework in Sect. 10.5, but first let's examine the physical principles that underlie general relativity.

10.3.1 Concepts of General Relativity

General relativity is a **geometric theory of gravity**: acceleration is a consequence of the curvature of spacetime. There are two key concepts governing the interaction between curvature and mass (as stated by Misner, Thorne and Wheeler [8]):

1. "Space acts on matter, telling it how to move."
2. "In turn, matter reacts back on space, telling it how to curve."

People often think of these in terms of a rubber sheet analogy. Imagine stretching a rubber sheet so it lies flat. This is a model of a 2-dimensional universe described by special relativity. Now place a bowling ball on the sheet. The bowling ball deforms the sheet; this is point #2 above. Then roll a ping-pong ball near the bowling ball. The curvature induced by the bowling ball controls how the ping-pong ball moves; this is point #1.

It is important to understand that this is an *analogy*. It is imperfect because it describes the 2-d rubber sheet as being curved into the third dimension. For the 3-d spatial universe, we would have to think of the curvature as extending into a fourth spatial dimension. I cannot picture such a thing! Also, the analogy works only if there is *external* gravity pulling on the bowling ball to distort the rubber sheet. In general relativity, everything needs to happen *within* the theory. So the rubber sheet is useful as a pictorial analogy, but please do not take it too literally. We will be more precise about describing curvature soon.

10.3.2 Principle of Equivalence

When Einstein was trying to figure out how to describe gravity and acceleration, he had an important thought: "If a person falls freely he will not feel his own weight." [9] To be more precise, let's go back to Newton for a moment. We have often used the equation

$$\frac{GMm}{r^2} = F = ma \tag{10.9}$$

and cancelled the m's from both sides. But it is not obvious that they have to be the same. The m on the left describes how an object *feels* the force of gravity; we might call it the "gravitational mass," m_g. The m on the right describes an object's inertia and how it *responds* to a force; we might call it the "inertial mass," m_i. We really ought to rewrite Eq. (10.9) as

$$\frac{GMm_g}{r^2} = F_g = m_i a$$

which yields the acceleration

$$a = \frac{GM}{r^2}\frac{m_g}{m_i}$$

It is an experimental fact that the ratio m_g/m_i is 1. Galileo is said to have shown this by dropping balls off the Leaning Tower of Pisa. Modern versions of the experiment reveal that the difference between gravitational and inertial masses is less than 1 part in 10^{12} (e.g., [10]). Therefore we can say to high precision that the acceleration due to gravity is independent of mass—all objects fall at the same rate.[4]

If that is true, then no experiment can reveal the acceleration because the equipment will fall at the same rate as the sample. By extension, no experiment can reveal that gravity is at work! Another way to say this is that a freely falling laboratory is equivalent to a lab floating in empty space. Within such a freely falling lab, we can apply the principles of special relativity. This simplifies things quite a lot.

Strictly speaking, this reasoning holds only in a region of space in which the acceleration due to gravity is uniform. Since objects on the surface of Earth fall towards the center of the planet, objects at different positions fall in different directions; that is enough to reveal the gravity. But if we pick a region that is small enough, these effects are negligible.

Einstein turned this idea into the foundation of his theory of gravity, calling it the **principle of equivalence**:

- All local, freely falling, non-rotating frames of reference are equivalent for performing physical experiments.

This is the fundamental principle that allows us to identify some physical aspects of general relativity.

10.3.3 Curvature of Spacetime

Let's apply the principle of equivalence to some thought experiments to understand how gravity affects spacetime. Consider a lab in freefall in a gravitational field where the acceleration due to gravity is g, as depicted in Fig. 10.1. Suppose a light source on the left-hand wall is pointed toward the right. By the principle of equivalence, the lab acts as a local inertial reference frame, so an observer in the lab would see the light travel in a straight line from one side to the other.

[4] This is not true of other forces. Consider the acceleration created by the electric force acting on an object with mass m and charge q near another charge Q: $a = Qq/mr^2$ does depend on mass.

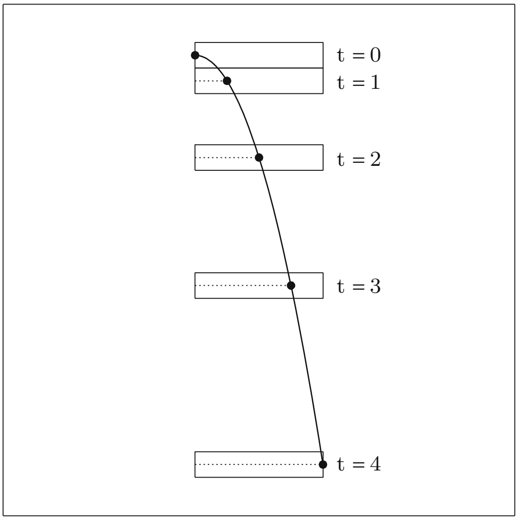

Fig. 10.1 Setup for gedanken experiment #1. A light ray travels horizontally in a lab that is in freefall with uniform acceleration. To an observer in the lab, the light ray travels straight across the room (*dotted lines*). But to an outside observer, the light ray follows a curved trajectory

What would be seen by an observer on the ground (who is stationary in the gravitational field)? As the light moves to the right in the lab, the lab accelerates downward, so the trajectory of the light looks like a parabola. Gravity has caused light to curve!

To be specific, let's write the equation of the trajectory. If the light starts at $(x, y) = (0, 0)$, its position as a function of time is

$$x = ct \quad \text{and} \quad y = -\frac{1}{2}gt^2$$

Eliminating t yields

$$y = -\frac{gx^2}{2c^2}$$

We know this trajectory is curved, but by how much? To quantify curvature, think about a circle:

$$x^2 + y^2 = R_c^2 \quad \Rightarrow \quad y = \left(R_c^2 - x^2\right)^{1/2}$$

If we know $y(x)$, we can extract the radius R_c as follows:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{x}{(R_c^2 - x^2)^{1/2}}$$

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = -\frac{R_c^2}{(R_c^2 - x^2)^{3/2}}$$

$$\text{at } x = 0: \quad \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = -\frac{1}{R_c}$$

(The minus sign means the circle is curved downward.) As a general rule, then, we can define the **radius of curvature** for a trajectory $y(x)$ as

$$R_c = \frac{1}{|\mathrm{d}^2y/\mathrm{d}x^2|}$$

Heuristically, R_c is the distance over which the trajectory deviates significantly from a straight line, so a smaller value corresponds to a greater curvature. For our gedanken experiment, the radius of curvature is found as follows:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{gx}{c^2} \quad \Rightarrow \quad \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = -\frac{g}{c^2} \quad \Rightarrow \quad R_c = \frac{c^2}{g}$$

If we assume Newtonian gravity for simplicity, the gravitational acceleration at a distance r from an object of mass M is

$$g = \frac{GM}{r^2}$$

so the radius of curvature for light is

$$R_c = \frac{c^2r^2}{GM}$$

Example: Earth

Near the surface of Earth, the acceleration due to gravity is $g = 9.80\,\mathrm{m\,s^{-2}}$ so the radius of curvature for light is

$$R_c = \frac{c^2}{g} = 9.17 \times 10^{15}\,\mathrm{m} = 0.97\,\mathrm{ly}$$

The radius of curvature is huge—far, far bigger than the size of Earth—which means the curvature is quite small. Nevertheless, it is significant enough to produce all the familiar effects of gravity.

Example: Black Hole

Near the event horizon,

$$r = R_S = \frac{2GM}{c^2} \quad \Rightarrow \quad R_c = \frac{c^2 R_S^2}{GM} = 2R_S$$

The radius of curvature is comparable to the size of the event horizon, which means gravity is strong.

Bottom line: we have found that gravity causes light to move on a curved trajectory. A similar analysis could be done for material particles. Operationally, then, what we mean when we say spacetime is curved is that objects follow curved trajectories. To summarize:

$$\text{objects follow curved trajectories} \quad \Leftrightarrow \quad \text{space is curved}$$

10.3.4 Gravitational Redshift and Time Dilation

Consider the same freely falling lab, only now put the light source on the floor and have it shine upward. By the time the light reaches a detector in the ceiling, the lab will be moving faster because of the acceleration. If the light moves a distance h, the time elapsed is $t = h/c$ and the lab's new speed is $u = -gt = -gh/c$ (where the minus sign means downward). By Eq. (10.6), there should be a Doppler blueshift of the form

$$\Delta\nu_{\text{Doppler}} = \nu_{\text{obs}} - \nu_{\text{em}} = \nu_{\text{em}} \left[\left(\frac{1 - u/c}{1 + u/c} \right)^{1/2} - 1 \right] \approx -\nu_{\text{em}} \frac{u}{c} \approx \nu_{\text{em}} \frac{gh}{c^2}$$

(assuming $u \ll c$). Here is the crux of this experiment: if there were a Doppler shift, we would know the lab is accelerating, and that would violate the equivalence principle. The only way out is to say that gravity causes the frequency of light to shift by just the right amount to cancel the Doppler shift. In other words, there must be a **gravitational redshift**

$$\Delta\nu_{\text{grav}} \approx -\nu_{\text{em}} \frac{gh}{c^2} \tag{10.10}$$

This actually makes sense physically: light loses energy as it moves against gravity, and since $E \propto \nu$ the frequency must decrease.

