General Relativity as a Geometric Theory of Gravity

5

- By focusing on the equivalence principle (EP), Einstein discovered that the gravitational field can be modeled as a curved spacetime. The mathematics of such a curved manifold is that of Riemannian geometry. We present in Section 5.1 some of its basic elements: Gaussian coordinates and the metric tensor.
- We use the calculus of variations to deduce the geodesic equation for curves of minimum length (i.e., extreme invariant spacetime interval). We also discuss how this equation describes a straight line.
- We next present a geometric description of the EP physics of gravitational time dilation. In this geometric theory, the metric $g_{\mu\nu}(x)$ plays the role of a relativistic gravitational potential.
- The identification of curved spacetime as the gravitational field naturally suggests that spacetime's geodesic equation is the GR equation of motion, which reduces to the Newtonian equation of motion in the limit of a nonrelativistic moving particle ($v \ll c$) in a static and weak gravitational field. This also clarifies how Newton's theory is extended by Einstein's general-relativistic theory of gravitation.

The road from the EP to GR can be viewed as follows. The equivalence of an accelerated frame to one with gravity means that we cannot say for sure that gravity causes a particle's acceleration. We could just as easily attribute the acceleration to some property of the reference frame itself. Thus Einstein proposed a geometric theory modeling gravity as a warping of spacetime. From this point of view, there is no gravitational force. Particles move freely through spacetime with gravity; however, the motion of such particles following straight paths in a warped spacetime may be nontrivial. Gravitational phenomena thereby reflect the curvature of spacetime.

In this chapter, we mainly study the GR equation of motion, the geodesic equation, which describes the motion of a test particle in a curved spacetime.

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

In Chapter 6, we will take up the GR field equation, the Einstein equation, which describes how a mass/energy source gives rise to a curved spacetime. But before presenting the geometric gravitational theory itself, we first provide a short mathematical introduction to some of the basic elements of a metric description of a curved space.

5.1 Metric description of a curved manifold

Differential geometry is the branch of mathematics that uses calculus techniques to study geometry. Its sub-branch Riemannian geometry concerns in particular the non-Euclidean n-dimensional spaces that can be described by distance measurements. Since most of us can only visualize, and only have any familiarity with, curved surfaces of two dimensions, we shall often use this simpler case, which was pioneered by Carl Friedrich Gauss (1777–1855), to illustrate the more general theory. The extension of Gauss's theory to higher dimensions was first[1] studied by his student, Bernhard Riemann (1826–1866).

[1] Non-Euclidean geometry was independently discovered by János Bolyai (1802–1860) and Nikolai Lobachevsky (1792–1856).

5.1.1 Gaussian coordinates and the metric tensor

Gaussian coordinates

Many of us intuitively think of a curved surface as a 2D surface embedded in 3D Euclidean space, for example a spherical surface (of radius R) described in 3D Cartesian coordinates (X, Y, Z) by

$$X^2 + Y^2 + Z^2 = R^2. \tag{5.1}$$

More generally, the embedding coordinates are subject to a constraint condition, $f(X, Y, Z) = 0$. This is an extrinsic geometric description; the space of interest (here the curved surface) is described using entities outside the space. We are most interested in an intrinsic geometric description, a characterization of the physical space without such external reference. Specifically, we would like to describe a 2D surface solely through measurements made by an inhabitant who never leaves that surface. Gauss introduced a generalized parametrization whose coordinates (x^1, x^2) are free to range over their respective domains without constraint:

$$X = X(x^1, x^2), \quad Y = Y(x^1, x^2), \quad Z = Z(x^1, x^2). \tag{5.2}$$

The Gaussian coordinates (x^1 and x^2, the number of which corresponds to the dimensionality of the embedded space) make the embedding coordinates (X, Y, and Z) superfluous; hence the geometric description can be purely intrinsic.

