Einstein Equation and its Spherical Solution

6

- At every spacetime point, one can construct a free-fall frame in which gravity is transformed away. However, in a finite-sized region, one can detect the residual tidal forces that stem from the second derivatives of the gravitational potential. The curvature of spacetime characterizes these tidal effects.
- The GR field equation, a set of coupled partial differential equations, directly relates spacetime's curvature to the mass/energy distribution. Its solution (the metric) determines the geometry of spacetime.
- We briefly introduce the concept of curvature in differential geometry.
- The nonrelativistic theory of tidal forces yields the Newtonian deviation equation, whose relativistic analog is the equation of geodesic deviation, which relates spacetime curvature to the relative motion of nearby particles.
- We explore the symmetries and contractions of the Riemann curvature in search of a rank-2 symmetric curvature tensor appropriate for the GR field equation.
- A spherically symmetric metric has two unknown scalar functions: g_{00} and g_{rr}. The Schwarzschild solution to the GR field equation yields $g_{00} = -g_{rr}^{-1} = -(1 - r^*/r)$, with $r^* = 2G_N M/c^2$. An embedding diagram allows us to visualize such a warped space.
- GPS time corrections and the precession of Mercury's perihelion are worked out as successful applications of GR.
- Gravitational deflection of light is calculated, showing comparable contributions from gravitational time dilation and from gravitational length contraction. This basic GR effect is the basis of the gravitational lensing technique, an important tool of modern cosmology.

In Chapter 5, we introduced Einstein's geometric theory of gravitation. Its equation of motion is identified with the geodesic equation. Here we present the GR field equation, Einstein's equation, whose key element is spacetime curvature. We introduce in Section 6.1 the geometry of curvature, which can be physically

A College Course on Relativity and Cosmology. First Edition. Ta-Pei Cheng.

interpreted as relativistic tidal gravity. The Newtonian theory of tidal forces is discussed in Section 6.2. After finding the GR field equation, we present its most important solution: the Schwarzschild exterior solution, which describes the spacetime outside a spherical source. Some notable examples of the geodesics in such a geometry are worked out.

The curvature of spacetime and Einstein's equation

With the identification of curved spacetime as the gravitational field and of its metric as the relativistic generalization of Newton's gravitational potential,

$$[\Phi(x)] \quad \longrightarrow \quad \big[g_{\mu\nu}(x)\big], \tag{6.1}$$

the GR generalization of Newton's equation of motion (4.9) is naturally identified with the geodesic equation (5.29):

$$\left[\frac{d^2\mathbf{r}}{dt^2} + \nabla\Phi = 0\right] \longrightarrow \left[\frac{d^2x^\sigma}{d\tau^2} + \Gamma^\sigma_{\lambda\rho}\frac{dx^\lambda}{d\tau}\frac{dx^\rho}{d\tau} = 0\right]. \tag{6.2}$$

Now our task is to find the GR generalization of Newton's field equation (4.7):

$$\big[\nabla^2\Phi = 4\pi G_{\rm N}\rho\big] \longrightarrow \quad [?]. \tag{6.3}$$

Let us lay out a strategy to find this GR field equation—the Einstein equation.

Seeking the GR field equation In Section 3.2.4, we learned that the mass density ρ is proportional to T_{00}, a component of the symmetric relativistic energy–momentum tensor. Thus, the right-hand side of the Einstein equation should be proportional to $T_{\mu\nu} = T_{\nu\mu}$. The GR field equation must be covariant and hence a tensor equation; the left-hand side (which incorporates the field, i.e., $g_{\mu\nu}$) must have the same tensor structure:

$$[\hat{O}g]_{\mu\nu} = \kappa\, T_{\mu\nu}, \tag{6.4}$$

where κ is some proportionality constant presumably related to $G_{\rm N}$. Namely, the left-hand side should likewise be a symmetric tensor of rank 2, $[\hat{O}g]_{\mu\nu} = [\hat{O}g]_{\nu\mu}$, resulting from some differential operator $\hat{O}$ acting on the metric $[g]$. Since we expect $[\hat{O}g]_{00}$ to have a Newtonian limit proportional to $\nabla^2\Phi$, $[\hat{O}]$ must be a second-derivative operator. Besides the $\partial^2 g$ terms, we also expect it to contain nonlinear terms of the $(\partial g)^2$ type. This follows from the fact that energy, just like mass, is a source of gravity, and gravitational fields themselves hold energy. Just as electromagnetic fields have an energy density quadratic in the fields $\sim (\mathbf{E}^2 + \mathbf{B}^2)$, the gravitational field energy density must be quadratic in the gravitational field strength ∂g. In terms of Christoffel symbols $\Gamma \sim \partial g$ (cf. (5.30)), we anticipate $[\hat{O}g]$ to contain not only $\partial\Gamma$ but also Γ^2 terms.

Thus, seeking the GR field equation reduces to finding such a tensor $[\hat{O}g]$:

$[\hat{O}g]$ must be
a symmetric rank-2 tensor, composed of terms like $(\partial^2 g), (\partial g)^2 \sim \partial\Gamma, \Gamma^2$.

(6.5)

In our discussion of the flatness theorem in Section 5.1.3, we identified the second derivative of the metric with the curvature of spacetime. We shall elaborate this geometric view of curvature in Section 6.1. We will then present in Section 6.2 another aspect of curvature—as relativistic tidal forces. This will be followed by a discussion of Einstein's equation itself in Section 6.3.

6.1 Curvature: a short introduction

Only a flat space can have a metric tensor whose elements are all constant. However, one cannot determine that a space is curved just because its metric elements are coordinate-dependent. Consider, for instance, a 2D flat surface. Its metric elements in Cartesian coordinates are all constant, $[g^{(x,y)}] = [\mathbb{I}]$, but they are not so in polar coordinates (r, ϕ),

$$[g^{(r,\phi)}]_{ab} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}. \tag{6.6}$$

Thus, to conclude that a space is curved, one has to ensure that there does not exist any coordinate system in which the metric is constant.

Gaussian curvature of a 2D curved surface

To characterize 2D flat and curved surfaces, Gauss discovered a unique invariant called the curvature K. It is composed of the second derivative of the metric tensor $\partial^2 g$ and the nonlinear term $(\partial g)^2$ in such a way, independent of the coordinate choice, $K = 0$ for flat surfaces and $K \neq 0$ for curved surfaces. Since this curvature K is expressed entirely in terms of the metric and its derivatives, it is an intrinsic geometric property. With no loss of generality, we shall quote Gauss's result for a diagonalized metric $g_{ab} = \mathrm{diag}(g_{11}, g_{22})$:

$$\begin{aligned} K = \frac{1}{2g_{11}g_{22}} \Bigg\{ & -\frac{\partial^2 g_{11}}{\left(\partial x^2\right)^2} - \frac{\partial^2 g_{22}}{\left(\partial x^1\right)^2} \\ & + \frac{1}{2g_{11}} \left[\frac{\partial g_{11}}{\partial x^1}\frac{\partial g_{22}}{\partial x^1} + \left(\frac{\partial g_{11}}{\partial x^2}\right)^2 \right] + \frac{1}{2g_{22}} \left[\frac{\partial g_{11}}{\partial x^2}\frac{\partial g_{22}}{\partial x^2} + \left(\frac{\partial g_{22}}{\partial x^1}\right)^2 \right] \Bigg\}. \end{aligned} \tag{6.7}$$

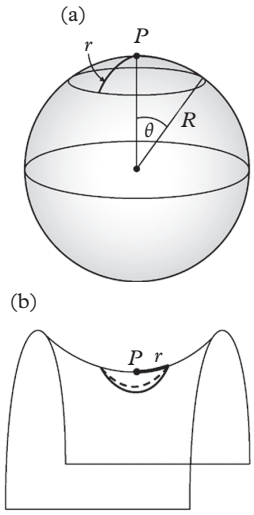

Figure 6.1 *A circle with radius r centered on a point P on a curved surface: (a) on a spherical surface with curvature $K = 1/R^2$; (b) on the middle portion of a saddle-shaped surface, which has negative curvature $K = -1/R^2$.*

[1] One can think of this 3D embedding space as Minkowski spacetime with one temporal and two spatial coordinates. The usual spacetime diagrams are attempts to represent the pseudo-Euclidean geometry in a 3D Euclidean space. In such a diagram, a pseudosphere is portrayed as a hyperbolic bowl of points displaced from the origin by a fixed proper time (a negative interval $x^2 + y^2 - c^2t^2 = -c^2\tau^2$). Such a bowl in Euclidean space would have positive curvature—for example, try to draw or envision in 3D spacetime a 2-sphere, the set of points displaced from the origin by a fixed positive (spacelike) interval. The hyperbolic tube appears to have negative curvature. As demonstrated by our discussion in Box 5.3, our Euclidean brains need lots of practice in order to grasp something as counterintuitive as pseudo-Euclidean geometry.

We will not present its derivation (see however Sidenote 11.7), but only indicate that for spaces of constant curvature (or for any infinitesimal surface),

$$K = \frac{k}{R^2}, \tag{6.8}$$

where R is the radius of curvature (which in the case of a spherical surface is simply the radius of the sphere) and k is the curvature signature: $k = 0$ for a flat surface; $k = +1$ for a spherical surface (called a 2-sphere in geometry); and $k = -1$ for a hyperbolic surface, a 2-pseudosphere.

Exercise 6.1 Gaussian curvature is coordinate-independent

Check the coordinate independence of the curvature (a) for a flat plane in Cartesian and polar coordinates and (b) for a spherical surface with radius R in polar coordinates and cylindrical coordinates, by plugging their respective metrics (6.6), (5.8), and (5.9) into the curvature formula (6.7).

Pseudosphere We have already displayed the embedding of a spherical surface in a 3D Euclidean space (with metric $g_{ij} = \delta_{ij} = \mathrm{diag}(1, 1, 1)$) as in (5.1); cf. Fig. 5.1. Unlike the cases of the plane ($k = 0$) or the sphere ($k = +1$), there is no simple way to visualize the whole ($k = -1$) pseudosphere, because its natural embedding is not into a 3D Euclidean space but into a flat 3D space with a pseudo-Euclidean metric $\eta_{ij} = \mathrm{diag}(1, 1, -1)$. In such a space[1] with coordinates (X, Y, Z), the embedding condition is $X^2 + Y^2 - Z^2 = -R^2$. While we cannot draw the whole pseudosphere in an ordinary 3D Euclidean space, the central portion of a saddle surface does have a negative curvature; see Fig. 6.1(b).

Exercise 6.2 Pseudosphere metric from embedding coordinates

(i) Recover the result (5.8) for the metric g_{ab} of a 2D $k = +1$ surface in the Gaussian coordinates $(x^{1,2} = \theta, \phi)$ by way of its 3D embedding coordinates $X^i(\theta, \phi)$ as shown in (5.3). From the invariant interval in the 3D embedding space, extended to the Gaussian coordinate differentials by the chain rule of differentiation, we have

$$ds^2 = \delta_{ij}\, dX^i\, dX^j = \delta_{ij} \frac{\partial X^i}{\partial x^a} \frac{\partial X^j}{\partial x^b}\, dx^a\, dx^b. \tag{6.9}$$

By comparing this with (5.5), you can identify the metric for the 2D space:

$$g_{ab} = \delta_{ij} \frac{\partial X^i}{\partial x^a} \frac{\partial X^j}{\partial x^b}, \tag{6.10}$$

which can be viewed as the transformation of the metric from Cartesian to polar coordinates.

(ii) For a $k = -1$ 2D pseudosphere, its hyperbolic Gaussian coordinates ($x^{1,2} = \psi, \phi$) can be related to the 3D embedding coordinates, in analogy with (5.3), by

$$X^1 = R\sinh\psi\cos\phi, \quad X^2 = R\sinh\psi\sin\phi, \quad X^3 = R\cosh\psi. \tag{6.11}$$

This embedding space has the metric $\eta_{ij} = \mathrm{diag}(1, 1, -1)$. Follow the above steps to deduce the 2D space metric

$$g_{ab}^{(\psi,\phi)} = R^2 \begin{pmatrix} 1 & 0 \\ 0 & \sinh^2\psi \end{pmatrix}. \tag{6.12}$$

(iii) Show that this metric leads, via (6.7), to a negative Gaussian curvature $K = -1/R^2$.

(iv) Show that while the circumference of a circle with radius r on a spherical surface is $2\pi R\sin(r/R)$ which is smaller than $2nr$ as shown in Fig. 6.1(a), on a hypersphere it is $2\pi R\sinh(r/R)$, and hence greater than $2\pi r$.

(v) A $k = +1$ 2D sphere has no boundary but a finite area: $A_2 = \int ds_\theta\, ds_\phi = 4\pi R^2$. Similarly demonstrate that a $k = -1$ 2D pseudosphere has no boundary but an area that is infinite: $\tilde{A}_2 = \int ds_\psi\, ds_\phi = \infty$.