The preceding analysis assumed a constant gravitational acceleration. To deal with the general case, we can use the gravitational acceleration $g = GM/r^2$ and change the height to $\mathrm{d}r$ and the frequency shift to $\mathrm{d}\nu$, obtaining

$$\mathrm{d}\nu \approx -\nu \frac{GM}{c^2 r^2} \mathrm{d}r$$

We can then integrate:

$$\int_{\nu_i}^{\nu_f} \frac{\mathrm{d}\nu}{\nu} \approx -\int_{r_i}^{r_f} \frac{GM}{c^2 r^2}\,\mathrm{d}r$$
$$\ln\frac{\nu_f}{\nu_i} \approx \frac{GM}{c^2}\left(\frac{1}{r_f} - \frac{1}{r_i}\right)$$
$$\frac{\nu_f}{\nu_i} \approx \exp\left[\frac{GM}{c^2}\left(\frac{1}{r_f} - \frac{1}{r_i}\right)\right] \approx 1 + \frac{GM}{c^2}\left(\frac{1}{r_f} - \frac{1}{r_i}\right)$$

where in the last step we use the Taylor series expansion $e^x \approx 1 + x$ for $x \ll 1$. It is convenient to take the "final" point to be at infinity, corresponding to an observer far from the object. This yields

$$\frac{\nu_\infty}{\nu(r)} \approx 1 - \frac{GM}{c^2 r}$$

The oscillations of the light act as a kind of clock, where the elapsed time is $t \propto \nu^{-1}$. We can therefore change the frequency equation into time,

$$\frac{t(r)}{t_\infty} \approx 1 - \frac{GM}{c^2 r}$$

This is **gravitational time dilation**: a clock in a gravitational field runs more slowly than a clock that is far away in empty space.

In this analysis we have made Taylor series approximations and computed the leading order relativistic effect. An exact analysis gives (see Sect. 10.6.1)

$$\frac{t(r)}{t_\infty} = \left(1 - \frac{2GM}{c^2 r}\right)^{1/2} \tag{10.11}$$

Gravitational time dilation becomes strong only when r gets close to $2GM/c^2$. We will see more about this when we study black holes.

Example: Surface of Earth

Clocks on the surface of Earth should run slower than clocks far away in empty space. How much slower? The difference in elapsed time is

$$\frac{\Delta t}{t_\infty} = \frac{t(r) - t_\infty}{t_\infty} \approx -\frac{GM_\oplus}{c^2 R_\oplus}$$
$$\approx -\frac{(6.67\times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}) \times (5.97\times 10^{24}\,\mathrm{kg})}{(3.0\times 10^{8}\,\mathrm{m\,s^{-1}})^2 \times (6.38\times 10^{6}\,\mathrm{m})}$$
$$\approx -7\times 10^{-10}$$

To put this number in context: if your life expectancy is 100 years, you would get to live about 2 s longer on Earth than if you were in space with no gravity. Please note, though, that your experience of time is unaffected by acceleration or gravity; you would not actually have more "time" to enjoy life. You would just appear to age slowly as seen by those living in weaker gravity, while to you they would seem to age quickly.

10.4 Applications of General Relativity

In the previous section we used gedanken experiments to discover the curvature of spacetime, gravitational redshift, and gravitational time dilation. Now let's consider several real experiments that have confirmed these predictions of general relativity.

10.4.1 Mercury's Perihelion Shift (1916)

When we studied planetary motion (Chaps. 3 and 4), we said a planet follows a perfectly elliptical orbit and traces it over and over again. Strictly speaking that is true only in an ideal two-body problem. When a planet's orbit is dominated by the Sun but perturbed by another planet, the situation is only approximately two-body. The resulting orbit can be thought of as an ellipse that precesses, or rotates a little each time the planet goes around (see Fig. 10.2). We can quantify the effect by measuring the shift in the perihelion position.[5]

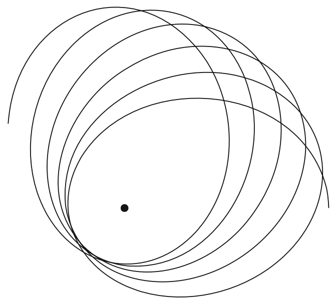

Fig. 10.2 An illustration of perihelion shift. Roughly speaking, the orbit is approximately elliptical but the ellipse rotates with time, which causes the location of perihelion to vary. The effect shown here is greatly exaggerated, with $\epsilon = 0.05$ (compared with $\epsilon = 8 \times 10^{-8}$ for Mercury)

[5]We could use any part of the orbit, but perihelion is distinctive.

In our Solar System, the largest perihelion shift is 560 arcsec/century for Mercury. Most of the shift can be attributed to perturbations from other planets, but after all the known planets are taken into account a shift of 43 arcsec/century remains unexplained. Historically, some people speculated that there might be another planet closer to the Sun than Mercury that caused the additional perihelion shift. The hypothetical planet was called Vulcan [11].

When Einstein considered Mercury's orbit in the context of general relativity, he discovered that it would be described by the equation of motion (see Sect. 10.6.4)

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} = -\frac{GM}{r^2} + \frac{\ell^2}{r^3} - \frac{3GM\ell^2}{c^2 r^4}$$

where $\ell = r^2 \mathrm{d}\phi/\mathrm{d}\tau$ is the specific angular momentum, which is conserved. The first term is standard Newtonian gravity, and the second term is the usual centrifugal term. The third term is new in general relativity, and it perturbs the orbit away from a pure ellipse. To see this, let's go all the way back to our analysis of Newtonian gravity in Chap. 3. Recall that we changed independent variables from time to angle, and we put $r = 1/u$. Repeating the analysis yields

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} = -\ell^2 u^2 \frac{\mathrm{d}^2 u}{\mathrm{d}\phi^2}$$

so the equation of motion becomes

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\phi^2} + u = \frac{GM}{\ell^2} + \frac{3GM}{c^2} u^2$$

Let's define

$$\epsilon = \frac{3(GM)^2}{c^2 \ell^2} \tag{10.12}$$

and then rewrite the equation of motion as

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\phi^2} + u = \frac{GM}{\ell^2} + \epsilon \frac{\ell^2}{GM} u^2$$

With $\epsilon = 0$ this would be Newtonian gravity and the solution would be an ellipse. For Mercury, ϵ is very small and so we can look for a solution that is perturbed away from an ellipse. The solution has the form [12]

$$u(\phi) \approx \frac{GM}{\ell^2} \left\{ 1 + e\cos[\phi(1-\epsilon)] + \epsilon \left[1 + e^2 \left(\frac{1}{2} - \frac{1}{6}\cos 2\phi \right) \right] + \mathscr{O}\left(\epsilon^2\right) \right\}$$

Notice the second term. In order for $\cos[\phi(1-\epsilon)]$ to complete a full cycle, ϕ has to range from 0 to $2\pi/(1-\epsilon) \approx 2\pi(1+\epsilon)$. Therefore it takes an extra azimuthal angle $\Delta\phi \approx 2\pi\epsilon$ for the planet to return to perihelion.

How strong is the effect? Using Eq. (3.11) we can rewrite ℓ in terms of orbital elements and obtain

$$\epsilon = \frac{3GM}{c^2 a(1-e^2)} \tag{10.13}$$

Mercury has $a = 0.387$ AU and $e = 0.206$, yielding

$$\epsilon = \frac{3 \times 6.67 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}} \times 1.99 \times 10^{30}\,\mathrm{kg}}{(3.0 \times 10^8\,\mathrm{m\,s^{-1}})^2 \times (0.387 \times 1.50 \times 10^{11}\,\mathrm{m}) \times (1 - 0.206^2)} = 8 \times 10^{-8}$$

The precession is an angle of about $2\pi\epsilon$ per orbit, so given Mercury's orbital period of $P = 87.97$ day the precession rate is

$$\frac{2\pi\epsilon}{P} = \frac{2\pi \times 8 \times 10^{-8}}{87.97 \times 86{,}400\,\mathrm{s}} = 6.6 \times 10^{-14}\,\mathrm{rad\,s^{-1}} = 43\,\mathrm{arcsec/century}$$

This was the first good explanation for Mercury's perihelion shift, and it convinced Einstein that he was on the right track with his new theory of gravity.

10.4.2 Bending of Light (1919)

We have already discussed Einstein's prediction for the bending of light by the Sun, and the measurements in 1919 and 1922 that confirmed the prediction. This was the second significant test of general relativity, and the first true prediction. (Einstein's explanation of Mercury's perihelion shift was "merely" an explanation of existing data.)

10.4.3 Gravitational Redshift on Earth (1960)

In Sect. 10.3.4 we discussed gravitational time dilation on the surface of Earth. While the effect is small, it turns out that we can measure the corresponding gravitational redshift. Atomic nuclei have energy levels just like atomic electrons, so they can produce emission or absorption lines in energy spectra. The difference is that nuclear lines generally involve much higher energies and are very narrow. For example, iron-57 has a spectral line with energy $E = 14.4$ keV and line width $\delta E \sim 10^{-11}$ keV. Because the line is so narrow, we can measure energies or frequencies very precisely. In 1959 and 1960, Pound and Rebka [13, 14] realized they could use this to measure gravitational redshift. They wanted of course to have light traverse as much vertical distance as possible; in the physics department at Harvard, the best option was a height of $h = 22.6\,\mathrm{m}$ in the stairwell. From Eq. (10.10) with $g = GM/r^2$, the fractional change in frequency as the light travels upward is

$$\begin{aligned}\left(\frac{\Delta\nu}{\nu}\right)_{\mathrm{up}} &= -\frac{GM_{\oplus}h}{c^2R_{\oplus}^2} \\ &= -\frac{(6.67\times10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}})\times(5.97\times10^{24}\,\mathrm{kg})\times(22.6\,\mathrm{m})}{(3.0\times10^{8}\,\mathrm{m\,s^{-1}})^2\times(6.38\times10^{6}\,\mathrm{m})^2} \\ &= -2.5\times10^{-15}\end{aligned}$$

As the light descends there would be a shift that has the same amplitude but the opposite sign. Pound and Rebka measured the combination of the upward and downward shifts as a way to remove any non-gravitational effects, finding

$$\left(\frac{\Delta\nu}{\nu}\right)_{\mathrm{down}} - \left(\frac{\Delta\nu}{\nu}\right)_{\mathrm{up}} = (5.1\pm0.5)\times10^{-15}$$

It seems astounding that we can measure relativistic effects to a few parts in 10^{15}. More recent experiments are even more precise.