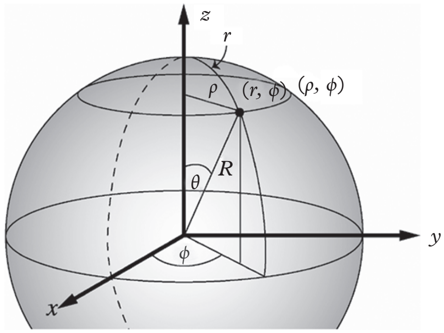

Figure 5.1 *Gaussian coordinates* (θ, ϕ) *and* (ρ, ϕ) *for the curved surface of a 2-sphere. The dashed line is the prime meridian. Entirely equivalent to the* (θ, ϕ) *system are the coordinates* (r, ϕ) *with the radial coordinate* $r = R\theta$ *measured from the north pole on the surface of the sphere (N.B., not from the center of the sphere).*

For the space of a 2-sphere of radius R, described extrinsically by (5.1) and illustrated in Fig. 5.1, using a 3D Euclidean embedding space, we provide two examples of Gaussian coordinate systems:

1. The polar coordinate system: To set up a Gaussian coordinate system $(x^1, x^2) = (\theta, \phi)$ to label points in this 2D surface, first pick a point on the surface (the north pole) and a longitudinal great circle through the pole (the prime meridian). The radial coordinate r of a point is the arclength between the point and the north pole. Instead of r, one can equivalently use the coordinate $\theta = r/R$ (the polar angle[2] or colatitude). The azimuthal angle ϕ (i.e., the longitude) is measured against the prime meridian. The coordinate domains are $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$. In this case, (5.2) is specified by
$$X = R\sin\theta\cos\phi, \quad Y = R\sin\theta\sin\phi, \quad Z = R\cos\theta. \tag{5.3}$$
One may easily verify that this parametrization and the following one, (5.4) are consistent with the extrinsic description, (5.1).
2. The cylindrical coordinate system: We can choose another set of Gaussian coordinates to label points in the 2D surface by defining a different radial coordinate $\rho = R\sin\theta$, with a domain $0 \leq \rho \leq R$. Thus $(x^1, x^2) = (\rho, \phi)$, so that (5.2) is specified by[3]
$$X = \rho\cos\phi, \quad Y = \rho\sin\phi, \quad Z = \pm\sqrt{R^2 - \rho^2}. \tag{5.4}$$

From now on, we will no longer use extrinsic coordinates such as (X, Y, Z). By coordinates, we shall always mean Gaussian coordinates such as (x^1, x^2) that label points on a 2D space. Since one can choose from any number of coordinate systems, and geometric relations should be independent of such choices, a proper formulation of geometry must be invariant under general coordinate transformations.

[2] It may be helpful to visualize the coordinate θ of a point as the angle formed (at the center of the sphere in the embedding space) between the polar (z) axis and the radial (R) axis as displayed in Fig. 5.1. Of course, the entire point of the intrinsic description is to discard the embedding space. The surface itself has no center. When we extend this method to higher dimensions or pseudo-Euclidean spaces, such mental/visual crutches will serve less well in any case.

[3] If the spherical surface is embedded in a 3D Euclidean space, ρ is interpreted as the perpendicular distance to the Z axis as shown in Fig. 5.1. Perhaps the term cylindrical coordinates becomes more understandable if, instead of ρ, we use Z directly: $(x^1, x^2) = (Z, \phi)$, where $\rho^2 = R^2 - Z^2$. While $Z = R\cos\theta$, with a domain of $-R \leqslant Z \leqslant R$, covers all latitudes, ρ covers only half of the sphere.

Exercise 5.1 Coordinate choice

Clearly the ideal choice of coordinate system often depends on the task at hand. Consider the calculation in the space of a 2D plane of the circumference of a circle of radius R ($2\pi R$ of course). It is easy in polar coordinates (r, θ), but rather complicated in Cartesian coordinates (x, y). Carry out the calculations in both coordinate systems.

It must be emphasized that the coordinates $\{x^a\}$ generally do not form a vector space (elements of which can be added and multiplied by scalars, etc.). They are labels of points in the curved space and are devoid of any physical significance in their own right. Cartesian coordinates in flat space are the exception. We learn in our first physics course that the displacement between two distant points is a vector. Indeed, we have already applied rotation and boost transformations to coordinate displacements. We cannot do this in curved spaces! However, we may treat infinitesimal displacements as vectors, because they reside in a flat space, as we will discuss in the context of the flatness theorem (Section 5.1.3).