In cosmology, we shall encounter 4D spacetimes with 3D spatial subspaces having constant curvature. Besides the 3D flat space, it is also possible to have 3D spaces with positive and negative curvature: 3-spheres and 3-pseudospheres. When they are embedded in 4D space with metric $g_{\mu\nu} = \mathrm{diag}(\pm 1, 1, 1, 1)$ and coordinates (W, X, Y, Z), these 3D spaces obey the constraint conditions

$$\pm W^2 + X^2 + Y^2 + Z^2 = \pm R^2. \tag{6.13}$$

Box 6.1 Metric tensors for spaces with constant curvature

Let us first collect the metric expressions in polar coordinates for the three 2D surfaces with constant curvature $K = k/R^2$. For the flat plane ($k = 0$), we have (6.6); for the sphere ($k = +1$), we have (5.8); and for the pseudosphere ($k = -1$), we have (6.12). In terms of the dimensionless radial coordinate $\chi \equiv r/R$ (previously we had $\chi = \theta$ and $\chi = \psi$), they all take a similar form:

continued

Box 6.1 *continued*

$$[ds^2]^{(k)}_{2D,\chi} = \begin{cases} R^2(d\chi^2 + \sin^2\chi\, d\phi^2) & \text{for } k = +1, \\ R^2(d\chi^2 + \chi^2\, d\phi^2) & \text{for } k = 0, \\ R^2(d\chi^2 + \sinh^2\chi\, d\phi^2) & \text{for } k = -1. \end{cases} \tag{6.14}$$

This similarity among the metrics suggests an even more compact expression that combines these three equations into one:[2]

$$[ds^2]^{(k)}_{2D,\chi} = R^2[d\chi^2 + k^{-1}(\sin^2\sqrt{k}\chi)d\phi^2]. \tag{6.15}$$

[2] Recall from Box 3.1 (in particular Sidenote 3) that $\sin i\chi = i\sinh\chi$. For $k = 0$, (6.15) is understood to mean the limit as $k \to 0$, so that $\sin^2\sqrt{k}\chi \to k\chi^2$.

3D spaces with constant curvature

We are particularly interested in 3D spaces of constant curvature, which will be relevant in our study of cosmology. The 2D spaces with constant curvature have metrics displayed in (6.14) with polar coordinates. We now make a heuristic argument for the generalization of the 2D results (6.15) to the corresponding 3D spaces.[3] We start with the observation that the 2D $k = 0$ metric for a flat plane in polar coordinates (r, ϕ) is $[ds^2]^{(k=0)}_{2D} = dr^2 + r^2\, d\phi^2$. For the $k = 0$ flat 3D space, the usual spherical coordinate system (r, θ, ϕ) involves an additional (polar) angle coordinate θ. We have the familiar metric $[ds^2]^{(k=0)}_{3D} = dr^2 + r^2\, d\Omega^2$, where $d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2$ is the differential solid angle. We now suggest that even for the curved spaces $k \neq 0$, we can similarly obtain the 3D metrics from their 2D counterparts by the simple substitution $d\phi^2 \to d\Omega^2$. We can thereby infer from (6.15) the metric for the $(k = 0, \pm 1)$ 3D spaces in spherical polar coordinates (χ, θ, ϕ):

$$[ds^2]^{(k)}_{3D,\chi} = R^2[d\chi^2 + k^{-1}(\sin^2\sqrt{k}\chi)\, d\Omega^2]. \tag{6.16}$$

[3] A rigorous derivation of these results would involve the mathematics of symmetric spaces, Killing vectors, and isometry. Our heuristic argument can also be supported by solving the Einstein equation for a homogeneous and isotropic 3D space; see (Cheng 2010, Section 14.4.1).

Extending in a similar way the result shown in (5.9), we can write down the metric for the three 3D spaces of constant curvature in cylindrical coordinates (ρ, θ, ϕ) using the dimensionless radial coordinate $\xi \equiv \rho/R$:

$$[ds^2]^{(k)}_{3D,\xi} = R^2\left(\frac{d\xi^2}{1 - k\xi^2} + \xi^2\, d\Omega^2\right). \tag{6.17}$$

Curvature measures the deviation from Euclidean relations On a flat surface, the familiar Euclidean geometrical relations hold. For example, the circumference of a circle with radius r is $S = 2\pi r$. The curvature measures how curved a surface is because it is directly proportional to the violation of Euclidean relations. Figure 6.1 shows two circles of radius r drawn on surfaces with

nonvanishing curvature. The circular circumference S differs from its flat-surface value $2\pi r$ by an amount proportional to the Gaussian curvature K:

$$\lim_{r\to 0}\frac{2\pi r - S}{r^3} = \frac{\pi}{3}K. \tag{6.18}$$

On a positively curved surface, the circumference is smaller than (or, on a negatively curved surface, larger than) that on a flat surface. Another simple example is that on a curved surface the angular excess ϵ of a polygon (the difference between the sum of the interior angles and the corresponding Euclidean sum) is proportional to its area σ, with the proportionality constant being the curvature K:

$$\epsilon = K\sigma. \tag{6.19}$$

Figure 6.2 illustrates this relation[4] for the case of a triangle on a sphere. The illustration also demonstrates that this angular excess can be measured by the directional change of a vector that is parallel-transported around this polygon. Such a 2D relation can be extended to higher-dimensional spaces, suggesting the way to define the curvature tensor in a general-dimensional space; see Section 11.2.1.

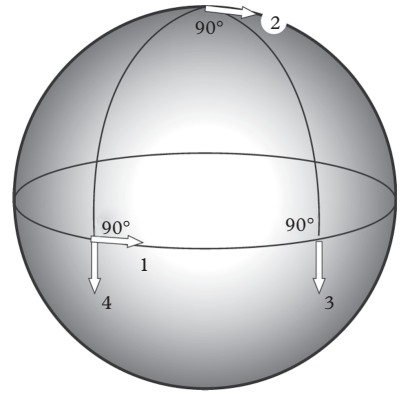

Figure 6.2 *A triangle with all interior angles of* 90° *on a spherical surface (radius R) has an angular excess* $\epsilon = \pi/2$. *This satisfies the relation* $\epsilon = K\sigma$, *with curvature* $K = 1/R^2$ *and the triangular area* $\sigma = \pi R^2/2$ *equal to one-eighth of the spherical area. The angular excess can be measured by the parallel transport of a vector around this triangle (vector-1, clockwise to vector-2, to -3, and finally back to the starting point as vector-4), which rotates the vector by* 90° *(the angular difference between vector-4 and vector-1).*

Riemann curvature tensor

Extending Gauss's discovery of the curvature (a single component) in 2D space, his pupil Riemann (with further contributions by Christoffel) established the existence of a rank-4 tensor, the Riemann curvature tensor, in an n-dimensional space:

$$R^{\mu}_{\ \lambda\alpha\beta} = \partial_\alpha\Gamma^{\mu}_{\lambda\beta} - \partial_\beta\Gamma^{\mu}_{\lambda\alpha} + \Gamma^{\mu}_{\nu\alpha}\Gamma^{\nu}_{\lambda\beta} - \Gamma^{\mu}_{\nu\beta}\Gamma^{\nu}_{\lambda\alpha}. \tag{6.20}$$

The Christoffel symbols Γ are first derivatives, as shown in (5.30), so this curvature $R = \partial\Gamma + \Gamma\Gamma$ is a nonlinear second-derivative function of the metric, with terms like $\partial^2 g$ and $(\partial g)^2$. It measures, independently of coordinate choice, the deviation from flat-space relations; thus $R^{\mu}_{\ \lambda\alpha\beta} = 0$ for a flat spacetime.[5] As we shall see, a linear combination of the contractions of this curvature tensor enters directly into the GR field equation, the Einstein equation.

There are many (mutually consistent) ways to derive the expression for the curvature tensor displayed in (6.20). One simple method is to generalize the 2D relation (6.19) to higher dimensions by calculating the change in a vector A parallel-transported around an infinitesimal parallelogram spanned by two infinitesimal vectors a^μ and b^ν. As the relative change of the vector yields the angular excess ($dA/A = \epsilon$ for small changes), the higher-dimensional generalization of (6.19) should then be

$$dA^\mu = -R^{\mu}_{\ \nu\lambda\rho}A^\nu a^\lambda b^\rho. \tag{6.21}$$

Namely, the vectorial change is proportional to the vector A^ν itself and to the area of a parallelogram spanned by two (infinitesimal) vectors a^λ and b^ρ.

[4] A somewhat less specific example would be a triangle with two right angles and a third unspecified angle θ; thus $\epsilon = \theta$. Such a triangle has an area θR^2, satisfying the $\epsilon = K\sigma$ relation. For a proof using a general spherical triangle with arbitrary interior angles, see (Cheng 2010, p. 92 in Section 5.3.2). Such a relation for triangles can be extended to polygons, as any polygon can be decomposed into triangles.

[5] It should be noted that in cosmology we shall often refer to a flat universe that has a nontrivial spacetime $R^{\mu}_{\ \lambda\alpha\beta} \neq 0$ but has a 3D subspace with a flat geometry (having a curvature signature $k = 0$).

The coefficient of proportionality $R^{\mu}{}_{\nu\lambda\rho}$ is a quantity with four indices, antisymmetric in λ and ρ because the area $a^{\lambda}b^{\rho}$ should be antisymmetric. We shall take this to be the definition of the curvature tensor of an n-dimensional space. For an explicit derivation, see Section 11.2.1

6.2 Tidal gravity and spacetime curvature

What is the physical significance of the curvature? Curvature involves second derivatives of the metric. Since the metric tensor can be regarded as the relativistic gravitational potential, its first derivatives are gravitational accelerations, so its second derivatives must then be relative accelerations (per unit separation) between neighboring particles. Namely, the second derivatives give rise to relativistic tidal forces.

6.2.1 Tidal forces—a qualitative discussion

Here we give an elementary and qualitative discussion of the (nonrelativistic) tidal force and its possible geometric interpretation. This will be followed by a more mathematical presentation in terms of the Newtonian deviation equation, which relates the relative motion due to tidal forces between two neighboring particles to the second derivatives of the gravitational potential. This in turn suggests the GR generalization, the equation of geodesic deviation, which relates the relative spacetime motion of two particles to the underlying spacetime curvature.

The EP states that in a freely falling reference frame, the physics is the same as that in an inertial frame with no gravity; SR applies, and the metric is the Minkowski metric $\eta_{\mu\nu}$. As shown in the flatness theorem (Section 5.1.3), one can only approximate $g_{\mu\nu}$ by $\eta_{\mu\nu}$ locally, that is, in an appropriately small region. Gravitational effects can always be detected in a finite-sized free-fall frame, as the gravitational field in reality is never strictly uniform; the second derivatives of the metric come into play (as the first derivatives vanish in a free-fall frame).

Consider the lunar gravitational attraction exerted on the earth. While the earth is in free fall toward the moon (and vice versa), there is still a detectable lunar gravitational effect on earth. Different points on earth feel slightly different gravitational pulls by the moon, as depicted in Fig. 6.3(a). The center-of-mass (CM) force causes the earth and everything on it to fall toward the moon, so this CM gravitational effect is canceled out in the freely falling terrestrial frame. But subtracting out this CM force leaves residual (relative) forces on various parts of the earth, as shown in Fig. 6.3(b). They are a stretching in the longitudinal direction and a compression in the transverse directions. This is just the familiar tidal gravity.[6] Namely, in the free-fall frame, the CM gravitational effect is transformed away, but there are still remnant tidal forces. They reflect the differences of the gravitational effects on neighboring points, and are thus proportional to the derivatives of the gravitational field (and hence to the second derivatives of

[6] The ocean is pulled away in opposite directions, giving rise to two tidal bulges. This explains why, as the earth rotates, there are two high tides in a day. This of course is a simplified description, as there are other effects (e.g., the solar tidal forces) that must be taken into account.