10.4.4 Gravitational Redshift from a White Dwarf (1971)

The gravitational redshift on Earth is small because Earth's gravity is weak. Even on the Sun the effect is small: $\Delta\nu/\nu = -2\times10^{-6}$ between the surface and a point at infinity. To get a larger shift we need an object that is massive but compact. A white dwarf star is typically about as massive as the Sun but only as large as Earth (see Sect. 17.2). The nearest white dwarf is Sirius B, so named because it is in a binary with the bright star Sirius. In 1971, Greenstein et al. [15] managed to measure the gravitational redshift of light from Sirius B. (See Hetherington [16] for more about the history.) Greenstein et al. analyzed the spectrum of Sirius B to infer that the star has a radius of

$$R = 0.0078\,R_{\odot} = 5.42\times10^{6}\,\mathrm{m} = 0.85\,R_{\oplus}$$

and a surface gravity of

$$g = 4.47\times10^{6}\,\mathrm{m\,s^{-2}}$$

Together, these imply a mass of

$$M = \frac{gR^2}{G} = 1.97 \times 10^{30}\,\mathrm{kg} = 0.99\,M_\odot$$

The gravitational redshift of a photon leaving the surface and traveling to infinity, expressed in terms of wavelength, is

$$\frac{\lambda(\infty)}{\lambda(r)} = \frac{\nu(r)}{\nu(\infty)} \approx 1 + \frac{GM}{c^2 r}$$

What we measure is the shift in wavelength,

$$z \equiv \frac{\lambda(\infty)}{\lambda(r)} - 1 \approx \frac{GM}{c^2 r} \approx \frac{gR}{c^2}$$

Given the properties of Sirius B, we predict a redshift of

$$z_{\mathrm{predicted}} \approx \frac{(4.47 \times 10^6\,\mathrm{m\,s^{-2}}) \times (5.42 \times 10^6\,\mathrm{m})}{(3.0 \times 10^8\,\mathrm{m\,s^{-1}})^2} \approx 2.7 \times 10^{-4}$$

The measured shifts in the spectral lines were

$$z_{\mathrm{measured}} = (3.0 \pm 0.5) \times 10^{-4}$$

10.4.5 Flying Clocks (1971)

In October 1971, Joseph Hafele and Richard Keating [6, 7] flew atomic clocks on airplanes around the Earth. Airborne clocks experience time dilation (relative to surface clocks) for two reasons: motion, because airplanes move at different speeds than the surface of Earth; and gravity, because gravity is a little weaker at the altitudes where planes fly.

Let's consider the motion first. Both the airplane and the surface of Earth follow curved trajectories, so strictly speaking they are not inertial reference frames, but we will still use the special relativistic expression (10.5) to estimate the time dilation due to motion. We will, however, reference our measurements to the center of Earth so we can treat the surface and airplane on equal footing. Earth's rotation causes a clock at the equator to have a speed relative to the center of Earth of

$$v_S = \frac{2\pi R_\oplus}{P_{\mathrm{rot}}} = \frac{2\pi \times (6.38 \times 10^6\,\mathrm{m})}{86{,}400\,\mathrm{s}} = 464\,\mathrm{m\,s^{-1}}$$

Suppose the airplane is flying east/west with speed v_A relative to Earth's center; then write

$$v_A = v_S + u$$

so u is the speed of the airplane relative to Earth's surface. Relative to Earth's center, the surface and airplane have relativistic factors

$$\gamma_S = \left(1 - \frac{v_S^2}{c^2}\right)^{-1/2} \quad \text{and} \quad \gamma_A = \left(1 - \frac{v_A^2}{c^2}\right)^{-1/2}$$

Let t_C be the duration of the airplane flight measured in the reference frame of Earth's center. Then the durations in the surface and airplane frames are

$$t'_S = \frac{t_C}{\gamma_S} \quad \text{and} \quad t'_A = \frac{t_C}{\gamma_A}$$

The difference between the time elapsed on the airplane and the time elapsed on the surface is

$$t'_A - t'_S = \frac{t_C}{\gamma_A} - t'_S = \left(\frac{\gamma_S}{\gamma_A} - 1\right) t'_S$$

The fractional change induced by the motion is

$$\left[\frac{t'_A - t'_S}{t'_S}\right]_{\text{motion}} = \left(\frac{1 - v_A^2/c^2}{1 - v_S^2/c^2}\right)^{1/2} - 1$$

Since the speeds are small compared with the speed of light, we can do a Taylor series expansion in v_S/c and v_A/c:

$$\left[\frac{t'_A - t'_S}{t'_S}\right]_{\text{motion}} \approx \left(1 - \frac{v_A^2}{2c^2}\right)\left(1 + \frac{v_S^2}{2c^2}\right) - 1 \approx \frac{v_S^2 - v_A^2}{2c^2}$$

Now we write $v_A = v_S + u$ and simplify:

$$\left[\frac{t'_A - t'_S}{t'_S}\right]_{\text{motion}} \approx \frac{v_S^2 - (v_S^2 + 2v_S u + u^2)}{2c^2} \approx -\frac{(2v_S + u)u}{2c^2}$$

To get some specific numbers, let's suppose the clocks flew on the Concorde, which used to reach a groundspeed of about 650 m s^{-1}. Then we find:

$$\text{eastbound,} \quad u = +650\,\text{m s}^{-1} : \quad \left[\frac{t'_A - t'_S}{t'_S}\right]_{\text{motion}} \approx -5.7 \times 10^{-12} \approx -490\ \text{ns/day}$$

$$\text{westbound,} \quad u = -650\,\text{m s}^{-1} : \quad \left[\frac{t'_A - t'_S}{t'_S}\right]_{\text{motion}} \approx 1.0 \times 10^{-12} \approx +90\ \text{ns/day}$$

Note that the time dilation due to motion depends on the direction in which the airborne clock flies. This is an important part of the relativistic prediction.

Now let's consider the effect of gravity. Using Eq. (10.11) but making a Taylor series expansion, we can write the gravitational time dilation as

$$\frac{t(r+h)}{t(r)} = \frac{t(r+h)/t_\infty}{t(r)/t_\infty} \approx \frac{1 - GM/c^2(r+h)}{1 - GM/c^2 r}$$
$$\approx 1 - \frac{GM}{c^2(r+h)} + \frac{GM}{c^2 r} \approx 1 + \frac{GMh}{c^2 r(r+h)}$$

To this point we have only assumed that r and $r+h$ are large compared with GM/c^2. For this experiment we can do an additional Taylor series expansion with $h \ll r$:

$$\frac{t(r+h)}{t(r)} \approx 1 + \frac{GMh}{c^2 r^2}$$

Then we identify $t(r) = t'_S$ with the surface and $t(r+h) = t'_A$ with the airplane, so we can write the time dilation induced by gravity as

$$\left[\frac{t'_A - t'_S}{t'_S}\right]_{\text{gravity}} \approx \frac{GMh}{c^2 r^2}$$

The Concorde flew at an altitude of about 20 km, so the gravitational time shift is

$$\left[\frac{t'_A - t'_S}{t'_S}\right]_{\text{gravity}} \approx 2.2 \times 10^{-12} \approx +190 \text{ ns/day}$$

This is the same for airplanes moving both east and west.

We have considered an idealized experiment (daylong Concorde flights over the equator) that captures the main ideas, but Hafele and Keating analyzed the actual flight paths. They found the following time shifts (measured in nanoseconds):

	Motion	Gravity	Net prediction	Measurement
Eastbound	-184 ± 18	144 ± 14	-40 ± 23	-59 ± 10
Westbound	96 ± 10	179 ± 18	275 ± 21	273 ± 7

(The uncertainties in the predictions include uncertainties in the flight parameters.) Notice that the east- and westbound flights have motion shifts with different signs, as we discussed. Also, they have different gravity shifts, presumably because the flights had different altitudes and/or durations. The key result is that the measurements confirm the predictions; relativistic time dilation can be measured in a controlled experiment.

10.4.6 Global Positioning System (1989)

Since 1989 we have had a widespread example of flying clocks: the Global Positioning System. GPS receivers take the time received from a satellite, compare it with the time on Earth, and use the difference (along with the known speed of light) to determine the distance to the satellite. Measuring distances to multiple satellites makes it possible to triangulate a position on Earth to high precision. The entire system rests on careful coordination between satellite and surface clocks, but relativity says they tick at different rates. Relativistic effects must therefore be taken into account for GPS to work. Let's estimate the size of those effects (see [17] for a more detailed discussion).

Each GPS satellite orbits about $h = 20{,}000\,\text{km} = 2 \times 10^7$ m above the surface of Earth. Its orbital speed is therefore

$$v = \left(\frac{GM}{r+h}\right)^{1/2} = 3.9 \times 10^3\,\text{m}\,\text{s}^{-1}$$

Each GPS satellite is moving faster than the surface of the Earth, so there is time dilation due to motion[6]:

$$\left[\frac{t'_A - t'_S}{t'_S}\right]_{\text{motion}} \approx \frac{v_S^2 - v_A^2}{2c^2} \approx -8.3 \times 10^{-11} \approx -7.2\,\mu\text{s/day}$$

where we again use $v_S = 464$ m/s as the velocity of the surface of Earth. There is also time dilation due to gravity (note that we no longer have $h \ll r$):

$$\left[\frac{t'_A - t'_S}{t'_S}\right]_{\text{gravity}} \approx \frac{GMh}{c^2 r(r+h)} \approx 5.3 \times 10^{-10} \approx 45.6\,\mu\text{s/day}$$

The net effect is that GPS satellites gain about 38 μs per day relative to clocks on the ground. If this difference were not taken into account, the time it takes the signal to travel from the satellite would be calculated incorrectly, so the distance to the satellite would be wrong, and the triangulation would be thrown off. How badly? After 1 day the time error would be $\Delta t = 38\,\mu\text{s}$, which would translate into an error in the distance to each satellite of $\Delta \ell = c\,\Delta t = 11$ km. This is not precisely the same as the error that a GPS receiver would make when triangulating from multiple satellites, but it does give a sense of the magnitude of the effect.

GPS is successful because the engineers who designed the system used the anticipated orbits to build clocks that would compensate for most of the relativistic effects. Also, each GPS receiver has a computer that performs relativistic calculations to determine additional corrections. Impressive, eh? General relativity at work!

[6]For comparison with Sect. 10.4.5, we retain the subscripts "A" for airplane and "S" for surface, respectively, even though the airplane is now a satellite.