The metric tensor

The basic idea of Riemannian geometry is that the geometry (angles, lengths, and shapes) of a space can be described by length measurements. To illustrate this for the case of a 2D spherical surface (of radius R), one first sets up a Gaussian coordinate system (e.g., polar or cylindrical coordinates) to label points on the globe, then measures the infinitesimal distances between neighboring points (Fig. 5.2). These length measurements are summarized in an entity called the metric g_{ab} (whose indices range over the coordinates). It relates length measurements to differentials in the chosen Gaussian coordinates at any given point in the space by (3.19), which may be written as a matrix product:

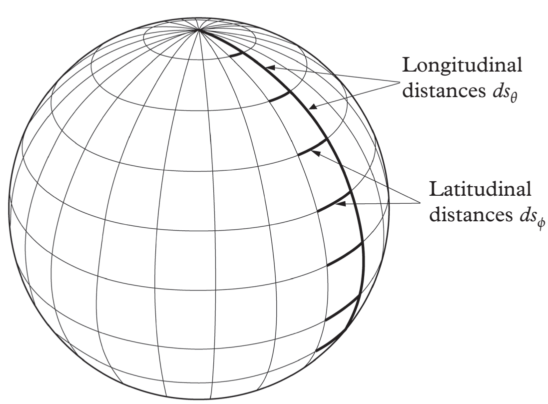

Figure 5.2 *Using distance measurements along links of constant longitude (ds_θ) and latitude (ds_ϕ) to specify the shape of the spherical surface.*

$$\begin{aligned} ds^2 &= g_{ab}\, dx^a\, dx^b \\ &= \begin{pmatrix} dx^1 & dx^2 \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \begin{pmatrix} dx^1 \\ dx^2 \end{pmatrix} \\ &= g_{11}(dx^1)^2 + 2g_{12}\, dx^1\, dx^2 + g_{22}(dx^2)^2. \end{aligned} \tag{5.5}$$

Note that the metric is always a symmetric matrix ($g_{12} = g_{21}$), because $dx^a\, dx^b = dx^b\, dx^a$. Also recall that the metric is directly related to the basis vectors of the space: $g_{ab} = \mathbf{e}_a \cdot \mathbf{e}_b$, cf. (3.8). ds^2 is formally interpreted as the infinitesimal squared length.[4]

The metric is an intrinsic geometric quantity Once the coordinate system has been chosen, the elements of the metric at a point $x = (x^1, x^2)$ can then be determined by infinitesimal distance measurements between x and nearby points. If we choose to measure the length ds_1 along the dx^1 direction (i.e., $dx^2 = 0$), then (5.5) reduces to $(ds_1)^2 = g_{11}(dx^1)^2$. Similarly, $(ds_2)^2 = g_{22}(dx^2)^2$. Let the measured length between x and the nearby point $(x^1 + dx^1, x^2 + dx^2)$ be ds_{12}. In this way, without invoking any extrinsic embedding space, the metric elements (g_{11}, g_{22}, g_{12}) at x can all be expressed in terms of measured lengths (ds_1, ds_2, ds_{12}) and coordinate differentials (dx^1, dx^2):

$$\begin{aligned} g_{11} &= \frac{{ds_1}^2}{(dx^1)^2}, \quad g_{22} = \frac{{ds_2}^2}{(dx^2)^2}, \\ g_{12} &= \frac{ds_{12}^2 - ds_1^2 - ds_2^2}{2\, dx^1\, dx^2} = -\frac{ds_1\, ds_2 \cos\alpha}{dx^1\, dx^2}. \end{aligned} \tag{5.6}$$

To reach the last equality, we have used the law of cosines $ds_{12}^2 = ds_1^2 + ds_2^2 - 2\, ds_1\, ds_2 \cos\alpha$, where α is the angle subtended by the axes. We expect that the off-diagonal metric elements should describe the deviation of the basis vectors from orthogonality ($g_{12} \equiv \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_2 \sim \cos\alpha$); thus, if the coordinates are orthogonal ($\alpha = \pi/2$), the metric matrix must be diagonal ($g_{12} = 0$). We emphasize again that the coordinates $\{x^a\}$ themselves do not measure distance. Only through the metric as in (5.5) are they connected to distance measurements.