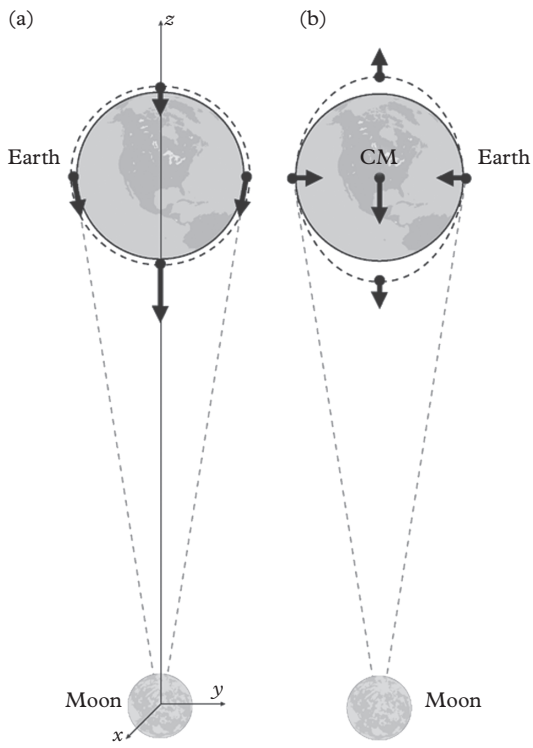

Figure 6.3 *Variations of the gravitational field as tidal forces. (a) Lunar gravitational forces on four representative points on earth. (b) After taking out the center-of-mass (CM) motion, the relative forces on earth are the tidal forces of longitudinal stretching and transverse compression.*

the potential Φ). Since tidal forces cannot be coordinate-transformed away, they should be regarded as the essence of gravitation.

6.2.2 Deviation equations and tidal gravity

Here we provide a quantitative description of gravitational tidal forces in the Newtonian framework, which suggests an analogous GR approach in which the curvature tensor plays the role of the second derivatives of the potential in producing tidal gravity.

Newtonian deviation equation for tidal forces

The tidal effect concerns the relative motion of particles in a nonuniform gravitational field. Let us consider two particles: one has trajectory $\mathbf{x}(t)$ and the other has $\mathbf{x}(t) + \mathbf{s}(t)$. That is, the locations of these two particles measured at the same time have a coordinate separation $\mathbf{s}(t)$. The respective equations of motion ($i = 1, 2, 3$) obeyed by these two particles, according to (4.9), are

$$\frac{d^2x^i}{dt^2} = -\frac{\partial\Phi(x)}{\partial x^i} \quad \text{and} \quad \frac{d^2x^i}{dt^2} + \frac{d^2s^i}{dt^2} = -\frac{\partial\Phi(x+s)}{\partial x^i}. \tag{6.22}$$

Consider the case in which the separation distance $s^i(t)$ is small, so we can approximate the gravitational potential $\Phi(x+s)$ by a Taylor expansion

$$\Phi(x+s) = \Phi(x) + \frac{\partial \Phi}{\partial x^j} s^j + \cdots . \tag{6.23}$$

From the difference of the two equations in (6.22), we obtain the Newtonian deviation equation that describes the separation between two nearby particle trajectories in a gravitational field:

$$\frac{d^2 s^i}{dt^2} = -\left(\frac{\partial^2 \Phi}{\partial x^i \partial x^j}\right) s^j. \tag{6.24}$$

Thus the relative acceleration per unit separation, $(d^2 s^i/dt^2)/s^j$, is given by the tidal force matrix, a rank-2 tensor having the second derivatives of the gravitational potential as its elements.

The spherically symmetric field As an illustrative application of (6.24), let us discuss the case of a spherical gravitational source of mass M (e.g., the gravity due to the moon on earth as shown in Fig. 6.3), $\Phi(x) = -G_{\rm N}M/r$, where the radial distance is related to the rectangular coordinates by $r = (x^2 + y^2 + z^2)^{1/2}$. Since $\partial r/\partial x^i = x^i/r$ and $\partial \Phi/\partial x^i = G_{\rm N} M x^i/r^3$, the tidal force tensor is

$$\frac{\partial^2 \Phi}{\partial x^i \partial x^j} = \frac{G_{\rm N} M}{r^3}\left(\delta_{ij} - \frac{3x^i x^j}{r^2}\right). \tag{6.25}$$

Consider a reference particle located along the z axis: $x^i = (0, 0, r)$. The Newtonian deviation equation (6.24) for the displacement of a second particle from the reference, with the tidal force tensor given by (6.25), becomes[7]

$$\frac{d^2}{dt^2}\begin{pmatrix} s_x \\ s_y \\ s_z \end{pmatrix} = \frac{-G_{\rm N} M}{r^3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}\begin{pmatrix} s_x \\ s_y \\ s_z \end{pmatrix}. \tag{6.26}$$

We see that there is an attractive tidal force between the two particles in the transverse directions, $d^2 s_{x,y}/dt^2 = -G_{\rm N} M s_{x,y}/r^3$, which leads to compression, while a tidal repulsion, $d^2 s_z/dt^2 = +2G_{\rm N} M s_z/r^3$, leads to stretching in the longitudinal (i.e., radial) direction.[8] This verifies the tidal effects illustrated in Fig. 6.3(b).

GR equation of geodesic deviation and the curvature tensor

We shall follow a similar approach in GR. The two equations of motion (6.22) will be replaced by the corresponding GR equations of motion (the geodesic equations). A Taylor expansion of their difference yields the equation of geodesic

[7] Since we have taken $x^i = (0, 0, r)$, the term $x^i x^j$ on the right-hand side can be worked out with $x^1 = x^2 = 0$ and $x^3 = r$.

[8] One notable feature of the tidal force matrix is that it is traceless. This follows from Newton's equation $\delta_{ij}\partial^2\Phi/\partial x^i \partial x^j = \nabla^2 \Phi = 0$ in the space exterior to the source.

deviation,[9] a generalization of (6.24) replacing the tidal force tensor by a contraction of the Riemann curvature tensor, given by (6.20), with two factors of the relative 4-velocity:

$$\left[\frac{\partial^2\Phi}{\partial x^i \partial x^j}\right] \quad \longrightarrow \quad \left[R^{\mu}{}_{\alpha\nu\beta}\dot{x}^{\alpha}\dot{x}^{\beta}\right]. \tag{6.27}$$

Thus the relative motion of nearby particles, like that of the converging particles in Fig. 5.4, can be attributed to curvature. Such considerations led Einstein to give gravity a direct geometric interpretation by identifying tidal gravity with the curvature of spacetime.

[9] See Section 11.2.2, in particular Eq. (11.67). To carryout this derivation one would have to use the concept of covariant derivative, which will be introduced in Chapter 11.

6.3 The GR field equation

We now discuss the contraction of the rank-4 Riemann curvature tensor (6.20) with the aim of finding, as outlined in (6.5), the appropriate rank-2 tensor for the GR field equation.

6.3.1 Einstein curvature tensor

There are six ways to contract a rank-4 tensor to produce a rank-2 tensor. However, because of the symmetries of the Riemann tensor, all of its contractions either vanish or are trivially related to one another.

Symmetries and contractions of the curvature tensor

We first note that the Riemann curvature tensor with all lower indices,

$$R_{\mu\nu\alpha\beta} = g_{\mu\lambda}R^{\lambda}{}_{\nu\alpha\beta}, \tag{6.28}$$

has the following symmetry features:

- It is antisymmetric with respect to the interchange of its first and second indices and of its third and fourth indices:

$$R_{\mu\nu\alpha\beta} = -R_{\nu\mu\alpha\beta}; \tag{6.29}$$

$$R_{\mu\nu\alpha\beta} = -R_{\mu\nu\beta\alpha}. \tag{6.30}$$

- It is symmetric with respect to the interchange of the pair made up of its first and second indices with the pair of third and fourth indices.

$$R_{\mu\nu\alpha\beta} = +R_{\alpha\beta\mu\nu}. \tag{6.31}$$

- It also has the cyclic symmetry[10]

$$R_{\mu\nu\alpha\beta} + R_{\mu\beta\nu\alpha} + R_{\mu\alpha\beta\nu} = 0. \tag{6.32}$$

Since the symmetry properties are tensor relations, they are not changed by coordinate transformations. Once they have been proved in one frame, we can

[10] Symmetries reduce the number of independent elements. Using (6.29)–(6.32), the reader is invited to verify the combinatorial result showing that out of the total of n^4 components of the Riemann tensor in an n-dimensional space, the number of independent components is $N_{(n)} = n^2(n^2 - 1)/12$. Thus $N_{(1)} = 0$; it is not possible for a one-dimensional inhabitant to see any curvature. $N_{(2)} = 1$; this is just the Gaussian curvature K for a curved surface. $N_{(4)} = 20$; there are 20 independent components in the curvature tensor for a 4D curved spacetime. Cf. (Cheng 2010, Section 13.3.2 and Problem 13.9).

then claim their validity in all frames. An obvious choice is the locally inertial frame with $\Gamma = 0$, $\partial\Gamma \neq 0$), where the curvature (6.20) takes a simpler form, $R_{\mu\nu\alpha\beta} = g_{\mu\lambda}(\partial_\alpha \Gamma^\lambda_{\nu\beta} - \partial_\beta \Gamma^\lambda_{\nu\alpha})$. In this way, we can use the fact that $g_{\mu\nu} = g_{\nu\mu}$ and $\Gamma^\mu_{\nu\lambda} = \Gamma^\mu_{\lambda\nu}$ to check easily the validity of the symmetry properties shown in (6.29)–(6.32).

Contractions of the curvature tensor

We contract the Riemann tensor, reducing its rank so it can be used in the GR field equation. Because of the symmetry properties shown above, these contractions are basically unique.

Ricci tensor $R_{\mu\nu}$ This is the Riemann curvature tensor with its first and third indices contracted:

$$R_{\mu\nu} \equiv g^{\alpha\beta} R_{\alpha\mu\beta\nu} = R^\beta{}_{\mu\beta\nu}. \tag{6.33}$$

It follows from (6.31) that the Ricci tensor is symmetric:

$$R_{\mu\nu} = R_{\nu\mu}. \tag{6.34}$$

If we had contracted a different pair of indices (or tried all six possibilities), the result would still be $\sigma R_{\mu\nu}$, with $\sigma = +1, 0$, or -1. Thus the rank-2 curvature tensor is essentially unique.

Ricci scalar R Contracting the Ricci tensor, we obtain the Ricci scalar field,

$$R \equiv g^{\alpha\beta} R_{\alpha\beta} = R^\beta{}_\beta. \tag{6.35}$$

Energy/momentum conservation and the Einstein tensor

These contractions show that there is indeed a ready-made symmetric rank-2 curvature tensor, the Ricci tensor $R_{\mu\nu}$, for the left-hand side of (6.4). Making this choice $[\hat{O}g]_{\mu\nu} = R_{\mu\nu}$, Einstein found that the resultant field equation did not reduce to (4.7) in the Newtonian limit. After a long struggle, he finally found (at the end of 1915) the correct tensor $[\hat{O}g]_{\mu\nu} = G_{\mu\nu}$, now known as the Einstein tensor,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}. \tag{6.36}$$

He realized that besides the Ricci tensor $R_{\mu\nu}$, there is another rank-2 symmetric curvature tensor, $Rg_{\mu\nu}$, the product of the Ricci scalar with the metric tensor itself, that satisfies the criteria for the geometry side of the field equation. In fact, any linear combination of the form $[\hat{O}g]_{\mu\nu} = R_{\mu\nu} + aRg_{\mu\nu}$ with a constant coefficient a will satisfy the search criteria (6.5). What is the correct value of a? To constrain the choice, Einstein invoked the conservation law of energy/momentum in the field system. This led him[11] to the combination (6.36).

[11] The calculation that Einstein undertook was rather complicated. After he published his GR papers in 1916, several eminent mathematicians, including David Hilbert and Felix Klein, were involved in this study. In particular, Emmy Noether (1882–1935) discovered her famous theorem connecting symmetries to conservation laws. This intense search lasting several years did bring about a greatly expedited proof showing that the Bianchi identity could lead directly to the Einstein tensor (6.36); cf. Section 11.2.3.

Remark 6.1 *The issue of energy/momentum conservation in GR is a delicate one. These conservation laws follow (via Noether's theorem) from spacetime translational symmetry. But in GR, space and time are dynamical, and a gravitational field's local energy/momentum densities are not well defined. Based on this, some would prefer to say that these conservation laws do not hold in GR. Since there can still be global conservation, we choose not to make such a strong statement.*

6.3.2 Einstein field equation

With the identification of $G_{\mu\nu}$ with the left-hand side of (6.4), Einstein (1916) arrived at his GR field equation:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \kappa T_{\mu\nu}. \tag{6.37}$$

The EP informs us that gravity can always be transformed away locally (by going to a reference frame in free fall); the immutable essence of gravity lies in its differentials (tidal forces). Thus the presence of the spacetime curvature in the GR field equation can be understood. This equation can be written in an alternative form. Contracting the two indices in (6.37) leads to $-R = \kappa T$, where T is the trace $g^{\mu\nu} T_{\mu\nu}$. In this way, we can rewrite the field equation (6.37) as

$$R_{\mu\nu} = \kappa \left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right). \tag{6.38}$$

Because the source distribution $T_{\mu\nu}$, the Ricci tensor $R_{\mu\nu}$, and the metric $g_{\mu\nu}$ are symmetric tensors, each has ten independent elements. Thus this seemingly simple field equation is actually a set of ten coupled nonlinear partial differential equations whose unknowns are the ten components[12] of the metric $g_{\mu\nu}(x)$. It is nonlinear because the solution potential can itself act as the source of gravitational potential.[13] We reiterate the central point of Einstein's theory: the spacetime geometry in GR, in contrast to SR, is not a fixed entity, but is dynamically ever-changing as dictated by the mass/energy distribution.