10.5 Mathematics of Relativity

Now we turn to the mathematical framework of relativity. While we will not delve into all of the details, we want to get to the point where we can analyze motion around a black hole.[7]

10.5.1 Spacetime Interval

Fundamentally, relativity is about the geometry of spacetime. How do we quantify geometry? The first step is to measure the distance between points. In familiar Euclidean geometry, if we have two points

$$(x, y, z) \quad \text{and} \quad (x + \mathrm{d}x, y + \mathrm{d}y, z + \mathrm{d}z)$$

then we define

$$\mathrm{d}\ell^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$$

and say that $\mathrm{d}\ell$ is the distance between the two points. If we have a curve, we imagine breaking it into a series of small segments, computing $\mathrm{d}\ell$ for each segment, and adding them up (by integrating).

The distance $\mathrm{d}\ell$ is the same in all coordinate systems—it is **invariant**. We can rotate or translate the coordinate system any way we like and still get the same distance between the points.

In the spacetime of special relativity, we add time to the mix by defining

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - (\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2) \tag{10.14}$$

This is the "distance" between two points in spacetime, which we call the **spacetime interval**. The expression for $\mathrm{d}s^2$ is known as the **metric** because it specifies how we "measure" intervals. Notice that space and time both enter the metric but with different signs. A key property of the spacetime interval is that it is invariant under the Lorentz transformation, so it is a good tool for characterizing the geometry of spacetime in special relativity.

In Euclidean geometry $\mathrm{d}\ell^2$ is non-negative. In special relativity, by contrast, $\mathrm{d}s^2$ can be positive, negative, or zero.[8] As we will see below, a light ray has $\mathrm{d}s^2 = 0$;

---

[7]Many books do give more details; *A First Course in General Relativity* by Bernard Schutz [2] is a good example.

[8]It is tempting to think that $\mathrm{d}s$ is a real-valued quantity such that $\mathrm{d}s^2$ must be non-negative. In relativity, $\mathrm{d}s^2$ is the quantity we work with, and it can be positive, zero, or negative. We may write $\sqrt{\mathrm{d}s^2}$ (see below), but we do not write $\mathrm{d}s$ by itself.

we call this a *lightlike interval*. For a clock sitting at a fixed position, the spacetime interval between any two ticks has $\mathrm{d}s^2 = c^2\,\mathrm{d}\tau^2$ where τ is the proper time; therefore we say that any positive spacetime interval is *timelike*. Conversely, for a ruler the spacetime interval between the two ends at any given time has $\mathrm{d}s^2 = -\mathrm{d}L^2$ where L is the proper length; therefore we say that any negative spacetime interval is *spacelike*. To summarize:

$$\mathrm{d}s^2 = \begin{cases} c^2\,\mathrm{d}\tau^2 > 0 & \text{timelike} \\ 0 & \text{lightlike} \\ -\mathrm{d}L^2 < 0 & \text{spacelike} \end{cases} \tag{10.15}$$

So far we have worked in Cartesian coordinates. Since many astrophysical objects are (approximately) spherical, it is good to be able to work in spherical coordinates (r, θ, ϕ) as well. In Euclidean geometry, the distance between nearby points in spherical coordinates can be written as

$$\mathrm{d}\ell^2 = \mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2$$

The extension to the spacetime interval of special relativity just adds time:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - (\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2) \tag{10.16}$$

We will see variants of the spatial piece several times in this chapter and the next.

Example: Straight Line

To help understand the spacetime interval, consider a light ray moving in a straight line. Suppose it moves along a line parallel to the x-axis but offset in the z-direction by an amount b. The Cartesian spacetime coordinates can be written

$$(t, x, y, z) = (t, ct, 0, b)$$

The spacetime interval for the light ray is the

$$\mathrm{d}s^2 \;=\; c^2\,\mathrm{d}t^2 - c^2\,\mathrm{d}t^2 \;=\; 0$$

This is a lightlike interval, as it should be.

Now consider spherical coordinates. Converting from (t, x, y, z) to (t, r, θ, ϕ) yields

$$r = \left(b^2 + c^2t^2\right)^{1/2} \qquad \theta = \tan^{-1}\left(\frac{ct}{b}\right) \qquad \phi = 0$$

which implies

$$\mathrm{d}r = \frac{ct}{(b^2 + c^2t^2)^{1/2}}\, c\,\mathrm{d}t \qquad \mathrm{d}\theta = \frac{b}{b^2 + c^2t^2}\, c\,\mathrm{d}t \qquad \mathrm{d}\phi = 0$$

The spacetime interval in spherical coordinates is then

$$\begin{aligned}\mathrm{d}s^2 &= c^2\,\mathrm{d}t^2 - \mathrm{d}r^2 - r^2\,\mathrm{d}\theta^2 \\ &= c^2\,\mathrm{d}t^2 - \frac{c^2t^2}{b^2 + c^2t^2}\, c^2\,\mathrm{d}t^2 - (b^2 + c^2t^2)\, \frac{b^2}{(b^2 + c^2t^2)^2} c^2\,\mathrm{d}t^2 \\ &= c^2\,\mathrm{d}t^2 - \left(\frac{c^2t^2}{b^2 + c^2t^2} + \frac{b^2}{b^2 + c^2t^2}\right) c^2\,\mathrm{d}t^2 \\ &= 0\end{aligned}$$

While spherical coordinates are less natural for this problem than Cartesian coordinates, they yield the same result. They will be more natural when we study black holes.

10.5.2 4-Vectors

We need to introduce vectors describing motion in four-dimensional spacetime. Let

$$\mathbf{X} = (ct, \mathbf{x}) \tag{10.17}$$

be a 4-d position vector that includes the time coordinate (with a factor of c so all components have dimensions of length). To compute the spacetime interval, we need to introduce a tensor that characterizes the metric. In special relativity, the tensor has the form

$$\mathsf{g} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{10.18}$$

Then we can write the spacetime interval as

$$\mathrm{d}s^2 = \sum_{\mu,\nu=1}^{4} g_{\mu\nu}\,\mathrm{d}X_\mu\,\mathrm{d}X_\nu$$

More generally, we use the tensor to define the dot product of any two 4-vectors:

$$\mathbf{U}\cdot\mathbf{V} = \sum_{\mu,\nu} g_{\mu\nu}\, U_\mu\, V_\nu \tag{10.19}$$

Given **X**, we can define the associated 4-velocity to be

$$\mathbf{V} = \frac{\mathrm{d}\mathbf{X}}{\mathrm{d}\tau} \tag{10.20}$$

Since this is defined using proper time, the spatial part of **V** is not the same as the measured velocity $\mathbf{v} = \mathrm{d}\mathbf{x}/\mathrm{d}t$. In a frame where the particle moves with measured velocity **v**, time dilation says the measured time is[9]

$$t = \gamma_v \tau \quad \text{where} \quad \gamma_v = \left(1 - \frac{|\mathbf{v}|^2}{c^2}\right)^{-1/2} \tag{10.21}$$

Therefore the 4-velocity can be written in terms of the measured velocity as

$$\mathbf{V} = (\gamma_v c, \gamma_v \mathbf{v}) \tag{10.22}$$

Why do we define 4-velocity in this way? We know that **X** transforms by the Lorentz transformation; then since τ is invariant, we realize that **V** must follow the Lorentz transformation as well. This clarifies the relation between reference frames[10]:

$$V_t = \gamma_u V_t' + \gamma_u \beta_u V_x' \tag{10.23a}$$
$$V_x = \gamma_u V_x' + \gamma_u \beta_u V_t' \tag{10.23b}$$
$$V_y = V_y' \tag{10.23c}$$
$$V_z = V_z' \tag{10.23d}$$

Our definition does mean that it takes a few extra steps to relate the measured velocities in different frames. In the primed frame, we can use Eq. (10.22) to write the 4-velocity in terms of the components of the measured velocity:

$$\mathbf{V}' = \frac{\{c, v_x', v_y', v_z'\}}{[1 - (v')^2/c^2]^{1/2}}$$

We can then use Eq. (10.23) to find the 4-velocity in the unprimed frame:

$$\mathbf{V} = \frac{\{\gamma_u(c + \beta_u v_x'), \gamma_u(v_x' + \beta_u c), v_y', v_z'\}}{[1 - (v')^2/c^2]^{1/2}}$$

---

[9] We put a subscript v on this γ to indicate that it is defined in terms of the particle's velocity and is not necessarily the same as the γ factor between arbitrary inertial frames (defined in Eq. 10.2).

[10] We put a subscript u on γ and β here to distinguish these factors, which relate arbitrary inertial frames, from γ_v in Eq. (10.21), which relates an arbitrary inertial frame to the particle's rest frame.

Inverting the relation between **v** and the spatial components of **V** gives

$$\{v_x, v_y, v_z\} = \frac{\{V_x, V_y, V_z\}}{[1 + (V_x^2 + V_y^2 + V_z^2)/c^2]^{1/2}}$$

Plugging in and simplifying yields

$$\begin{aligned} v_x &= \frac{v'_x + u}{1 + uv'_x/c^2} \\ v_y &= \frac{v'_y}{\gamma_u(1 + uv'_x/c^2)} \\ v_z &= \frac{v'_z}{\gamma_u(1 + uv'_x/c^2)} \end{aligned}$$

This is the same transformation we found by a different approach in Eq. (10.4).