As we have already noted prior to this subsection, because an infinitesimally small area on a curved surface can be treated as a flat plane (per the flatness theorem), flat-space geometric relations such as the law of cosines and the Pythagorean theorem apply.

The metrics for a 2-sphere Here we will work out a concrete example by writing down the metric for a 2-sphere in the polar coordinate system. One finds that the latitudinal distances $ds_\phi = R \sin\theta\, d\phi$ (subtended by some fixed $d\phi$ between two points having the same radial distance/latitude, $ds_\theta = 0$) become ever smaller as one approaches the poles ($\theta \to 0, \pi$). Meanwhile, the longitudinal distances ds_θ (subtended by $d\theta$ between two points having the same longitude $d\phi = 0$) have the same value, $ds_\theta = R\, d\theta$, over the whole range of θ and ϕ. (See Fig. 5.2.)

[4] Equation (5.5) is understood to mean $ds^2 = \sum_{a,b} g_{ab}\, dx^a\, dx^b$; i.e., the Einstein summation convention has been employed. For Cartesian coordinates in a Euclidean space, it is simply the Pythagorean theorem, $ds^2 = dx^2 + dy^2 + \cdots$. Recall the result worked out in Exercise 3.1 that a contraction between symmetric and antisymmetric tensors vanishes. Thus, if the metric had an antisymmetric part $g^A_{ab} = -g^A_{ba}$, it would not contribute to the length, because $g^A_{ab}\, dx^a\, dx^b = 0$.

Such distance measurements completely describe this spherical surface. These distance measurements can be compactly expressed in terms of the metric tensor elements. Because we have chosen orthogonal coordinates, $g_{\theta\phi} = \mathbf{e}_\theta \cdot \mathbf{e}_\phi = 0$, the infinitesimal length between two nearby points with coordinate displacement $(d\theta, d\phi)$ can be expressed using the Pythagorean theorem as

$$\begin{aligned}[ds^2]_{(\theta,\phi)} &= (ds_\theta)^2 + (ds_\phi)^2 \\ &= R^2\, d\theta^2 + R^2 \sin^2\theta\, d\phi^2. \end{aligned} \tag{5.7}$$

Matching terms of (5.7) and (5.5) yields the metric tensor for this (θ, ϕ) coordinate system:

$$g_{ab}^{(\theta,\phi)} = R^2 \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix}. \tag{5.8}$$

Exercise 5.2 Cylindrical coordinate metric

Find the metric tensor for the cylindrical coordinates (ρ, ϕ) on a 2-sphere.

Suggestion: *From Fig. 5.1, note that the radial coordinate is related to the polar angle by $\rho = R\sin\theta$; then show that*

$$g_{ab}^{(\rho,\phi)} = \begin{pmatrix} R^2/(R^2 - \rho^2) & 0 \\ 0 & \rho^2 \end{pmatrix}. \tag{5.9}$$

Coordinate transformation in a curved space

A key difference between curved and flat spaces is that curved space must necessarily have position-dependent coordinate bases, while a flat space can have constant (Cartesian) coordinates. As a consequence, the metric and coordinate transformation matrices are position-dependent in any curved space.[5]

[5] The contrast between flat and curved spaces will also be discussed in Section 6.1, as well as in the introductory paragraph of Chapter 11. For examples of a transformation as a matrix of partial derivatives, see Exercise 2.4 and Box 5.2.

Because a flat space can still have curvilinear coordinates (such as polar coordinates), the main features of coordinate transformation can be illustrated in a flat space with a curvilinear coordinate system. Take the simplest case of a 2D plane. Consider the transformation from Cartesian coordinates $\{x^a\} = (x, y)$ to polar coordinates $\{x^a\} \to \{x'^a\} = (r, \theta)$. A coordinate transformation can in general be written as a matrix of partial derivatives:

$$dx^a = [\Lambda]^a{}_b dx'^b \qquad \text{with} \qquad [\Lambda]^a{}_b = \frac{\partial x^a}{\partial x'^b}; \tag{5.10}$$

this expression follows from an application of the chain rule of differentiation. In this 2D example, taking derivatives of the relations $x = r\cos\theta$ and $y = r\sin\theta$ leads to

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \begin{pmatrix} dr \\ d\theta \end{pmatrix}. \tag{5.11}$$

The elements of the transformation matrix are derivatives, for example,

$$[\Lambda]^1{}_1 = \cos\theta = \frac{\partial x}{\partial r} = \frac{\partial x^1}{\partial x'^1} \quad \text{and} \quad [\Lambda]^1{}_2 = \frac{\partial x^1}{\partial x'^2}. \tag{5.12}$$

We note that, unlike the rotation matrix (1.5) or the Lorentz transformation (2.12), the transformation matrix elements $[\Lambda]^a{}_b$ are position-dependent—here they depend on $\{x'^a\} = (r, \theta)$.