The Newtonian limit for a general source

One can examine one aspect of the correctness of this proposed GR field equation by checking its Newtonian limit. A straightforward calculation similar to that carried out in Section 5.3.1 shows that (6.38) does reduce to Newton's field equation (4.7) for a particle moving nonrelativistically in a weak and static field.[14] This also gives us the identification

$$\kappa = 8\pi G_{\mathrm{N}}/c^4. \tag{6.39}$$

The proportionality constant κ, and hence Newton's constant G_{N}, is the conversion factor between energy density on the right-hand side and the geometric quantity on the left-hand side of Einstein's equation (6.37) or (6.38).

[12] There is subtle point here: actually not all ten components of the metric are independent. The principle of GR requires that a solution of the field equation be the same in any choice of coordinates. This means that if $g_{\mu\nu}(x)$ is a solution, so must $g_{\mu\nu}(x')$, where the x'^{μ} are related to the x^{μ} by any general coordinate transformations. If there are only six independent metric elements, do the ten field equations overdetermine them? We shall resolve this puzzle when we discuss the Bianchi identity and the energy/momentum conservation constraint of Einstein equations at the end of Section 11.3.2.

[13] In contrast to electrically neutral photons, the quanta of the electromagnetic field, gravitons, the quanta of the gravitational field, carry energy (i.e., gravitational charge) and thus are themselves sources of the gravitational field.

[14] See Exercise 11.6.

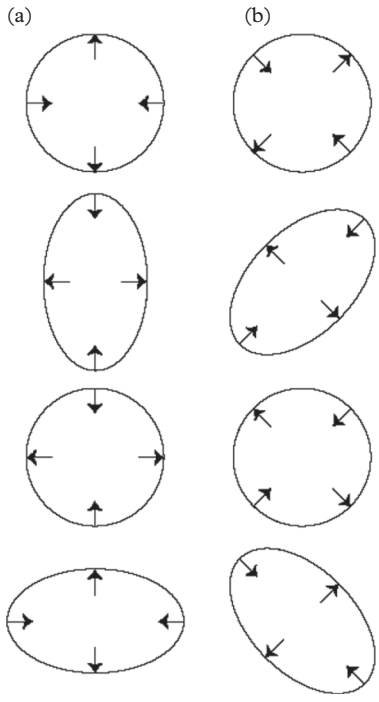

Figure 6.4 *Tidal effects of gravitational waves on a circle of test particles. The two columns represent the effects of waves in two polarization states shown in (6.44): (a) the plus-polarization and (b) the cross-polarization states. Similar to the effects of the transverse forces shown in Fig. 6.3, the circle is alternately stretched and squeezed in directions perpendicular to the wave propagation direction.*

[15] Namely, we will ignore any further gravitational waves generated by the gravitational field itself.

[16] For a derivation, see (Cheng 2010, Section 15.1.2). Just as we have chosen a particular gauge (the Lorentz gauge, $\partial^\mu A_\mu = 0$) to simplify the writing of the electromagnetic wave equation, we have made a particular coordinate choice, the transverse-traceless (TT) gauge, in writing the gravitational wave equation.

In taking the above limit, we have taken the source to be nonrelativistic matter with its energy–momentum tensor completely dominated by the $T_{00} = \rho c^2$ term. Effectively, we have taken the source to be a swarm of noninteracting particles (dust). In many situations, with cosmology being the notable example, we consider the source of gravity to be a plasma having mass density ρ and pressure p. Namely, we need to take $T_{\mu\nu}$ as the energy–momentum tensor for an ideal fluid as in (3.84). A calculation entirely similar to the Newtonian-limit calculation mentioned above leads to

$$\nabla^2 \Phi = 4\pi G_N \left(\rho + 3\frac{p}{c^2}\right). \tag{6.40}$$

In this way, Einstein's relativistic theory makes it clear that not only mass but also pressure can be a source of the gravitational field.

6.3.3 Gravitational waves

Newton's theory of gravitation is a static theory; the field due to a source is established instantaneously. Thus the field depends explicitly on the spatial coordinates, but not on time. Einstein's relativistic theory treats space and time on an equal footing. As in Maxwell's theory, the field propagates outward from the source with a finite speed c. Thus, just as one can shake an electric charge to generate electromagnetic waves, one can shake a mass to generate gravitational waves. Since gravity is ordinarily such a weak force, we shall only deal here with the case of a weak gravitational field whose metric field is almost flat:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad \text{with } h_{\mu\nu} \ll O(1). \tag{6.41}$$

This approximation linearizes[15] Einstein's theory. In this limit, gravitational waves may be viewed as small curvature ripples propagating in a background of flat spacetime. Just as the wave equation yields the electromagnetic potential $A_\mu(x)$ due to a current density $j_\mu(x)$,

$$\Box A_\mu = \frac{4\pi}{c} j_\mu, \tag{6.42}$$

the perturbed gravitational field obeys[16]

$$\Box h_{\mu\nu} = -\frac{16\pi G_N}{c^4} T_{\mu\nu}. \tag{6.43}$$

Its vacuum solution is a transverse wave having two independent polarization states and traveling at the speed of light. For a plane wave solution (with this particular choice of coordinates), we can display the field's tensor structure:

$$h_{\mu\nu}(z,t) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} e^{i\omega(z-ct)/c}. \tag{6.44}$$

Gravitational waves thus have two independent polarizations: the plus (h_+) and the cross ($h_\times$) states.[17] As we discussed previously, the gravitational field (whether static or time-dependent) affects the relative positions of test bodies. A gravitational wave alternately compresses and elongates space in directions perpendicular to that of its propagation. These tidal effects deform a circle of test particles into an ellipse whose orientation oscillates (as shown in Fig. 6.4) or rotates (by a combination of oscillations with offset phases). Contrast that with an electromagnetic wave, which pushes around a test charge (or a circle of charges in unison), causing it to oscillate back and forth (linear polarization) or in circles (by a combination of linear polarizations with offset phases). A major effort is underway to detect gravitational waves using sensitive interferometers that can measure the minute compressions and elongations of orthogonal lengths that are caused by the passage of such waves (Fig. 6.5).

In the meantime, there is already indirect[18] but convincing evidence for the existence of gravitational waves as predicted by GR. This has come from observations, spanning more than 25 years, of the orbital motion of the relativistic Hulse–Taylor binary pulsar system (PSR 1913+16). The orbiting binary stars (the shaking mass source) generate gravitational waves. Even though the binary is 21 000 light-years away from us, the basic parameters of the system can be deduced by carefully monitoring the radio pulses emitted by the pulsar, which effectively acts as an accurate and stable clock. From this record, we can verify a number of GR effects. GR predicts that gravitational radiation will carry away energy from the system, reducing the orbital period at a calculable rate.[19] The observed orbit rate decrease has been found to be in splendid agreement with the prediction by Einstein's theory (Fig. 6.6).

[17] Because the gravitation metric field is a symmetric tensor of rank 2, we expect its quantum, the graviton, to be a spin-2 particle. While a massive spin-2 particle has $2s + 1 = 5$ spin projection states, the graviton, being massless (as demonstrated by gravity's long-range nature), has only two independent helicity states (projections of spin in the direction of motion). In short, quantum theory leads us to expect exactly what GR gives us: two independent polarizations of gravitational waves.

[18] At the time of this writing, the BICEP2 Collaboration (2014) announced that they had obtained more direct evidence of gravitational waves. The *B*-mode of radiation polarization in the cosmic microwave background (CMB) was observed. This is a signature of gravitational waves generated in the big bang that left their imprint on the particle plasma at the cosmic epoch when the CMB was created. (See further comments in Chapter 10, especially Box 10.3.)

[19] For a more detailed calculation, see Eq. (15.71) in (Cheng 2010).

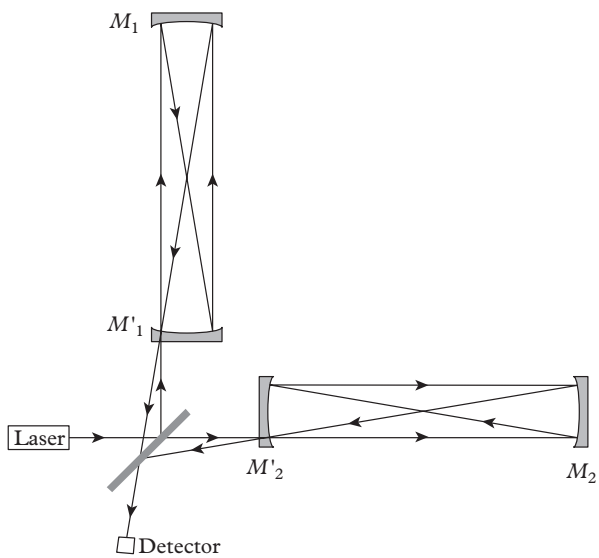

Figure 6.5 *Schematic diagram of a gravitational wave Michelson interferometer. The four mirrors* $M_{1,2}$, $M'_{1,2}$ *and the beam-splitter mirror are freely suspended. The two arms are optical cavities that increase the optical paths by many factors. A minute length change of the two arms (one expands and the other contracts) will show up as changes in the fringe pattern of the detected light.*

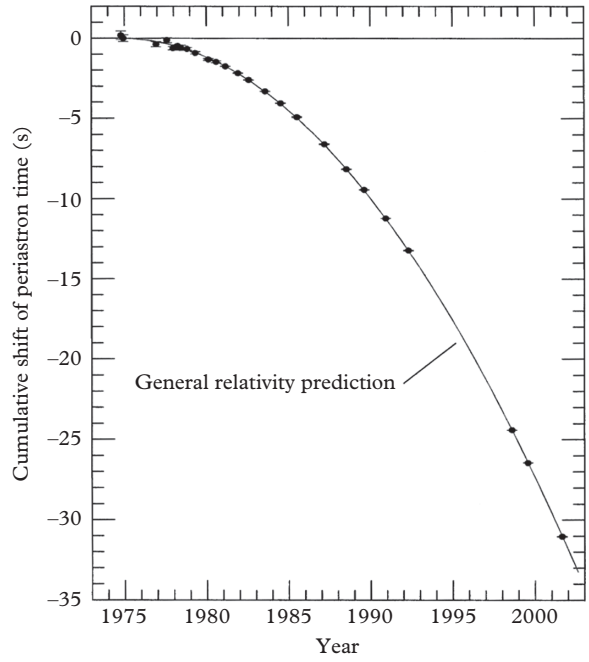

Figure 6.6 *Gravitational radiation damping causes orbital decay of the Hulse–Taylor binary pulsar. Plotted here is the accumulating shift in the epoch of periastron. The parabola is the GR prediction, and observations are depicted by data points. In most cases, the measurement uncertainties are smaller than the line widths. The data gap in the 1990s reflects the downtime when the radio telescope at the Arecibo Observatory was being upgraded. The graph is reproduced with permission from (Weisberg and Taylor 2003). ©2003 Astronomical Society of the Pacific Conference Series.*

6.4 Geodesics in Schwarzschild spacetime

Given the source mass distribution, we can solve the Einstein equation to find the metric $g_{\mu\nu}(x)$ and therefore the spacetime geometry. Here we shall concentrate on the important case of the gravitational field outside a spherical mass. The solution was obtained by Karl Schwarzschild (1873–1916) almost immediately after Einstein finally formulated his field equation in 1915. It is called the Schwarzschild metric. We note that its Newtonian analog is the gravitational potential

$$\Phi(r) = -\frac{G_{\rm N}M}{r}, \tag{6.45}$$

which is the solution to the field equation $\nabla^2\Phi(x) = 4\pi G_{\rm N}\rho(x)$ with a spherically symmetric mass density $\rho(x)$ having a total mass M inside a sphere with radius r.

In this section, we describe the metric and geometry of Schwarzschild spacetime. We apply this to GPS and two other interesting examples of geodesic motion: the deflection of a light by a massive body (e.g., the sun) and the precession of an elliptic orbit (e.g., that of the planet Mercury). For such processes within the solar system, the relativistic corrections are very small. We will study in Chapter 7 the strong gravity of black holes and in Chapters 8–10 cosmology, for which the GR description of gravitation is indispensable.

Box 6.2 Form of a spherically symmetric metric tensor

If a gravitational source is spherically symmetric, the spacetime it generates (as the solution of the Einstein equation) must also have this symmetry. Naturally, we will pick a spherical coordinate system (t, r, θ, ϕ) whose center coincides with that of the spherical source. Here we shall demonstrate explicitly that a spherically symmetric metric tensor has only two unknown scalar functions: g_{00} and g_{rr}.