10.5.3 Relativistic Momentum and Energy

We also need to generalize the concepts of energy and momentum. We define the 4-momentum to be

$$\mathbf{P} = m\mathbf{V} = (\gamma_v mc, \gamma_v m\mathbf{v}) \tag{10.24}$$

We then take the relativistic versions of energy and momentum to be the time and space parts of **P**, respectively:

$$\mathbf{P} = \left(\frac{E}{c}, \mathbf{p}\right) \tag{10.25}$$

(The factor of c is included so E has dimensions of energy.) To understand what E represents, consider the dot product of **P** with itself. Using Eq. (10.24) along with the definition of the dot product in Eq. (10.19), we have

$$\mathbf{P}\cdot\mathbf{P} = \gamma_v^2 m^2c^2 - \gamma_v^2 m^2 v^2 = \gamma_v^2\left(1 - \frac{v^2}{c^2}\right)m^2c^2 = m^2c^2$$

where we use Eq. (10.21) to simplify. If instead we computed the dot product using Eq. (10.25), we would find

$$\mathbf{P}\cdot\mathbf{P} = \left(\frac{E}{c}\right)^2 - p^2$$

In other words, the relativistic energy and momentum are related by

$$\left(\frac{E}{c}\right)^2 - p^2 = m^2c^2 \quad \Rightarrow \quad E^2 = p^2c^2 + m^2c^4 \tag{10.26}$$

In the particle's rest frame, $p = 0$ so we recover the famous relation $E = mc^2$ for the rest mass energy. In the non-relativistic limit, $p \ll mc$ so we can make a Taylor series expansion:

$$E = mc^2\left(1 + \frac{p^2}{m^2c^2}\right)^{1/2} \approx mc^2 + \frac{p^2}{2m}$$

The first term is the rest-mass energy, while the second term is the Newtonian kinetic energy. The bottom line is that we can interpret E as the total energy in relativity.

There is one more useful relation we can derive. Again combining Eqs. (10.24) and (10.25), we can write

$$\frac{p}{E} = \frac{v}{c^2}$$

Using Eq. (10.26) to rewrite E yields

$$v = \frac{pc^2}{(p^2c^2 + m^2c^4)^{1/2}} \tag{10.27}$$

This is the relativistic version of the relation between momentum and velocity. In the non-relativistic limit, $p \ll mc$ so Eq. (10.27) reduces to the familiar relation $v \approx p/m$.

10.6 Black Holes

To this point we have discussed situations in which gravity is "weak" and we can make Taylor series expansions. We now move into the regime of "strong" gravity and examine a surprising and bizarre prediction of general relativity: black holes. While we are particularly interested in the strange physics near a black hole's event horizon, our analysis actually applies outside any spherical object in GR.

10.6.1 Schwarzschild Metric

To begin, we need to specify the spacetime geometry through the metric. To understand the form of the metric, recall from Eq. (10.11) the expression for gravitational time dilation,

$$\frac{\Delta t(r)}{\Delta t(\infty)} = \left(1 - \frac{2GM}{c^2 r}\right)^{1/2}$$

Presumably this factor appears in the time term of the metric. It also appears in the space term (basically from the curvature we discussed in Sect. 10.3.3). The full metric outside any spherical object of mass M is

$$ds^2 = \left(1 - \frac{2GM}{c^2 r}\right) c^2\,dt^2 - \left(1 - \frac{2GM}{c^2 r}\right)^{-1} dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta\, d\phi^2 \quad (10.28)$$

This is called the **Schwarzschild metric** after the German mathematician Karl Schwarzschild, who discovered it as a solution of the equations of Einstein's general theory of relativity.[11]

We think of the coordinates (t, r, θ, ϕ) as quantities that would be measured by an observer far from the object, and we refer to them as "coordinate time," "coordinate radius," etc. They are different from quantities measured by an observer near the object; understanding the difference is one of our goals.

Notice that something funny happens to the metric when r approaches the **Schwarzschild radius** $R_S = 2GM/c^2$: the time term vanishes, while the radial term diverges. In the early twentieth century, all known astrophysical objects had sizes $R \gg R_S$, so Einstein and other prominent figures such as Arthur Eddington assumed the weirdness was merely a mathematical curiosity, not a physical reality. It was only later, after Subramanyan Chandrasekhar and Robert Oppenheimer showed that stars could collapse to become comparable to or even smaller than the Schwarzschild radius, that physicists began to take the strange predictions seriously.

The Schwarzschild metric deviates from the flat spacetime from special relativity (Eq. 10.16) only to the extent that R_S/r is nonzero. This allows us, finally, to specify what we mean by "weak" or "strong" gravity:

$$r \gg R_S \quad \rightarrow \quad \frac{R_S}{r} \ll 1 \quad \rightarrow \quad \text{“weak field”}$$

$$r \sim R_S \quad \rightarrow \quad \frac{R_S}{r} \sim 1 \quad \rightarrow \quad \text{“strong field”}$$

[11] Historical aside (drawn from *Black Holes and Time Warps* by Kip Thorne [18]): When Einstein's general theory of relativity was published on Nov. 25, 1915, Schwarzschild was serving in the German army on the Russian front in the first World War. He managed to obtain Einstein's paper, read it, apply it to stars, discover a solution to the complicated equations Einstein had derived, write a paper of his own, and send it to Einstein—all in time for Einstein to present the paper on Schwarzschild's behalf at a meeting on Jan. 13, 1916. Unfortunately, Schwarzschild died on May 11 of illness contracted during his service.

To give some examples, let's quantify the Schwarzschild radius:

$$\begin{aligned} R_S &= \frac{2GM_\odot}{c^2} \times \frac{M}{M_\odot} \\ &= \frac{2 \times (6.67 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}) \times (1.99 \times 10^{30}\,\mathrm{kg})}{(3.0 \times 10^8\,\mathrm{m\,s^{-1}})^2} \times \left(\frac{M}{M_\odot}\right) \\ &= 3\,\mathrm{km} \times \left(\frac{M}{M_\odot}\right) \end{aligned} \tag{10.29}$$

We retain the mass dependence but express M in solar masses so we can quickly evaluate the Schwarzschild radius for different astrophysical objects. Here are typical numbers for some systems we have studied already or will encounter:

	$M/M_\odot$	R_S	R	R_S/R
Earth	3×10^{-6}	0.009 m	6.4×10^6 m	1.4×10^{-9}
Sun	1	3 km	7×10^8 m	4×10^{-6}
White dwarf	1	3 km	6×10^6 m	5×10^{-4}
Neutron star	1.4	4 km	10 km	0.4

10.6.2 Spacetime Geometry

To begin to see some of the weird properties of a black hole, consider the spacetime interval between ticks on a stationary clock. If the clock does not move then $\mathrm{d}r = \mathrm{d}\theta = \mathrm{d}\phi = 0$, so the spacetime interval is

$$\mathrm{d}s^2 = \left(1 - \frac{R_S}{r}\right) c^2\,\mathrm{d}t^2 = \begin{cases} > 0 \text{ (timelike)} & \text{for } r > R_S \\ < 0 \text{ (spacelike)} & \text{for } r < R_S \end{cases}$$

The spacetime interval *changes sign* at the Schwarzschild radius, switching from spacelike to timelike. This is important because no physical object can experience a spacelike interval; to do so, it would have to move faster than the speed of light. We seem to have a paradox: a stationary clock inside the Schwarzschild radius would have a spacelike interval, which is not allowed. To resolve the paradox, we conclude that *it is impossible to remain stationary inside the Schwarzschild radius*. In fact, objects inside the Schwarzschild radius are inexorably drawn to the central singularity, just as on Earth we are inexorably drawn forward in time.

10.6.3 Particle in a Circular Orbit

As we set out to study motion in general relativity, it is good to start with the simple case of circular orbits. Such an orbit stays in a plane, and we can choose our coordinates so this is the equatorial plane:

$$r = \text{constant} \qquad \theta = \frac{\pi}{2} \qquad \phi = \omega\, t$$

where ω is the coordinate angular speed. The period of the orbit in coordinate time is $P = 2\pi/\omega$. The spacetime interval for the orbit is:

$$\mathrm{d}s^2 = \left(1 - \frac{R_S}{r}\right) c^2\, \mathrm{d}t^2 - r^2\omega^2\, \mathrm{d}t^2 = \left(1 - \frac{R_S}{r} - \frac{r^2\omega^2}{c^2}\right) c^2\, \mathrm{d}t^2$$

With this we can determine the proper time (see Eq. 10.15):

$$\tau_{\text{circ}} = \frac{1}{c}\int_{\text{one orbit}} \sqrt{\mathrm{d}s^2} = \left(1 - \frac{R_S}{r} - \frac{r^2\omega^2}{c^2}\right)^{1/2} P$$

This is the time that would be measured on a clock that is executing the circular orbit. Note that it is not a simple integral over $\mathrm{d}t$; we must account for the motion using the spacetime interval.

We have not yet specified the radius. We can find it by applying Fermat's principle of least time: for a given angular speed, the particle will "choose" the radius that minimizes the proper time. Operationally, we want to find the radius that minimizes τ, so we want to solve

$$0 = \frac{\mathrm{d}\tau}{\mathrm{d}r} = \frac{1}{2}\left(1 - \frac{R_S}{r} - \frac{r^2\omega^2}{c^2}\right)^{-1/2} \left(\frac{R_S}{r^2} - \frac{2r\omega^2}{c^2}\right) P$$

The solution is

$$r = \left(\frac{c^2 R_S}{2\omega^2}\right)^{1/3}$$

It is more convenient to write the relation as

$$\omega = \left(\frac{c^2 R_S}{2r^3}\right)^{1/2} = \left(\frac{GM}{r^3}\right)^{1/2}$$

The coordinate velocity is then

$$v = \omega r = \left(\frac{GM}{r}\right)^{1/2}$$

This is the same expression we had in Newtonian gravity (see Eq. 7.7). In other words, a distant observer would measure the same orbital size and orbital velocity, and hence the same orbital period, as in Newtonian gravity. But a clock following the circular orbit would measure the proper time, which is different:

$$\tau_{\rm circ} = \left(1 - \frac{R_S}{r} - \frac{r^2\omega^2}{c^2}\right)^{1/2} P = \left(1 - \frac{3R_S}{2r}\right)^{1/2} P$$

Out of curiosity, what about a clock *at rest* at the same radius? Such a clock has $dr = d\phi = d\theta = 0$ and hence

$$ds^2 = \left(1 - \frac{R_S}{r}\right) c^2\, dt^2 \quad \Rightarrow \quad \tau_{\rm rest} = \frac{1}{c}\int \sqrt{ds^2} = \left(1 - \frac{R_S}{r}\right)^{1/2} P$$

This is identical to the gravitational time dilation for a clock at rest in a gravitational field that we examined in Sect. 10.3.4.