Exercise 5.3 Transformation in curved space

Find the coordinate transformation matrix $[\Lambda]$ *(showing its coordinate dependence) that changes the polar coordinates* (θ, ϕ) *to the cylindrical coordinates* (ρ, ϕ) *on a 2-sphere:*

$$\begin{pmatrix} d\rho \\ d\phi \end{pmatrix} = [\Lambda] \begin{pmatrix} d\theta \\ d\phi \end{pmatrix}. \tag{5.13}$$

5.1.2 The geodesic equation

As previously stated, in Riemannian geometry, the spatial geometry is determined by (infinitesimal) length measurements, which are codified in the metric tensor. Namely, once the (Gaussian) coordinate system has been chosen, the metric elements $g_{ab}(x)$ relate the coordinate differentials to the measured lengths at all x. In this way, the geometry of the space can be determined from the metric.[6] Here we shall work out such an example: how to find, from the metric, the equation that describes the shortest curve between two fixed points in a space.

Any curve in a space can be written in the form[7] $x^a(\tau)$, where τ is some parameter (which might be, but need not be, proper time) having some range, for example $[\tau_i, \tau_f]$. We are interested in finding, for a given space, the shortest curve, called the geodesic, that connects initial $x^a(\tau_i)$ and final $x^a(\tau_f)$ positions. (A discussion of the geodesic as a straight line can be found in Box 5.2.) The squared length (invariant interval) ds^2 of an infinitesimal segment of any curve is given by (5.5). We integrate $ds = \sqrt{|ds^2|}$ (the length of a spacelike segment or the proper time of a timelike one) along this curve, changing to Greek indices as is conventional for 4D spacetime (also denoting $\dot{x}^\mu = dx^\mu/d\tau$):

$$s = \int ds = \int \sqrt{|g_{\mu\nu} dx^\mu dx^\nu|} = \int L(x, \dot{x})\, d\tau, \tag{5.14}$$

[6] A note of caution: while the metric determines the geometry, geometry may not fix the metric. For example, a metric with constant elements describes a space with zero curvature, but a space with no curvature does not necessarily imply a constant metric. More specifically, a flat plane can be described by polar coordinates, whose metric is position-dependent.

[7] Namely, as the parameter τ varies $(\tau_1, \tau_2, \tau_3, \ldots)$, one obtains a continuous set of coordinates $x^a(\tau_1), x^a(\tau_2), x^a(\tau_3), \ldots$. An example of such a parametric description of a curve would be the trajectory of a particle parameterized by its time.

where

$$L(x, \dot{x}) = \sqrt{|g_{\mu\nu}\dot{x}^{\mu}\dot{x}^{\nu}|}. \tag{5.15}$$

To determine the shortest line in the curved space, we impose the extremization condition[8] for variation of the path with endpoints fixed:

$$\delta s = \delta \int L(x, \dot{x})\, d\tau = 0, \tag{5.16}$$

which the calculus of variations can translate (cf. Box 5.1) into a partial differential equation—the Euler–Lagrange equation:

$$\frac{d}{d\tau}\frac{\partial L}{\partial \dot{x}^{\mu}} - \frac{\partial L}{\partial x^{\mu}} = 0. \tag{5.17}$$

[8] In contrast to the extremization of a function $f(x)$, in which a single variable x is varied resulting in the condition $df/dx = 0$, here one varies an entire function $x(\tau)$ (all possible curves running from $x(\tau_i)$ to $x(\tau_f)$), resulting in the condition (5.16), which yields the Euler–Lagrange equation (5.17).