General considerations of spherical symmetry

The infinitesimal invariant interval $ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu$ must be quadratic in $d\mathbf{r}$ and dt without singling out any particular angular direction.[20] That is, ds^2 must be composed of terms having two powers of $d\mathbf{r}$ and/or dt; furthermore, the vectors $\mathbf{r}$ and $d\mathbf{r}$ must appear in the form of dot products so as not to spoil the spherical symmetry:

$$ds^2 = A\, d\mathbf{r}\cdot d\mathbf{r} + B(\mathbf{r}\cdot d\mathbf{r})^2 + C\, dt\,(\mathbf{r}\cdot d\mathbf{r}) + D\, dt^2, \tag{6.46}$$

where A, B, C, and D are scalar functions of t and $\mathbf{r}\cdot\mathbf{r}$. In a spherical coordinate system with orthonormal basis vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$, we have

$$\mathbf{r} = r\hat{\mathbf{r}} \quad \text{and} \quad d\mathbf{r} = dr\,\hat{\mathbf{r}} + r\, d\theta\,\hat{\boldsymbol{\theta}} + r\sin\theta\, d\phi\,\hat{\boldsymbol{\phi}}. \tag{6.47}$$

Thus

$$\mathbf{r}\cdot\mathbf{r} = r^2, \qquad \mathbf{r}\cdot d\mathbf{r} = r\, dr,$$
$$\text{and} \quad d\mathbf{r}\cdot d\mathbf{r} = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{6.48}$$

The invariant interval is now

$$ds^2 = A[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)] + Br^2\, dr^2 + Cr\, dr\, dt + D\, dt^2, \tag{6.49}$$

or, relabeling the scalar function $A + Br^2 = B'$ as B,

$$ds^2 = A[r^2(d\theta^2 + \sin^2\theta\, d\phi^2)] + B\, dr^2 + Cr\, dr\, dt + D\, dt^2. \tag{6.50}$$

Simplification by coordinate choices

From our discussion in Chapter 5, we learned that the Gaussian coordinates, as labels of points in the space, can be freely chosen. In the same way, the coordinates in Riemannian geometry have no intrinsic significance until their connection to the physical length ds^2 is specified by the metric function.[21]

continued

[20] Here we adopt the notation $x^\mu = (ct, \mathbf{r})$.

[21] This is sometimes stated as "in Riemannian geometry, coordinates have no metric significance." See in particular (5.5) and the introductory paragraph in Section 5.1.2, where the geodesic equation was first derived.

Box 6.2 *continued*

Thus we are free to choose new coordinates, thereby altering the metric to make it as simple as possible. Of course, in our particular case, the new coordinates should also have spherical symmetry.

1. Pick a new time coordinate to eliminate the cross term $dt\,dr$: we introduce a new coordinate t' such that

$$t' = t + f(r). \tag{6.51}$$

Differentiating, we have $dt' = dt + (df/dr)\,dr$. Squaring both sides and solving for dt^2,

$$dt^2 = dt'^2 - \left(\frac{df}{dr}\right)^2 dr^2 - 2\frac{df}{dr}\,dr\,dt. \tag{6.52}$$

Now the cross term $dt\,dr$ in (6.50) has a coefficient $Cr - 2D(df/dr)$, which can be eliminated[22] by choosing an $f(r)$ that satisfies the differential equation

$$\frac{df}{dr} = +\frac{Cr}{2D}. \tag{6.53}$$

Similarly, the $(df/dr)^2\,dr^2$ term in (6.52) can be absorbed into a new scalar function $B' = B - (df/dr)^2 D$.

2. Scale the radial coordinate to simplify the angular coefficient: We can set the function A in (6.50) to unity by choosing a new radial coordinate,

$$r'^2 = A(r,t)r^2, \tag{6.54}$$

so that the first term on the right-hand side of (6.50) is just $r'^2(d\theta^2 + \sin^2\theta\,d\phi^2)$. The effect of this new radial coordinate on the other terms in ds^2 can again be absorbed by the relabeling of scalar functions. We leave this as an exercise for the reader to complete.

With these new radial and time coordinates (relabeled as r and t), we are left with only two unknown scalar functions in the metric. In this way, the line element interval takes the form

$$ds^2 = g_{00}(r,t)c^2\,dt^2 + g_{rr}(r,t)\,dr^2 + r^2(d\theta^2 + \sin^2\theta\,d\phi^2), \tag{6.55}$$

yielding the metric matrix shown in (6.56).

[22] The absence of any linear factor dt means that the metric is also time-reversal invariant (i.e., changing t to $-t$ does not affect ds^2). In fact, one can see that the metric matrix should be diagonal, because spherical symmetry of the problem implies that the solution should be symmetric under the reflections of each coordinate: $r \to -r$, $\theta \to -\theta$, and $\phi \to -\phi$.

6.4.1 The geometry of a spherically symmetric spacetime

Box 6.2 demonstrated that the metric tensor for spherically symmetric spacetime has only two unknown scalar functions. For these particular coordinates[23] $(x^0 = ct, r, \theta, \phi)$, the functions are the tensor elements $g_{00}(t, r)$ and $g_{rr}(t, r)$. The metric is diagonal, with $g_{\theta\theta} = r^2$ and $g_{\phi\phi} = r^2 \sin^2\theta$:

$$g_{\mu\nu} = \begin{pmatrix} g_{00} & & & \\ & g_{rr} & & \\ & & r^2 & \\ & & & r^2 \sin^2\theta \end{pmatrix}. \tag{6.56}$$

[23] This system is called the Schwarzschild coordinate system. We will see other spherically symmetric coordinate systems in Chapter 7.

The metric functions $g_{00}(t, r)$ and $g_{rr}(t, r)$ can be obtained by solving the Einstein field equation. (cf. the penultimate subsection of Chapter 11). Its solution obtained by Schwarzschild is as follows:

$$g_{00}(t, r) = -\frac{1}{g_{rr}(t, r)} = -1 + \frac{r^*}{r}. \tag{6.57}$$

Thus the invariant interval is

$$ds^2 = -\left(1 - \frac{r^*}{r}\right) c^2\, dt^2 + \left(1 - \frac{r^*}{r}\right)^{-1} dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2. \tag{6.58}$$

There is warping in the radial as well as in the time coordinate. For points that are far away from the gravitational source, $g_{\mu\nu}$ approaches the flat-spacetime limit:

$$\lim_{r\to\infty} g_{00}(t, r) \to -1 \qquad \text{and} \qquad \lim_{r\to\infty} g_{rr}(t, r) \to 1. \tag{6.59}$$

Thus the deviation from flat Minkowski spacetime is measured by the ratio r^*/r. The constant length r^* can be fixed by our Newtonian-limit result (5.41):

$$g_{00} = -\left(1 + \frac{2\Phi(x)}{c^2}\right) = -\left(1 - \frac{r^*}{r}\right). \tag{6.60}$$

We then have, from (6.45),

$$r^* = \frac{2G_N M}{c^2}. \tag{6.61}$$

This Schwarzschild radius is generally a very small lengthscale. For example, the solar and terrestrial Schwarzschild radii are, respectively,

$$r^*_\odot \simeq 3 \text{ km} \qquad \text{and} \qquad r^*_\oplus \simeq 9 \text{ mm}. \tag{6.62}$$

For the Schwarzschild exterior solution to apply, the radial coordinate value r must be at least R, the radius of the source mass. Thus the respective maximum solar and terrestrial ratios (i.e., for the minimum r) are

$$\frac{r^*_\odot}{R_\odot} = O(10^{-6}) \qquad \text{and} \qquad \frac{r^*_\oplus}{R_\oplus} = O(10^{-9}), \tag{6.63}$$

which means that the GR modification r^*/r, which enters into (6.60), is very small. Nevertheless, such tiny corrections have been found to agree with observations whenever high-precision measurements have been performed.

Example 6.1 Curved spacetime and the operation of GPS

As the first application of the Schwarzschild geometry, let us revisit the time dilation corrections required for the proper operation of the GPS system, previously discussed in Section 4.3.2. Recall that the relevant issue is the comparison of the proper time rates recorded by the clock on the satellite ($d\tau_S$) and by the ground clock ($d\tau_\oplus$). Their relation can be efficiently deduced by a consideration of the relevant invariant intervals (i.e., proper time rates): $ds^2_S = -c^2\, d\tau^2_S$ and $ds^2_\oplus = -c^2\, d\tau^2_\oplus$. Since both clocks are in circular orbits (hence $dr = 0$), we have[24]

$$ds^2_\oplus = -\left(1 + 2\frac{\Phi_\oplus}{c^2}\right) c^2\, dt^2 + r^2\, d\phi^2 = -c^2\, d\tau^2_\oplus, \tag{6.64}$$

where t is the (faraway) coordinate time, and we have also plugged in $r^*/r = -2\Phi/c^2$. Dividing both sides by $-c^2\, dt^2$ and taking the square root yields,[25] for the terrestrial time interval,

$$\frac{d\tau_\oplus}{dt} = 1 + \frac{\Phi_\oplus}{c^2} - \frac{v^2_\oplus}{2c^2}, \tag{6.65}$$

where we have also substituted the tangential speed $v = r\, d\phi/dt$. We can carry out similar manipulations on the satellite time $d\tau_S$. In this way, we obtain their ratio

$$\frac{d\tau_\oplus}{d\tau_S} = 1 + \frac{\Phi_\oplus}{c^2} - \frac{\Phi_S}{c^2} - \frac{v^2_\oplus}{2c^2} + \frac{v^2_S}{2c^2}, \tag{6.66}$$

which is exactly the same result we obtained in (4.36). The previous derivation in Chapter 4 may have left the impression that the SR time dilation (due to relative motion) is distinct from GR predictions. But SR correction (as a kinematic result) is naturally included in GR; both effects should be contained in one general formulation.

[24] We choose to define the polar angle $\theta = \pi/2$ (hence $d\theta = 0$ and $sin\theta = 1$) for each orbit plane.

[25] Throughout our presentation, we often make the approximation of $(1 \pm x)^n \simeq 1 \pm nx$ when $x \ll 1$.

Interpreting the coordinates

Our spherically symmetric metric (6.56) is diagonal. In particular, $g_{i0} = g_{0i} = 0$ (with spatial indices $i = 1, 2, 3$), which means that for a given t, we can discuss the spatial subspace separately. To set up the coordinates, we can first imagine the situation where gravity is absent. For a fixed t, this spherically symmetric coordinate system can be visualized as a series of 2-spheres having different radial coordinate values of r, with their common center at the origin of the spherically symmetric source. Each 2-sphere has a surface area $4\pi r^2$ and can be pictured as a lattice of rigid rods arranged in a grid corresponding to various (θ, ϕ) values, with synchronized clocks attached at each grid point (i.e., each point in this subspace has the same coordinate time).

- In the absence of gravity,[26] the spacetime is flat as in Fig. 6.7(a). The coordinate r is the proper radial distance ρ defined as

$$dr^2 = ds_r^2 \equiv d\rho^2, \tag{6.67}$$

  where ds_r^2 is the invariant interval ds^2 with $dt = d\theta = d\phi = 0$. The coordinate t is the proper time τ for an observer at a fixed location:

$$-c^2\, dt^2 = ds_t^2 \equiv -c^2\, d\tau^2, \tag{6.68}$$

  where ds_t^2 is the invariant interval ds^2 with $dr = d\theta = d\phi = 0$. Thus, the coordinates (r, t) can be interpreted as the radial distance and time measured by an observer.

- When gravity is turned on, the spacetime is warped.[27] In the spatial radial direction, the proper radial distance no longer equals the radial coordinate:

$$d\rho = \sqrt{g_{rr}}\, dr. \tag{6.69}$$

  Consequently, the spherical surface area $4\pi r^2$ no longer bears the Euclidean relation with the proper radius ρ. Similarly, the proper time

$$d\tau = \sqrt{-g_{00}}\, dt \tag{6.70}$$

  differs from the coordinate time, because $g_{00} \neq -1$. This gravitational time dilation signifies the warping of the spacetime in the time direction.

An embedding diagram, which may help to visualize warped space, is discussed in Box 6.3.

[26] This also applies for an observer far from the gravitational source, where the spacetime is approximately flat. However, while dr is the measured distance between two radially separated points in the faraway region, r is not the radial distance to the origin, as the spacetime near the origin is curved.

[27] We recall that (6.69) and (6.70) are just the standard expressions for the metric elements in terms of defined coordinates and invariant intervals discussed in (5.6).

Box 6.3 Embedding diagram

A convenient way to visualize warped space is to use an embedding diagram. Since it is difficult to visualize the full three-dimensional curved space (embedded in a 4D space), we shall concentrate on the two-dimensional subspace corresponding to a fixed polar angle $\theta = \pi/2$ (and at some given instant of time). That is, we will focus on a 2D space slicing through the middle of the source. In the absence of gravity, this is just a flat plane as depicted in Fig. 6.7(a).