We see that time is complicated! The time you measure depends on where you are and how you are moving. These are both effects that we have seen already (time dilation in special and general relativity), but it is interesting to see how they manifest themselves here.

Example: Circular Orbit Around Sgr A*

Imagine we were in a spaceship orbiting the black hole at the center of the Milky Way at $r = 3\,R_S$. If we take $M_{\rm bh} = 4 \times 10^6\, M_\odot$ then from Eq. (10.29) the Schwarzschild radius is $R_S = 1.18 \times 10^{10}$ m, and so the radius of the orbit is $r = 3R_S = 3.54 \times 10^{10}$ m. The orbital period as measured by a distant observer (i.e., in coordinate time) is the same as in Newtonian gravity:

$$P = \frac{2\pi}{\omega} = 2\pi \left(\frac{r^3}{GM}\right)^{1/2} = 1{,}800\,{\rm s} = 30\,{\rm min}$$

However, our clocks on the spaceship show the proper time, and in our frame one orbital period takes

$$\tau_{\rm circ} = \left(1 - \frac{3R_S}{2r}\right)^{1/2} P = 1{,}290\,{\rm s} = 21\,{\rm min}$$

If we had friends in a space station that is sitting at a fixed spot with $r = 3\,R_S$ (i.e., not orbiting but stationary), they would measure our orbital period as

$$\tau_{\rm rest} = \left(1 - \frac{R_S}{r}\right)^{1/2} P = 1{,}490\,{\rm s} = 25\,{\rm min}$$

Again, time depends on where you are and how you are moving.

10.6.4 General Motion Around a Black Hole

Now we allow general motion.[12] Let's briefly review the Newtonian case as a point of reference. As we saw in Sect. 3.1, in spherical symmetry the motion is confined to a plane, which we can define to be the equatorial plane. The equation of motion for the one-body problem is then

$$\left[\frac{d^2r}{d\tau^2} - r\left(\frac{d\phi}{d\tau}\right)^2\right]\hat{\mathbf{r}} + \frac{1}{r}\frac{d}{d\tau}\left(r^2\frac{d\phi}{d\tau}\right)\hat{\boldsymbol{\phi}} = -\frac{GM}{r^2}\,\hat{\mathbf{r}}$$

(In general relativity, the natural time coordinate for studying motion is the proper time, so we write τ here.) The angular component of the equation of motion implies

$$r^2\,\frac{d\phi}{d\tau} = \text{constant} \equiv \ell \tag{10.30}$$

where ℓ is the specific angular momentum. This is conservation of angular momentum (which we have seen many times now). The radial component of the equation of motion looks like

$$\frac{d^2r}{d\tau^2} - \frac{\ell^2}{r^3} = -\frac{GM}{r^2}$$

We rewrite this as

$$\frac{d^2r}{d\tau^2} = -\frac{d\Phi_{\rm Newt}}{dr} \tag{10.31}$$

where we define the effective potential

$$\Phi_{\rm Newt} = -\frac{GM}{r} + \frac{\ell^2}{2r^2} + \frac{c^2}{2} \tag{10.32}$$

The first term is the familiar Newtonian gravitational potential. The second term is the centrifugal term. The last term is just a constant that we add because it will prove to be convenient in the relativistic case.

The effective potential is useful because we can think of it as a surface and use our intuition to understand what would happen to a ball on that surface. Some examples are shown by the dashed curves in Fig. 10.3. For $\ell = 0$ the ball would roll all the way down to $r = 0$. For any nonzero value of ℓ, however, the centrifugal term causes an upturn at small radius. This creates a stationary point that corresponds to a constant radius and hence a circular orbit. In the Newtonian effective potential, the

[12] This presentation draws from the book by Schutz [2].

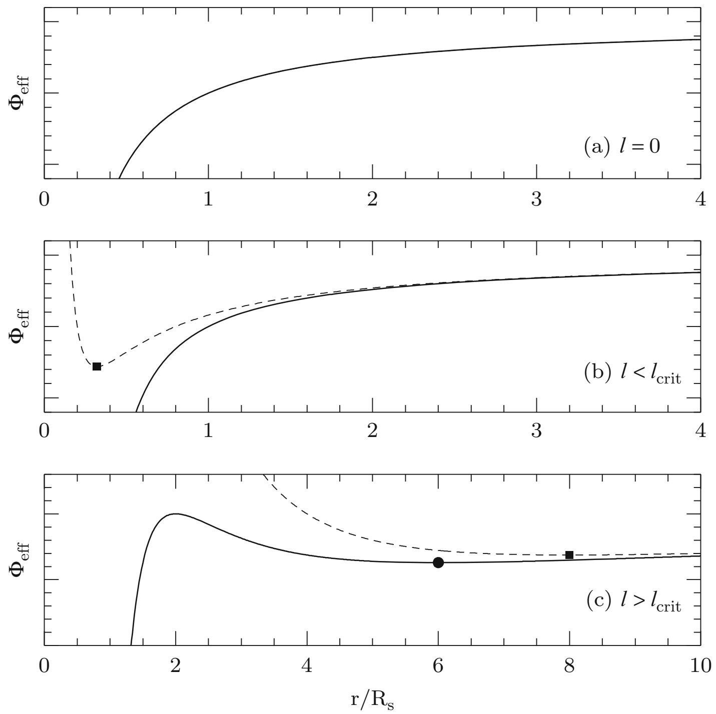

Fig. 10.3 Examples of effective potentials for a massive particle in Newtonian gravity (*dashed*) and GR (*solid*). The three panels correspond to different values of the specific angular momentum, ℓ. Note that the *bottom panel* has a different horizontal scale. Points indicate local minima (*squares* for Newtonian gravity and *circles* for GR)

stationary point is a minimum of Φ, so the orbit is stable: if you put the ball near the minimum but give it a little kick, it will oscillate around the minimum but remain confined.

In general relativity the equation of motion can be written in the form of (10.31) but with a different effective potential. To find that potential, recall that special relativity has a relation between energy, momentum, and mass: $(E/c)^2 - p^2 = m^2$ (Eq. 10.26). With the Schwarzschild metric the analogous relation has factors of $(1 - R_S/r)$:

$$\left(1 - \frac{R_S}{r}\right)^{-1} \left(\frac{m\tilde{E}}{c}\right)^2 - \left(1 - \frac{R_S}{r}\right)^{-1} \left(m\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 - \frac{m^2\ell^2}{r^2} = m^2c^2 \qquad (10.33)$$

where ℓ is the specific angular momentum and $\tilde{E} = E/m$ is the energy per unit mass, both of which are well defined only if the moving particle has a nonzero rest mass. (We consider a massless particle below.) We can divide through by m^2 and rearrange to write

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 = \frac{\tilde{E}^2}{c^2} - \left(1 - \frac{R_S}{r}\right)\left(c^2 + \frac{\ell^2}{r^2}\right) \tag{10.34}$$

Take the derivative $\mathrm{d}/\mathrm{d}\tau$, divide through by $2\,\mathrm{d}r/\mathrm{d}\tau$, and use $R_S = 2GM/c^2$:

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} = -\frac{GM}{r^2} + \frac{\ell^2}{r^3} - \frac{3GM\ell^2}{c^2 r^4}$$

The first two terms match the Newtonian case, but the third term is new in GR. We can capture all of the terms in the same form as Eq. (10.31) by introducing the effective potential

$$\text{(massive)} \qquad \Phi_{\mathrm{GR}} = -\frac{GM}{r} + \frac{\ell^2}{2r^2} + \frac{c^2}{2} - \frac{GM\ell^2}{c^2 r^3}$$
$$= \frac{1}{2}\left(1 - \frac{R_S}{r}\right)\left(c^2 + \frac{\ell^2}{r^2}\right)$$

If the particle is massless (e.g., a photon), the analysis is slightly different because we cannot define the energy and angular momentum per unit mass. Nevertheless, light does carry both energy and momentum, and we can keep the same form of the equations if we define ℓ and $\tilde{E}$ to be the total angular momentum and energy, respectively. Also, we need to be careful with the derivative term in Eq. (10.33) because τ and m are both zero for photons. We can, however, define a new parameter λ that runs along the photon's trajectory in spacetime such that the derivative $\mathrm{d}r/\mathrm{d}\lambda$ is well defined. The upshot is that Eq. (10.33) is replaced for a massless particle by

$$\left(1 - \frac{R_S}{r}\right)^{-1}\left(\frac{\tilde{E}}{c}\right)^2 - \left(1 - \frac{R_S}{r}\right)^{-1}\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 - \frac{\ell^2}{r^2} = 0 \tag{10.35}$$

or

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 = \frac{\tilde{E}^2}{c^2} - \left(1 - \frac{R_S}{r}\right)\frac{\ell^2}{r^2} \tag{10.36}$$

As before, we take the derivative $\mathrm{d}/\mathrm{d}\lambda$, divide through by $2\,\mathrm{d}r/\mathrm{d}\lambda$, and use $R_S = 2GM/c^2$:

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\lambda^2} = \frac{\ell^2}{r^3} - \frac{3GM\ell^2}{c^2 r^4}$$

In this case we define the effective potential to be

$$\text{(massless)} \qquad \Phi_{\mathrm{GR}} = \frac{1}{2}\left(1 - \frac{R_S}{r}\right)\frac{\ell^2}{r^2}$$

We can combine the expressions for the massive and massless cases into a single effective potential if we write

$$\Phi_{\rm GR} = \frac{1}{2}\left(1 - \frac{R_S}{r}\right)\left(\tilde{m}c^2 + \frac{\ell^2}{r^2}\right) \tag{10.37}$$

and put $\tilde{m} = 1$ for a massive particle and $\tilde{m} = 0$ for the massless case.