In the presence of gravity, this 2D surface is curved. We would like to visualize the warping of this 2D subspace. A useful technique is to imagine that the curved 2D surface is embedded in a fictitious 3D Euclidean space, as in Fig. 6.7(b). For the particular case of the external Schwarzschild geometry, the line element interval with $dt = d\theta = 0$ is

$$ds^2 = \left(1 - \frac{r^*}{r}\right)^{-1} dr^2 + r^2\, d\phi^2. \tag{6.71}$$

We now imagine this 2D surface as a subspace of a 3D Euclidean space with a cylindrical polar coordinate system: $x^i = (r, \phi, z)$. We have chosen the polar radial (perpendicular) distance from the cylindrical axis of symmetry (the z axis) of this 3D space to coincide with the Schwarzschild r coordinate, and the azimuthal angle to coincide with the Schwarzschild ϕ angle. Just as a line-curve is represented in 3D space by $x^i(\lambda)$, with a single curve parameter λ, a 2D surface is represented by $x^i(\lambda, \sigma)$, with two surface parameters. In our case, we naturally choose the Schwarzschild coordinates (r, ϕ) as parameters. The spherical symmetry of the Schwarzschild space implies a cylindrical symmetry of the embedding space; there is no ϕ dependence in the metric. Thus, the only nontrivial relation is $z(r)$, which can be worked out as follows. We have the Euclidean interval

$$ds^2 = dz^2 + dr^2 + r^2\, d\phi^2 = \left[\left(\frac{dz}{dr}\right)^2 + 1\right] dr^2 + r^2\, d\phi^2. \tag{6.72}$$

A comparison of (6.72) with (6.71) leads to $dz = \pm[r^*/(r - r^*)]^{1/2}\, dr$, which can be integrated to yield $z = \pm 2[r^*(r - r^*)]^{1/2}$, or

$$z^2 = 4r^*(r - r^*), \tag{6.73}$$

which is a sideways parabola in the (r, z) plane for any fixed ϕ. Sweeping the parabola in a circle around the z axis (to include all ϕ) yields the 3D representation of the curved surface as shown in Fig. 6.7(b).

Distances between points on this surface are equal to the proper distances in Schwarzschild space. Note that radial proper lengths are stretched (compared with the coordinate separation) near the throat at $r = r^*$. The surface does not extend to $r < r^*$, where the radial paths have timelike (negative) intervals. It is not quite correct to assume that slices of spacetime with constant t must be entirely spatial. We will see in Chapter 7 that different coordinates can better describe this region around a black hole, an object massive and dense enough to fit within its own Schwarzschild radius r^*.

Finally, note that the upper and lower (positive- and negative-z) halves of Fig. 6.7(b) provide redundant descriptions of Schwarzschild space. The redundant half could be used to describe an extra universe inside the Schwarzschild radius. But even if such a universe existed, it would be causally disconnected from ours.

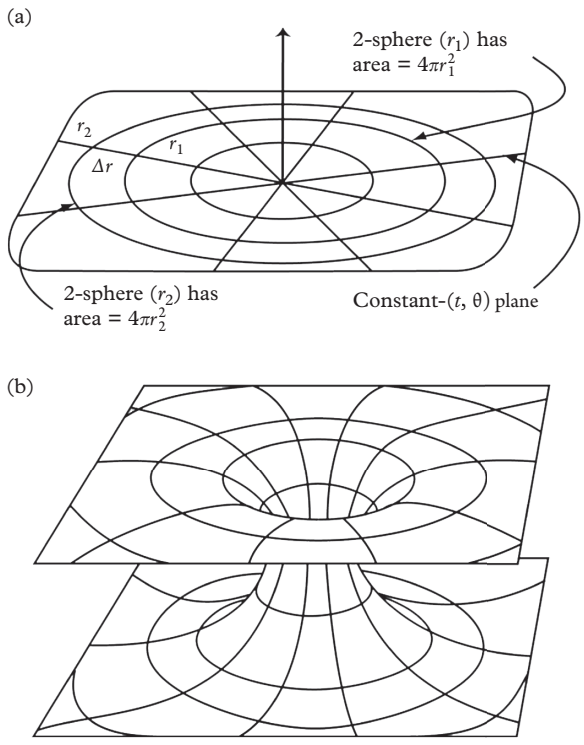

Figure 6.7 *(a) The $\theta = \pi/2$ plane (r, ϕ) cutting across the spherical source, displayed in a flat coordinate space without gravity. (b) In a fictitious 3D embedding space, the physical 2D subspace of (a) is shown as a curved surface, which truly portrays the proper distance between points in the Schwarzschild space (6.71).*

Spherically symmetric metric is time-independent

We note from (6.57) that the scalar metric functions are time-independent, even though we never assumed a static source. This turns out to be a general result (Birkhoff's theorem): whenever the source is spherically symmetric, the resultant exterior spacetime must be time-independent. We shall not provide a mathematical proof of this theorem, but will note that the same result holds for the Newtonian theory as well—as it should. Recall that the gravitational field outside a spherical source is identical to the gravitational field of a point source having all its mass concentrated at its center. This proof depends only on the spherical symmetry of the setup, irrespective of any possible time dependence. Thus, whether the spherical mass is pulsating or exploding or whatever, the resultant field is the same as long as the spherical symmetry is maintained. The analogous statement in electromagnetism is that there is no monopole radiation (only dipole radiation, quadrupole radiation, etc.)

6.4.2 Curved spacetime and deflection of light

In the time dilation effects on the operation of GPS (Example 6.1) and in the gravitational redshift (Example 5.1), the GR effects are manifestations of the

bending of time: $g_{00} \neq -1$. However, GR treats space and time on an equal footing, and involves bending in both the temporal and the spatial directions: $g_{00} \neq -1$ and $g_{rr} \neq 1$. Here we work out the GR result for light deflection, which goes beyond Einstein's 1911 calculation, in which only gravitational time dilation was taken into account.

To an observer far from the source, using the coordinate time and radial distance (t, r), the effective speed of light $c(r)$ can be related[28] to the locally measured universal light speed $c = d\rho/d\tau$ by

$$c(r) \equiv \frac{dr}{dt} = \frac{d\rho}{d\tau}\sqrt{\left|\frac{g_{00}(r)}{g_{rr}(r)}\right|} = c\sqrt{\left|\frac{g_{00}(r)}{g_{rr}(r)}\right|}, \tag{6.74}$$

where the proper distance and time (ρ, τ) are related to the coordinate distance and time (r, t) by (6.69) and (6.70). The Schwarzschild solution (6.57) informs us that $g_{rr} = -g_{00}^{-1}$. Thus, alongside the relation between proper and coordinate time (4.40), we also have a relation between the proper and coordinate length intervals:

$$d\rho = \left(1 - \frac{\Phi(r)}{c^2}\right) dr. \tag{6.75}$$

Clearly, besides gravitational time dilation, we also have a gravitational length contraction effect.[29] The influences of spatial and temporal warping on $c(r)$ in (6.74) are of the same size and in the same direction. The deviation of $c(r)$ from c (thus the change of index of refraction $n(r)$) is twice Einstein's 1911 result of taking into account only $g_{00} \neq -1$ as in (4.41):

$$\begin{aligned}[c(r)]_{\rm GR} \equiv \frac{dr}{dt} &= \frac{d\rho}{1-\Phi(r)/c^2}\frac{1+\Phi(r)/c^2}{d\tau} \\ &= \frac{1+\Phi(r)/c^2}{1-\Phi(r)/c^2}\frac{d\rho}{d\tau} = \left(1 + \frac{2\Phi(r)}{c^2}\right) c,\end{aligned} \tag{6.76}$$

According to (4.43)–(4.48), the deflection angle $\delta\phi$ is directly proportional to $d[c(r)]$, and thus is twice as large:

$$[\delta\phi]_{\rm GR} = 2[\delta\phi]_{\rm EP} = \frac{4G_{\rm N}M}{c^2 r_{\rm min}}. \tag{6.77}$$

The GR prediction verified in the 1919 solar eclipse expeditions If a light ray is deflected by an angle $\delta\phi$ as shown in Fig. 4.8, the light source at S will appear to the observer at O to be located at S'. Since the deflection is inversely proportional to $r_{\rm min}$, one can maximize the bending by minimizing $r_{\rm min}$. For light grazing the surface of the sun, $r_{\rm min} = R_\odot$ and $M = M_\odot$, the predicted solar deflection of light is $[\delta\phi]_{\rm GR} = 1.74''$.

[28] An alternative derivation is through a lightlike worldline, $ds^2 = g_{00}c^2\,dt^2 + g_{rr}\,dr^2 = 0$, in a fixed direction, $d\theta = d\phi = 0$. One may wonder about the validity of imposing this fixed angle in the calculation of the light deflection angle $\delta\phi$. Section 4.3.2 showed us that such a calculation requires two steps: first we calculate the speed of light $c(r)$ in terms of gravitational potential; then we relate the differential deflection $d\phi$ to this result for $c(r)$. To calculate $c(r)$, we can work with any propagation direction.

[29] Recalling the derivation given in Box 4.1 of gravitational time dilation via an arrangement of three identically constructed clocks, one may be tempted to derive gravitational length contraction from the SR length contraction with a similar arrangement involving three identically constructed rods. Such a derivation was criticized in (Rindler 1968) because it inevitably assumes some spacetime geometry, making such a derivation before the introduction of a curved spacetime description of gravity less useful even for pedagogical purposes.

The deflection angle (about 1/4000 of the angular width of the sun as seen from earth), predicted by Einstein's equations in 1915, was not easy to detect. One needed a solar eclipse against a background of several bright stars (so that some could be used as reference points). On May 29, 1919, there was an solar eclipse. Two British expeditions were organized by A.S. Eddington (1882–1944) and the Astronomer Royal, F.W. Dyson (1868–1939): one to Sobral in northern Brazil and the other to the island of Principe off the coast of West Africa. The report that Einstein's prediction was successful in these tests created a worldwide sensation—partly for scientific reasons and partly because the world was amazed that so soon after the "Great War" (World War I) the British should finance and conduct an expedition to test a theory proposed by a German citizen.[30]

[30] Eddington was a Quaker conscientious objector. With the backing of Dyson, he was able to continue his research during the war, and, through Willem de Sitter in neutral Holland, obtained Einstein's GR papers in 1916.

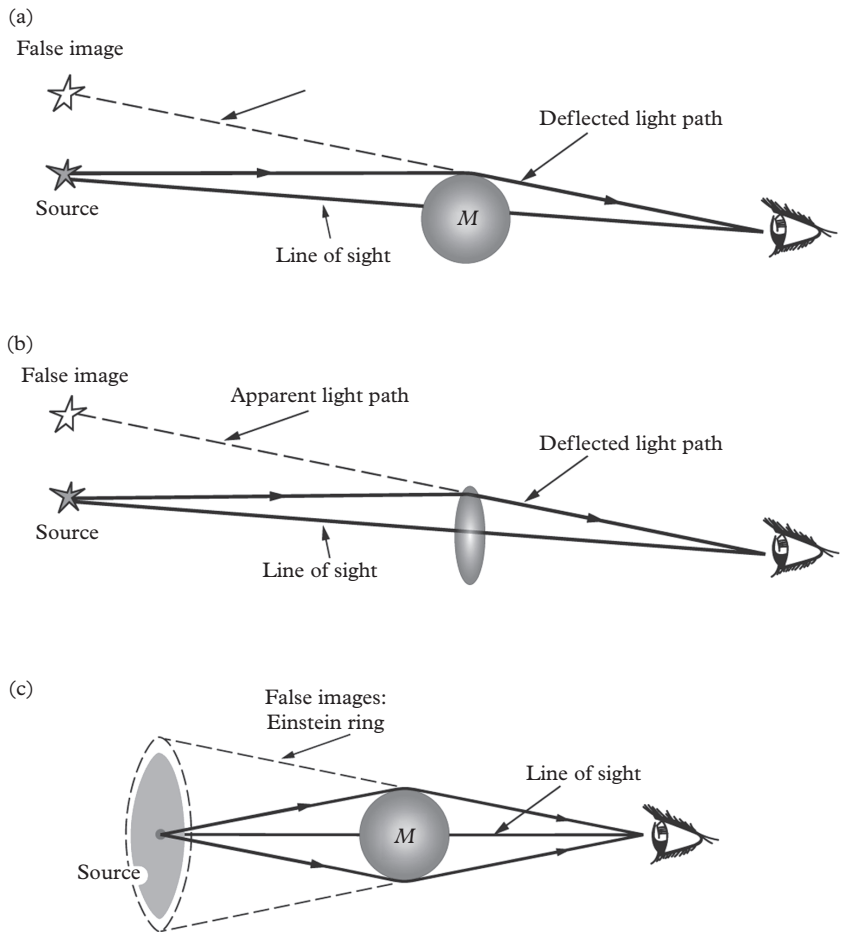

Figure 6.8 *Bending of a light ray as a lensing effect. (a) Light from a distant star is deflected by a lensing mass M lying close to the line of sight from the observer to the source. As a result, the source star appears to be located at a different position. (b) The bending of light in (a) is analogous to that caused by a glass lens. (c) If the line of sight passes directly through the center of a symmetrical lensing mass distribution, the false image appears as a ring, the* Einstein ring. *If the alignment is not perfect or the lensing mass is not spherical, the images may be arcs, as seen in Fig. 6.9.*

Figure 6.9 *Gravitational lensing effects due to the galaxy cluster Abell 2218. Nearly all of the bright objects in this picture taken by the Hubble Space Telescope are galaxies in this cluster, which is so massive and so compact that it lenses the light from galaxies that lie behind it into multiple images that appear as long, faint arcs. The image is reproduced with permission from (Fruchter et al. 2001) NASA/ERO - STScI and ST-ECF Team.*

Gravitational lensing The gravitational deflection of a light ray discussed above (Fig. 6.8a) has some similarity to the bending of light by a glass lens (Fig. 6.8b). When the source and observer are sufficiently far from the lensing mass, bending from both sides of the mass can produce separate images[30a]—or even a ring image (Fig. 6.8c) if the source, lensing mass, and observer are perfectly aligned. Thus one generally finds stretched arc images of the light source (see, e.g., the photograph in Fig. 6.9). In fact, gravitational lensing has become a powerful tool of modern astrophysics in mapping out astronomical mass distributions. In cosmology, the theoretical predictions are mostly given in terms of mass distributions rather than, say, the distributions of stars and galaxies. Thus lensing results are particularly useful. For an example showing that such distributions can be different, see the discussion of the Bullet Cluster of galaxies in Chapter 8; cf. Fig. 8.4. The key input in the various lensing equations is the deflection result (6.77).

[30a] More spectacularly, the image of a distant bright source (such as a supernova) if lensed by a galactic cluster can appear to us as multiple events separated in time, because the different optical paths it takes to reach our telescope.

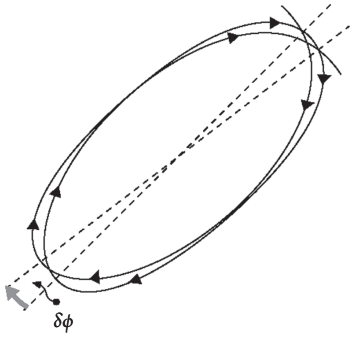

Figure 6.10 *A perturbed $1/r^2$ attraction leads to an open elliptical orbit that may be described as an elliptical orbit with a precessing axis. For planetary motion, this is usually stated as the precession of the point of closest approach to the sun, the perihelion.*

6.4.3 Precession of Mercury's orbit

In this section, we shall discuss the motion of a test mass in Schwarzschild spacetime. A particle under the Newtonian $1/r^2$ gravitational attraction has an elliptical orbit. Here we calculate the GR correction to such a trajectory.

Celestial mechanics based on the Newtonian theory of gravitation has been remarkably successful. However, it was realized around 1850 that there was a discrepancy between the theory and the observed precession of the perihelion of the planet Mercury. The pure $1/r^2$ force law of Newton predicts a closed elliptical orbit for a planet, i.e., an orbit with its major axis fixed in space. However, perturbations due to the presence of other planets and astronomical objects lead to a trajectory that is no longer closed. Since the perturbation is small, such a deviation from the closed orbit can be described as an elliptical orbit with a precessing axis, or equivalently a precessing perihelion (the point of the orbit closest to the sun); see Fig. 6.10.

For the case of Mercury, such planetary perturbations could account for most[31] of its perihelion advance of 574″ per century. However, there was still a discrepancy of 43″ left unaccounted for. Following a similar situation involving Uranus that eventually led to the prediction and discovery of the outer planet Neptune in 1846, a new planet, named Vulcan, was predicted to lie inside Mercury's orbit; however, it was never found. This is the perihelion precession problem that Einstein solved by applying his new theory of gravitation. As we shall presently see, GR implies a small correction to the $1/r^2$ force law, which just accounts for the missing 43″ per century advance of Mercury's orbit.

[31] Most of the raw observational value of 5600″ per century is due to the effect of the rotation of our earth-based coordinate system. This leaves the planetary perturbations to account for the remaining 574″, to which Venus contributes 277″, Jupiter 153″, Earth 90″, and Mars and the rest of the planets 10″.

Lagrangian of the geodesic equation

A particle's motion in a curved spacetime such as the Schwarzschild spacetime is determined by the GR equation of motion, the geodesic equation. As discussed in Section 5.1.2, this is the Euler–Lagrange equation

$$\frac{\partial L}{\partial x^\mu} = \frac{d}{d\tau}\frac{\partial L}{\partial \dot{x}^\mu}, \tag{6.78}$$

where the Lagrangian is the particle's 4-velocity squared,

$$L(x, \dot{x}) = g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu. \tag{6.79}$$

As is appropriate for a massive test particle, we have picked its proper time τ to be the curve parameter and have used the notation $\dot{x}^\mu = dx^\mu/d\tau$. The dependence of L on the coordinates x^μ is through the metric tensor $g_{\mu\nu}$.

As in Newtonian mechanics, because angular momentum is conserved under the action of a central force, the trajectory will always remain in the plane spanned by the particle's initial velocity and the radial vector $\mathbf{r}$ connecting the force center to the test particle. Let us call this the $\theta = \pi/2$ plane. By inserting the Schwarzschild metric, the Lagrangian (6.79) becomes

$$L(x, \dot{x}) = -\left(1 - \frac{r^*}{r}\right) c^2 \dot{t}^2 + \left(1 - \frac{r^*}{r}\right)^{-1} \dot{r}^2 + r^2 \dot{\phi}^2. \tag{6.80}$$

In principle, one can obtain the orbit $r(\phi)$ by solving the equation of motion (6.78). Since this equation involves the second-order time derivative, it is simpler if we proceed via the energy balance equation, which involves only the first derivative. The energy balance equation, as will be demonstrated below, is simply the normalization condition of the Lagrangian (i.e., the 4-velocity invariant (6.79)):[32]

$$L(x, \dot{x}) = -c^2. \tag{6.81}$$

[32] Even though this $L = -c^2$ result was first obtained in (3.29) in the context of a flat spacetime, it is still valid for a general metric $g_{\mu\nu}$, as the proper time is always defined by $-c^2\, d\tau^2 \equiv g_{\mu\nu}\, dx^\mu\, dx^\nu$. Dividing both sides by $d\tau^2$ immediately leads to (6.81).

Two constants of motion

Our next step is to express the various terms of (6.80) in terms of constants of motion (i.e., conserved quantities along the particle's path). If the metric $g_{\mu\nu}$ is independent of a particular coordinate, say x^α (for a particular index α), then $\partial L/\partial x^\alpha = 0$. This can be translated, through (6.78), into a conservation statement: $d(\partial L/\partial \dot{x}^\alpha)/d\tau = 0$. In the present case of planetary motion on a fixed plane ($d\theta = 0$), the Lagrangian (6.80) does not depend on t or ϕ; we have two constants of motion. The metric's independence of t leads to the conservation of $\partial L/\partial \dot{x}^0 = \partial(g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu)/\partial\dot{x}^0 \sim g_{00}\dot{x}^0$; we will call this conserved quantity

$$\kappa \equiv -g_{00}c\dot{t} = \left(1 - \frac{r^*}{r}\right)c\dot{t}. \tag{6.82}$$

The independence of ϕ points to the conservation of $\partial L/\partial\dot{\phi} \sim g_{\phi\phi}\dot{\phi}$:

$$l \equiv m g_{\phi\phi}\dot{\phi} = mr^2\dot{\phi}. \tag{6.83}$$

[33] As discussed in Section 3.2.2, the key attribute of energy and momentum is that they are conserved quantities and have the correct flat-spacetime and Newtonian limits; cf. (3.40) and Example 3.1.

Clearly κ is to be identified with the particle's energy[33] (in units of mc), as in the flat-spacetime limit ($r^*/r \to 0$) it reduces directly to the SR expression (3.40) of energy $c\dot{t} = E/mc$; l is the orbital angular momentum $|\mathbf{r}\times\mathbf{p}| = mr^2\dot{\phi}$.

Energy balance equation

Substituting the constants l and κ into (6.80), we can then write the $L = -c^2$ equation (6.81) as

$$-\frac{\kappa^2}{1 - r^*/r} + \frac{\dot{r}^2}{1 - r^*/r} + \frac{l^2}{m^2r^2} = -c^2. \tag{6.84}$$

After multiplying by $\frac{1}{2}m(1 - r^*/r)$, and using $r^* = 2G_NM/c^2$ on the right-hand side, this may be written as

$$-\frac{m\kappa^2}{2} + \frac{1}{2}m\dot{r}^2 + \left(1 - \frac{r^*}{r}\right)\frac{l^2}{2mr^2} = -\frac{1}{2}mc^2\left(1 - \frac{2G_NM}{c^2r}\right), \tag{6.85}$$

or

$$\frac{1}{2}m\dot{r}^2 + \left(1 - \frac{r^*}{r}\right)\frac{l^2}{2mr^2} - \frac{G_NmM}{r} = \frac{m\kappa^2}{2} - \frac{1}{2}mc^2. \tag{6.86}$$

[34] In Schwarzschild spacetime, as κ is the particle energy E/mc, $\mathcal{E}$ can be identified as the corresponding kinetic energy: $\mathcal{E} = m(\kappa^2 - c^2)/2 = (E^2 - m^2c^4)/2mc^2$, because $E = (m^2c^4 + 2mc^2\mathcal{E})^{1/2} \simeq mc^2 + \mathcal{E}$ in the Newtonian limit when $\mathcal{E} \ll E$.

Replacing the constant κ by another energy quantity[34]

$$\frac{\mathcal{E}}{m} = \frac{\kappa^2 - c^2}{2} \tag{6.87}$$

simplifies the right-hand side of (6.86):

$$\frac{1}{2}m\dot{r}^2 + \left(1 - \frac{r^*}{r}\right)\frac{l^2}{2mr^2} - \frac{G_NmM}{r} = \mathcal{E}. \tag{6.88}$$

Except for the $(1 - r^*/r)$ factor, this is just the energy balance equation for the nonrelativistic central force problem: (radial and rotational) kinetic energy plus potential energy equals the total (Newtonian) energy $\mathcal{E}$. The extra term,

$$-\frac{r^*}{r}\frac{l^2}{2mr^2} = -\frac{G_N M l^2}{mc^2 r^3}, \tag{6.89}$$

may be regarded here as a small GR correction to the Newtonian potential energy $-G_N mM/r$ by a $1/r^4$ type of force.

The precessing orbit

This relativistic energy equation (6.88) can be cast in the form of an orbit equation. We can solve for $r(\phi)$ using standard perturbation theory (see Box 6.4). With e being the eccentricity of the orbit, $\alpha = l^2/G_N Mm^2 = (1 + e)r_{\min}$, and $\epsilon = 3r^*/2\alpha$, the solution is

$$r = \frac{\alpha}{1 + e\cos[(1-\epsilon)\phi]}. \tag{6.90}$$

Thus the planet returns to its perihelion $r_{\min}$ not at $\phi = 2\pi$ but at $\phi = 2\pi/(1-\epsilon) \simeq 2\pi + 3\pi r^*/\alpha$. The perihelion advances (i.e., the whole orbit rotates in the same sense as the planet itself) per revolution by (Fig. 6.10)

$$\delta\phi = \frac{3\pi r^*}{\alpha} = \frac{3\pi r^*}{(1+e)r_{\min}}. \tag{6.91}$$

With the solar Schwarzschild radius $r^*_\odot = 2.95\,\text{km}$, Mercury's eccentricity $e = 0.206$, and its perihelion $r_{\min} = 4.6 \times 10^7$ km, the numerical value of the advance is

$$\delta\phi = 5.0 \times 10^{-7} \text{ radians per revolution} \tag{6.92}$$

or $5.0 \times 10^{-7}\,\text{rad} \times (180^\circ/\pi\,\text{rad}) \times (60'/\text{deg}) \times (60''/\text{min}) = 0.103''$ per revolution. In terms of the advance per century,

$$0.103'' \times \frac{100 \text{ years/century}}{\text{Mercury's period of } 0.241 \text{ years}} = 43'' \text{ per century}. \tag{6.93}$$

This agrees with the observational evidence.