Sample GR potentials are shown by the solid curves in Fig. 10.3. There are several important points to make:

- At large radius, the new term $GM\ell^2/c^2r^3$ from GR is small, so Newtonian gravity is a good approximation. GR effects are significant only at small radii.
- For ℓ above some critical value $\ell_{\rm crit}$, there is a minimum in the potential curve, which corresponds to a stable circular orbit. (You can find $\ell_{\rm crit}$, along with the location of the stable circular orbit, in Problem 10.7.)
- For $\ell > \ell_{\rm crit}$, there is also a maximum in the GR potential curve. It corresponds to a second allowed circular orbit for a given angular momentum, but one that is unstable. This is new in GR.
- For $\ell < \ell_{\rm crit}$, there is no minimum in the potential curve, and hence no stable circular orbit. Thus, there exists some *smallest* stable circular orbit in GR. This is another new feature (Newtonian gravity allows arbitrarily small circular orbits).
- The GR potential turns over at small radius, so if a particle gets too close to the black hole it cannot help but fall in. The ability to fall all the way to $r = 0$ with finite angular momentum is yet another difference from Newtonian gravity.

To learn more about the motion, let's return to Eqs. (10.34) and (10.36). We can write both in the form

$$\frac{{\rm d}r}{{\rm d}\tau} = \pm\left(\frac{\tilde{E}^2}{c^2} - 2\Phi\right)^{1/2} \tag{10.38}$$

This is the key equation of motion in the radial direction. We still need one more ingredient: an equation of motion for time (since t depends on position and motion). This equation involves again involves the factor $(1 - R_S/r)$:

$$\frac{{\rm d}t}{{\rm d}\tau} = \left(1 - \frac{R_S}{r}\right)^{-1}\frac{\tilde{E}}{c^2} \tag{10.39}$$

If we want an equation for r in terms of coordinate time (as opposed to proper time), we can combine Eqs. (10.38) and (10.39) to obtain

$$\frac{{\rm d}r}{{\rm d}t} = \frac{{\rm d}r/{\rm d}\tau}{{\rm d}t/{\rm d}\tau} = \pm c\left(1 - \frac{R_S}{r}\right)\frac{(\tilde{E}^2 - 2c^2\Phi)^{1/2}}{\tilde{E}} \tag{10.40}$$

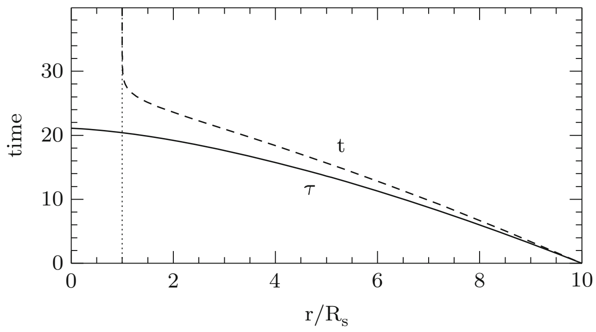

Fig. 10.4 The *dashed* and *solid lines* show how the coordinate time and proper time (respectively) flow as an object falls into a black hole. The horizontal axis is radius scaled by the Schwarzschild radius; the *dotted line* indicates the event horizon. The vertical axis is time scaled by R_S/c. In this example, time is defined to be 0 at $r/R_S = 10$ even though the particle started from rest at infinity

You can explore various aspects of motion in the Schwarzschild metric in Problems 10.6 and 10.7. Here let's consider a particle falling from rest into a black hole. We know the particle falls straight in, so one of the constants of motion is $\ell = 0$. If the particle starts from rest ($\mathrm{d}r/\mathrm{d}\tau = 0$) at infinity, then $\tilde{E} = c^2$. In Eq. (10.40) we choose the minus sign since we know radius must decrease with time. We can then rewrite the equation of motion as

$$c\,\mathrm{d}t = -\left(1 - \frac{R_S}{r}\right)^{-1}\left(\frac{r}{R_S}\right)^{1/2}\mathrm{d}r$$

Integrating both sides yields $t(r)$ as shown in Fig. 10.4. The curious result is that time goes to infinity as r approaches R_S. As seen by an observer far away, the particle never actually reaches the black hole!

This bizarre result only applies to the coordinate time. Repeating the analysis using the equation of motion (10.38) reveals that proper time—which is what you would see on your watch if you fell into a black hole—is perfectly well behaved. It *is* possible to fall into a black hole, but no one on the outside can see it.

10.6.5 Gravitational Deflection

We are now equipped to derive the relativistic deflection angle that we previously quoted in Sect. 9.1.1.[13] Consider the setup in Fig. 10.5, which is modified from Fig. 9.1 to have the particle move from right to left so the azimuthal angle ϕ

[13]This presentation follows the analysis given by Keeton and Petters [19].

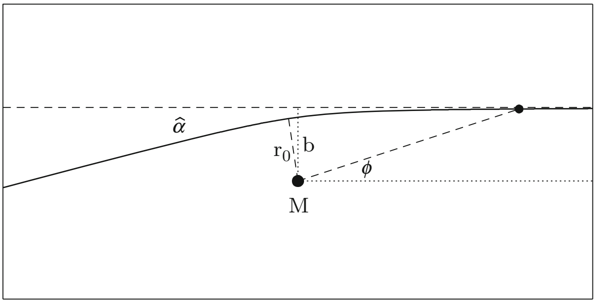

Fig. 10.5 Setup for calculating the bending angle $\hat{\alpha}$. The azimuthal angle ϕ is defined from the x-axis as usual, and the particle moves from right to left so ϕ increases monotonically from $\phi \to 0$ as $\tau \to -\infty$ to $\phi \to \pi + \hat{\alpha}$ as $\tau \to +\infty$. The impact parameter is b while the distance of closest approach is r_0

increases monotonically from an initial value of $\phi \to 0$ as $\tau \to -\infty$. If there were no deflection, ϕ would go to π for $\tau \to +\infty$; thus the angle of deflection is the amount by which $\Delta\phi$ exceeds π:

$$\hat{\alpha} = \Delta\phi - \pi$$

We need an equation of motion for ϕ. Since Eqs. (10.30) and (10.38) both have r on the right-hand side, it is useful to let r be the independent variable by writing

$$\frac{\mathrm{d}\phi}{\mathrm{d}r} = \frac{\mathrm{d}\phi/\mathrm{d}\tau}{\mathrm{d}r/\mathrm{d}\tau} = \pm\frac{\ell}{r^2}\left(\frac{\tilde{E}^2}{c^2} - 2\Phi\right)^{-1/2}$$

To solve this equation, we break the trajectory into two pieces. In the first half, r decreases from ∞ down to the point of closest approach r_0, and $\mathrm{d}\phi/\mathrm{d}r < 0$ so we take the minus sign. In the second half, r increases from r_0 out to ∞, and $\mathrm{d}\phi/\mathrm{d}r > 0$ so we take the plus sign. Then:

$$\begin{aligned}\hat{\alpha} &= -\int_{\infty}^{r_0}\frac{\ell}{r^2}\left(\frac{\tilde{E}^2}{c^2} - 2\Phi\right)^{-1/2}\mathrm{d}r + \int_{r_0}^{\infty}\frac{\ell}{r^2}\left(\frac{\tilde{E}^2}{c^2} - 2\Phi\right)^{-1/2}\mathrm{d}r - \pi \\ &= 2\int_{r_0}^{\infty}\frac{\ell}{r^2}\left(\frac{\tilde{E}^2}{c^2} - 2\Phi\right)^{-1/2}\mathrm{d}r - \pi \end{aligned} \tag{10.41}$$

Now consider the constants of motion. At very early times the trajectory is nearly a straight line with $x = -v\tau$ and $y = b$, so the polar coordinates are

$$r = \left(b^2 + v^2\tau^2\right)^{1/2} \quad \text{and} \quad \phi = -\tan^{-1}\left(\frac{b}{v\tau}\right)$$

and the constants of motion are

$$\ell = \lim_{r\to\infty}\left[r^2\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right] = bv$$

$$\frac{\tilde{E}^2}{c^2} = \lim_{r\to\infty}\left[\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + \left(1-\frac{R_S}{r}\right)\left(\tilde{m}c^2+\frac{\ell^2}{r^2}\right)\right] = v^2+\tilde{m}c^2$$

Thus we can rewrite Eq. (10.41) as

$$\hat{\alpha} = 2\int_{r_0}^{\infty}\frac{bv}{r^2}\left[\tilde{m}c^2+v^2-\left(1-\frac{R_S}{r}\right)\left(\tilde{m}c^2+\frac{b^2v^2}{r^2}\right)\right]^{-1/2}\mathrm{d}r - \pi \quad (10.42)$$

There is one final ingredient: we need to relate the impact parameter b to the distance of closest approach r_0. We use the fact that $\mathrm{d}r/\mathrm{d}\tau = 0$ at the point of closest approach to put $\tilde{E}^2/c^2 = 2\Phi(r_0)$ or

$$\tilde{m}c^2+v^2 = \left(1-\frac{R_S}{r_0}\right)\left(\tilde{m}c^2+\frac{b^2v^2}{r_0^2}\right)$$

which can be solved to find

$$b = \frac{r_0}{v}\left(\frac{v^2+\tilde{m}c^2R_S/r_0}{1-R_S/r_0}\right)^{1/2} \quad (10.43)$$

Plug this into Eq. (10.42) and make the change of variables $r = r_0/w$:

$$\hat{\alpha} = 2\int_0^1\left[\frac{s\tilde{m}c^2+v^2}{s\tilde{m}c^2w(1-w)[1-s(1+w)]+v^2[1-w^2-s(1-w^3)]}\right]^{1/2}\mathrm{d}w - \pi$$

where $s \equiv R_S/r_0$. This is the general expression for the deflection angle in the Schwarzschild metric, but it is not terribly enlightening. We can make more progress if the trajectory never gets very close to the central object. In that case $s \ll 1$ and we can expand the integrand as a Taylor series in s:

$$\begin{aligned}\hat{\alpha} &\approx \int_0^1\left[\frac{2}{\sqrt{1-w^2}}+\frac{\tilde{m}c^2+v^2(1+w+w^2)}{v^2(1+w)\sqrt{1-w^2}}s+\mathscr{O}\left(s^2\right)\right]\mathrm{d}w - \pi \\ &\approx \left(2+\frac{\tilde{m}c^2}{v^2}\right)s + \mathscr{O}\left(s^2\right)\end{aligned}$$

Recall that $s = R_S/r_0$, but from Eq. (10.43) we can replace this with $s \approx R_S/b \approx 2GM/c^2b$ at the order of approximation to which we are working. This yields

$$\hat{\alpha} \approx \left(2+\frac{\tilde{m}c^2}{v^2}\right)\frac{2GM}{c^2b} \quad (10.44)$$

In the case of a massless particle (e.g., a photon), $\tilde{m} = 0$ and so we have

$$\hat{\alpha} \approx \frac{4GM}{c^2 b} \qquad \text{(massless)}$$

which is the result that we used in Chap. 9 to build the theory of gravitational lensing. In the case of a massive particle $\tilde{m} = 1$, and if the particle is non-relativistic then $v/c \ll 1$ and the second term in parentheses in Eq. (10.44) dominates the first term to yield

$$\hat{\alpha} \approx \frac{2GM}{v^2 b} \qquad \text{(massive and non-relativistic)}$$

which is the same result that we derived using Newtonian gravity in Sect. 9.1.1.