This calculation explaining the perihelion advance of the planet Mercury from first principles, and the correct prediction for the bending of starlight around the sun, were both obtained by Einstein in an intense two-week period in November, 1915. This gave Einstein great of joy.[34a] This moment of elation was characterized by his biographer Abraham Pais as "by far the strongest emotional experience in Einstein's scientific life, perhaps, in all his life". Afterward, he wrote to Arnold Sommerfeld in a now-famous letter:

[34a] "For a few days, I was besides myself with joyous excitement," Einstein wrote.

This last month I have lived through the most exciting and the most exacting period of my life; and it would be true to say this, it has been the most fruitful. Writing letters has been out of the question. I realized that up until now my field equations of gravitation have been entirely devoid of foundation. When all my confidence in the old theory vanished, I saw clearly that a satisfactory solution could only be reached by linking it with the Riemann variations. The wonderful thing that happened then was that not only did Newton's theory result from it, as a first approximation, but also the perihelion motion of Mercury, as a second approximation. For the deviation of light by the Sun I obtained twice the former amount.

We have already discussed the doubling of the light deflection angle. Recall that for several years Einstein had struggled with the Ricci-tensor-based field equation that could not reproduce the Newtonian limit. The Riemann variation to which he referred above is the linear combination that we now called the Einstein tensor. The mathematics involving the curvature tensor can be found in Chapter 11.

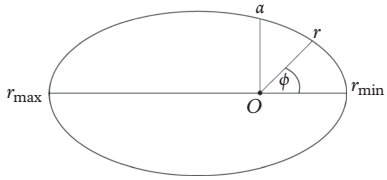

Figure 6.11 *Points on an elliptical orbit around a mass at the origin O are labeled by their coordinates* (r, ϕ)*, with some notable positions at* $(r_{\min}, 0)$*,* $(r_{\max}, \pi)$*, and* $(\alpha, \pi/2)$*.*

Box 6.4 The orbit equation and its perturbation solution

We solve the relativistic energy equation (6.88) as a standard central force problem. The relevant kinematic variables are shown in Fig. 6.11.

The orbit equation

To obtain the orbit $r(\phi)$, we first change all the time derivatives into derivatives with respect to the angle ϕ using the angular momentum equation (6.83):

$$d\tau = \frac{mr^2}{l}\, d\phi. \tag{6.94}$$

We then make the change of variable $u \equiv 1/r$ (and thus $u' \equiv du/d\phi = -u^2\, dr/d\phi = -m\dot{r}/l$). In this way, (6.88) turns into

$$u'^2 + u^2 - \frac{2}{\alpha}u - r^* u^3 = C, \tag{6.95}$$

where $\alpha = l^2/(G_N M m^2)$ and the constant $C = 2m\mathcal{E}/l^2$. This is the equation we need to solve in order to obtain the planet's orbit $r(\phi)$.

Zeroth-order solution

Split the solution $u(\phi)$ into an unperturbed part u_0 and a small (of the order of $r^* u^2$) correction u_1: $u = u_0 + u_1$, with

$$u_0'^2 + u_0^2 - \frac{2}{\alpha}u_0 = C. \tag{6.96}$$

This unperturbed orbit equation can be solved by differentiating with respect to ϕ and dividing the resulting equation by $2u_0'$:

$$u_0'' + u_0 = \alpha^{-1}, \tag{6.97}$$

which is a simple harmonic oscillator equation in the variable $u_0 - \alpha^{-1}$, with ϕ replacing the time variable and $\omega = 1$ the angular frequency. It has the solution $u_0 - \alpha^{-1} = D\cos\phi$. We choose to write the constant $D \equiv e/\alpha$, so that the unperturbed solution takes on the well-known form of a conic section,

$$r = \frac{\alpha}{1 + e\cos\phi}. \tag{6.98}$$

It is clear (see Fig. 6.11) that $r = \alpha/(1+e) = r_{\min}$ (perihelion) at $\phi = 0$, and $r = \alpha/(1-e) = r_{\max}$ (aphelion) at $\phi = \pi$. Geometrically, e is called the eccentricity of the orbit. The radial distance at $r(\phi = \pi/2) = \alpha$ can be expressed in terms of perihelion (or aphelion) and eccentricity as

$$\alpha = (1+e)r_{\min} = (1-e)r_{\max}. \tag{6.99}$$

Relativistic correction

We now plug the full solution $u = u_0 + u_1$ into the perturbed orbit equation (6.95):

$$(u_0' + u_1')^2 + (u_0 + u_1)^2 - \frac{2}{\alpha}(u_0 + u_1) - r^*(u_0 + u_1)^3 = C, \tag{6.100}$$

and separate out the leading and the next-to-leading terms (with $u_1 = O(r^*u^2) = O(r^*/\alpha^2)$):

$$\left(u_0'^2 + u_0^2 - \frac{2}{\alpha}u_0 - C\right) + \left(2u_0'u_1' + 2u_0u_1 - \frac{2}{\alpha}u_1 - r^*u_0^3\right) = 0, \tag{6.101}$$

where we have dropped the higher-order terms proportional to u_1^2, $u_1'^2$, or $u_1 r^*$. After using (6.96), we can then pick out the first-order equation:

$$2u_0'u_1' + 2u_0u_1 - \frac{2}{\alpha}u_1 = r^*u_0^3, \tag{6.102}$$

where

$$u_0 = \frac{1 + e\cos\phi}{\alpha}, \qquad u_0' = -\frac{e}{\alpha}\sin\phi. \tag{6.103}$$

The equation for u_1 is then

$$-e\sin\phi\,\frac{du_1}{d\phi} + e\cos\phi\,u_1 = \frac{r^*(1 + e\cos\phi)^3}{2\alpha^2}. \tag{6.104}$$

continued

Box 6.4 *continued*

One can verify that it has the solution

$$u_1 = \frac{r^*}{2\alpha^2}\left[(3+2e^2) + \frac{1+3e^2}{e}\cos\phi - e^2\cos^2\phi + 3e\phi\sin\phi\right]. \qquad (6.105)$$

The first two terms have the form of the zeroth-order solution, $A + B\cos\phi$; thus they represent unobservably small corrections. The third term, whose period (in ϕ) is exactly half that of the unperturbed orbit, is likewise unimportant. We only need to concentrate on the ever-increasing fourth term,[35] whose effects accumulate over many orbits. Plugging this into $u = u_0 + u_1 = 1/r$, we obtain

$$r = \frac{\alpha}{1 + e\cos\phi + \epsilon e\phi\sin\phi}, \qquad (6.106)$$

where $\epsilon = 3r^*/2\alpha$ is a small quantity. By approximating $\cos\epsilon\phi \simeq 1$ and $\sin\epsilon\phi \simeq \epsilon\phi$, so that

$$e(\cos\phi + \epsilon\phi\sin\phi) \simeq e\cos(\phi - \epsilon\phi), \qquad (6.107)$$

we can put the solution (6.106) in the more transparent form shown in (6.90).

[35] While the detection of the higher-order effects represented by the first three terms would require measurements at impossibly high accuracy, the fourth term, while equally small, is a new effect (a correction to a zero) that is much easier to measure.

In Section 4.3.2, we derived using Huygens' method the gravitational angular deflection $\delta\phi$ of a light ray according to the equivalence principle (accounting for the curvature only in the time direction). In Exercise 6.3, you are asked to obtain $\delta\phi$ by following the more standard procedure of directly applying the geodesic equation—just as we have done in calculating the precession of an elliptical orbit. The steps to solve these two problems are the same. But for the light geodesic $x^\mu(\tau)$, the curve parameter τ cannot be the proper time. Still we can choose to define it so that the light's 4-momentum $p^\mu = \dot{x}^\mu \equiv dx^\mu/d\tau$, which is a null 4-vector. Thus the geodesic Lagrangian (6.79), $L = p_\mu p^\mu = 0$, is the energy balance equation, instead of $L = -c^2$ as in (6.81) for a particle with mass.

Exercise 6.3 Light deflection from solving the geodesic equation

Take the following steps to obtain the result for the bending of light shown in (6.77):

(a) Identify the two constants of motion.

(b) Express the $L = 0$ equation in terms of these constants.

(c) By changing the curve parameter differential $d\tau \to d\phi$ and changing the radial distance variable to its inverse $u \equiv 1/r$, you should find that the light trajectory obeys[36]

$$u'' + u - \epsilon u^2 = 0, \tag{6.108}$$

where $u'' = d^2u/d\phi^2$ and $\epsilon = 3r^*/2$.

(d) Solve (6.108) by perturbation:[37] *$u = u_0 + \epsilon u_1$. Suggestion: Parameterize the first-order perturbation solution as $u_1 = \alpha + \beta \cos 2\phi$; then fix the constants α and β.*

(e) From this solution of the orbit $r(\phi)$ for the light trajectory, one can deduce the angular deflection $\delta\phi$ result (6.77) by comparing the directions of the initial and final asymptotes ($r = \infty$ at $\phi_i = \pi/2 + \delta\phi/2$ and $\phi_f = -\pi/2 - \delta\phi/2$) as shown in Fig. 6.12(b).

[36] This is the massless particle version of (6.95).

[37] The unperturbed solution u_0 leads to the trajectory as shown in Fig. 6.12(a).

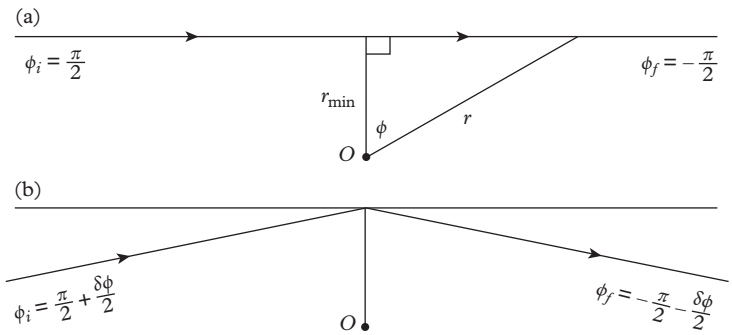

Figure 6.12 *(a) Light's geodesic trajectory is straight, $r = r_{\min}/\cos\phi$, in the absence of gravity. (b) The light geodesic is bent by an angle $\delta\phi$ by a source mass at O. Thus the trajectory has asymptotes ($r = \infty$) at $\phi_i = \pi/2 + \delta\phi/2$ and $\phi_f = -\pi/2 - \delta\phi/2$ or $\phi_f = 3\pi/2 - \delta\phi/2$.*

Review questions

1. A curved surface necessarily has a position-dependent metric. Why is this not a sufficient condition? Illustrate this point with an example of a flat space with a position-dependent metric. Gaussian curvature K provides a good criterion for finding out whether a surface is curved or not. Show that $K = 0$ for your example, and explain why $K = 1/R^2$ for a spherical surface, regardless of what coordinates one uses.

2. What are the three surfaces of constant curvature? Write out the embedding equation of a 3-sphere in a 4D space, as well as the equation of a 3-pseudosphere.

Is the embedding space for the latter a Euclidean space?

3. The Gaussian curvature measures how curved a space is, because it controls the amount of deviation from Euclidean relations. What is angular excess? How is it related to Gaussian curvature? Give a simple example of a polygon on a spherical surface that clearly illustrates this relation. Give another example of such a non-Euclidean relation, showing that the deviation from flatness is proportional to the curvature.

4. What are tidal forces? How are they related to the gravitational potential? Explain how in GR the tidal forces are identified with the curvature of spacetime.

5. Give a qualitative description of the GR field equation. Describe its tensor structure and its various terms: what are its source term and curvature term? What is the solution to this field equation?

6. What is the form of the spacetime metric (when written in terms of the spherical Schwarzschild coordinates) for a spherically symmetric spacetime? Explain very briefly how such a spacetime is curved in space as well as in time.

7. Present a simple proof of Birkhoff's theorem for Newtonian gravity. Explain how it implies that there is no monopole radiation.

8. Write down the metric elements for the Schwarzschild spacetime (i.e., the Schwarzschild solution). If the metric element g_{00} is related to the gravitational potential by $g_{00} = -(1 + 2\Phi/c^2)$ in the Newtonian limit, demonstrate that the Newtonian result $\Phi = -G_{\rm N}M/r$ is contained in this Schwarzschild solution.

9. A light ray is bent in the presence of a gravitational field. How does the feature $g_{rr} = -g_{00}^{-1}$ in the Schwarzschild metric lead to a bending of the light ray in GR that is twice as much as predicted by Einstein's 1911 EP calculation?

10. Explain qualitatively how GR causes the perihelion of an elliptical orbit to precess.

11. Suppose the metric does not depend on a certain coordinate, say the ϕ angle of the (t, r, θ, ϕ) coordinates. Use the geodesic equation to show that this leads to a constant of motion $\sim g_{\phi\phi}\dot{\phi}$.