10.7 Other Effects

Many other aspects of relativity lie beyond the scope of this book, but two are worth mentioning briefly. First, we have studied black holes that are static and spherically symmetric, but most objects in the universe rotate. Roy Kerr [20] found a solution to Einstein's equations that describes a rotating black hole. The spin modifies spacetime in the vicinity of the black hole, which affects the motion of any matter in an accretion disk and the properties of light emitted from the disk. Observations of spectral lines from black hole accretion disks can therefore be used to measure black hole spin and probe the Kerr metric (see [21] for a review).

Second, relativity predicts that accelerating masses create ripples in spacetime that propagate as **gravitational waves**. Gravitational radiation from accelerating masses is somewhat analogous to electromagnetic radiation from accelerating charges, although the analogy is not precise. The waves are predicted to carry energy away from a binary star system and cause the stars' orbits to decay. (You can explore this process in Problem 10.8.) The energy loss scales with orbital separation as $P \propto a^{-5}$, so it mainly affects close binaries. Two particular systems show clear evidence for orbital decay due to gravitational radiation: the "binary pulsar" PSR B1913+16, discovered in 1974 by Hulse and Taylor [22] (for which they received the 1993 Nobel Prize in Physics); and the "double pulsar" J0737−3039, discovered in 2003 by Burgay et al. [23]. These systems provide strong if indirect tests of predictions for gravitational radiation [24, 25]. The next goal is to detect gravitational waves directly. Projects such as the Laser Interferometer Gravitational-Wave Observatory as well as Virgo, AURIGA, and MiniGRAIL are trying to create new ways for us to observe the effects of strong gravity in extreme events throughout the universe.

Problems

10.1. If the Sun were replaced by a black hole with the same mass, would Earth's orbit change significantly? Why or why not?

10.2. The Michelson-Morley experiment showed that the speed of light does not depend on the speed of the source. Use the velocity transformation (10.4) to explain the result. Specifically:

(a) Suppose a source moving horizontally with speed u emits a light ray going in the horizontal direction. What is the speed and direction of the light ray as measured by a stationary observer?
(b) Suppose a source moving horizontally with speed u emits a light ray going in the vertical direction (in the source's reference frame). What is the speed and direction of the light ray as measured by a stationary observer?

10.3. Muons are elementary particles produced when cosmic rays collide with atoms in Earth's upper atmosphere. Muons are unstable and decay, so the number of muons as a function of time has the form $N(t) = N_0\, e^{-t/\tau}$ where $\tau = 2.20 \times 10^{-6}$ s and N_0 is the number at $t = 0$. In 1963, Frisch and Smith [5] put a muon detector at the top of Mt. Washington (1,907 m above sea level) and counted 563 muons per hour coming down through the atmosphere. Then they took their detector to sea level and counted 408 muons per hour. From the muon energies they inferred a speed of $0.995\,c$.

(a) If there were no time dilation, how many muons should have been measured at sea level?
(b) How does time dilation affect the experiment?
(c) Use the experimental data to determine the muons' relativistic γ factor.

10.4. Let's see how light from the Hβ transition of hydrogen (wavelength 486.13 nm) is affected by the relativistic Doppler effect and time dilation.

(a) Consider debris from a supernova moving directly toward an observer on Earth with a speed $v = 18{,}000\,\mathrm{km\,s^{-1}}$. At what wavelength would the Hβ spectral line from the debris be observed?
(b) Imagine instead the debris is moving perpendicular to our line of sight (i.e., "in the plane of the sky") with a transverse velocity $v = 18{,}000\,\mathrm{km\,s^{-1}}$. Now at what wavelength would the Hβ line be observed? Hint: you can consider the light to be a "clock" with a frequency ν.
(c) What would the predicted wavelengths have been for parts (a) and (b) if we had ignored special relativity and used the "classical" Doppler formula $\Delta\lambda/\lambda = v_{\mathrm{radial}}/c$?

10.5. The rest mass energy of a proton is about 938 MeV. The Large Hadron Collider is designed to accelerate protons to an energy of about 7 TeV. How fast do such protons move? Hint: write $v = (1 - \delta)c$ and find δ.

10.6. Suppose a probe that emits a flash of green light ($\lambda = 500\,\text{nm}$) once every second is dropped into the black hole at the center of the Milky Way, starting at rest from $r = 2R_S$. Use $M = 4 \times 10^6\, M_\odot$ (see Sect. 3.2.1).

(a) In the probe's reference frame, how much time elapses as it falls from its starting point to the event horizon? From the event horizon to the center? Hint: by changing variables, you can express the integral in a form that can be evaluated using Sect. A.7.
(b) Describe qualitatively what you would see from a fixed vantage point far from the black hole as the probe falls in.
(c) What is the wavelength of the first flash, as measured by a distant observer? What about the last flash emitted by the probe before it crosses the event horizon? Hint: you will need to use a numerical root finder to solve for radius corresponding to the last integer second before the probe crosses the event horizon.

10.7. For a particle moving in the Schwarzschild metric, the effective potential is given by Eq. (10.37). Consider a circular orbit.

(a) For a massive particle ($\tilde{m} = 1$), there are two possible circular orbits. What are their radii? What is the smallest value of ℓ for which the answer is physical? What is the radius of the smallest possible circular orbit?
(b) For a photon ($\tilde{m} = 0$), show that it is possible to have a circular orbit at a particular radius, but the orbit is unstable.

10.8. In a binary star system, the accelerating masses create gravitational radiation that removes energy from the system. For stars of mass M_1 and M_2 in nearly circular orbits separated by distance a, the power emitted in gravitational radiation is [26]

$$P = \frac{32G^4}{5c^5}\frac{(M_1 M_2)^2(M_1 + M_2)}{a^5}$$

(a) Explain conceptually what happens to the stars' orbits.
(b) Using conservation of energy, derive a differential equation for the semimajor axis as a function of time. Solve the equation to find how long it takes to go from some initial semimajor axis $a = a_0$ to $a = 0$.
(c) The binary pulsar system J0737−3039 has an orbital period of $P = 0.102$ day, and the stars move in nearly circular orbits with speeds of about $310\,\text{km s}^{-1}$. When will the stars merge? (Assume the stars have the same mass. Problem 4.5 gives more precise parameters, but we use simpler approximations here.)

References

1. B.W. Carroll, D.A. Ostlie, *An Introduction to Modern Astrophysics*, 2nd edn. (Addison-Wesley, San Francisco, 2007)
2. B.F. Schutz, *A First Course in General Relativity* (Cambridge University Press, Cambridge, 1985)
3. P. Galison, *Einstein's Clocks, Poincaré's Maps: Empires of Time* (W.W. Norton, New York, 2003)
4. A. Einstein, *Einstein's 1912 Manuscript on the Special Theory of Relativity: A Facsimile* (George Braziller, New York, 1996)
5. D.H. Frisch, J.H. Smith, Am. J. Phys. **31**, 342 (1963)
6. J.C. Hafele, R.E. Keating, Science **177**, 166 (1972)
7. J.C. Hafele, R.E. Keating, Science **177**, 168 (1972)
8. C. Misner, K. Thorne, J. Wheeler, *Gravitation* (W.H. Freeman, San Francisco, 1973)
9. A. Pais, *Subtle Is the Lord: The Science and the Life of Albert Einstein* (Oxford University Press, Oxford/New York, 1982)
10. S. Schlamminger, K.Y. Choi, T.A. Wagner, J.H. Gundlach, E.G. Adelberger, Phys. Rev. Lett. **100**(4), 041101 (2008)
11. R. Baum, W. Sheehan, *In Search of Planet Vulcan: The Ghost in Newton's Clockwork Universe* (Basic Books, New York, 2003)
12. C. Pollack, Mercury's Perihelion (2003), http://www.math.toronto.edu/~colliand/426_03/-Papers03/C_Pollock.pdf
13. R.V. Pound, G.A. Rebka, Phys. Rev. Lett. **3**, 439 (1959)
14. R.V. Pound, G.A. Rebka, Phys. Rev. Lett. **4**, 337 (1960)
15. J.L. Greenstein, J.B. Oke, H.L. Shipman, Astrophys. J. **169**, 563 (1971)
16. N.S. Hetherington, Q. J. R. Astron. Soc. **21**, 246 (1980)
17. N. Ashby, Living Rev. Relativ. **6**, 1 (2003)
18. K.S. Thorne, *Black Holes & Time Warps: Einstein's Outrageous Legacy* (W.W. Norton, New York, 1994)
19. C.R. Keeton, A.O. Petters, Phys. Rev. D **72**(10), 104006 (2005)
20. R.P. Kerr, Phys. Rev. Lett. **11**, 237 (1963)
21. J.M. Miller, Annu. Rev. Astron. Astrophys. **45**, 441 (2007)
22. R.A. Hulse, J.H. Taylor, Astrophys. J. Lett. **195**, L51 (1975)
23. M. Burgay et al., Nature **426**, 531 (2003)
24. J.M. Weisberg, D.J. Nice, J.H. Taylor, Astrophys. J. **722**, 1030 (2010)
25. M. Kramer, et al., Science **314**, 97 (2006)
26. S.L. Shapiro, S.A. Teukolsky, *Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects* (Wiley, New York, 1986)