Box 5.1 Euler–Lagrange equation from action extremization

In physics, one often uses the calculus of variations to formulate equations of motion and field equations from the least-action principle. Namely, these equations are derived as the Euler–Lagrange equations from the extremization of some action integral:

$$S = \int_{\tau_1}^{\tau_2} L(x, \dot{x}) d\tau. \tag{5.18}$$

The integrand, called the Lagrangian, is the difference between the kinetic energy and the potential energy for a classical system parameterized by time:

$$L(x, \dot{x}) = T(\dot{x}) - V(x). \tag{5.19}$$

The principle of least action states that the action is extremal with respect to the variation of the trajectory $x^{\mu}(\tau)$ with its endpoints fixed at initial and final positions $x^{\mu}(\tau_1)$ and $x^{\mu}(\tau_2)$. This extremization requirement can be translated into a partial differential equation as follows. The variation of the Lagrangian is

$$\delta L(x, \dot{x}) = \frac{\partial L}{\partial x^{\mu}}\,\delta x^{\mu} + \frac{\partial L}{\partial \dot{x}^{\nu}}\,\delta \dot{x}^{\nu}. \tag{5.20}$$

Thus the condition for extremization[9] of the action integral becomes

$$\begin{aligned} 0 = \delta S = \delta \int_{\tau_1}^{\tau_2} L(x, \dot{x})\, d\tau &= \int_{\tau_1}^{\tau_2} \left(\frac{\partial L}{\partial x^{\mu}}\,\delta x^{\mu} + \frac{\partial L}{\partial \dot{x}^{\nu}}\frac{d}{d\tau}\delta x^{\nu} \right) d\tau \\ &= \int_{\tau_1}^{\tau_2} \left(\frac{\partial L}{\partial x^{\mu}} - \frac{d}{d\tau}\frac{\partial L}{\partial \dot{x}^{\mu}} \right) \delta x^{\mu}\, d\tau. \end{aligned} \tag{5.21}$$

[9] Extremization means minimization or maximization. This allows us to avoid the square roots and absolute values in (5.14) and (5.15) by integrating ds^2 (rather ds) as our action, cf. (5.23). For timelike curves such as the trajectories of a particle with mass, the action is negative and maximized (least negative) by the geodesic path. Taylor and Wheeler (2000) gave it the evocative name “the principle of extremal aging.”

To reach the last expression, we have performed an integration by parts on the second term:

$$\int_{\tau_1}^{\tau_2} \frac{\partial L}{\partial \dot{x}^\nu}\, d(\delta x^\nu) = \left[\frac{\partial L}{\partial \dot{x}^\nu}\,\delta x^\nu(\tau)\right]_{\tau_1}^{\tau_2} - \int_{\tau_1}^{\tau_2} (\delta x^\nu)\, d\left(\frac{\partial L}{\partial \dot{x}^\nu}\right). \tag{5.22}$$

The first term on the right-hand side can be discarded, because the endpoint positions are fixed: $\delta x^\nu(\tau_1) = \delta x^\nu(\tau_2) = 0$. Since δS must vanish for arbitrary variations $\delta x^\mu(\tau)$, the expression in parentheses on the right-hand side of (5.21) must also vanish. The result is the Euler–Lagrange equation (5.17). For the simplest case of $L = \frac{1}{2}m\dot{x}^2 - V(\mathbf{x})$, the Euler–Lagrange equation is just the familiar $\mathbf{F} = m\mathbf{a}$ equation, as it yields $m\ddot{\mathbf{x}} + \nabla V = 0$.

The geodesic determined by the Euler–Lagrange equation As a mathematical exercise, one can show that the same Euler–Lagrange equation (5.17) following from (5.15) follows as well from a Lagrangian of the form

$$L(x, \dot{x}) = g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu, \tag{5.23}$$

which, without the square root, is much easier to work with. With L in this form, the derivatives in (5.17) become

$$\frac{\partial L}{\partial \dot{x}^\mu} = 2g_{\mu\nu}\dot{x}^\nu, \qquad \frac{\partial L}{\partial x^\mu} = \frac{\partial g_{\lambda\rho}}{\partial x^\mu}\dot{x}^\lambda\dot{x}^\rho, \tag{5.24}$$

where we have used the fact that the metric function $g_{\mu\nu}$ depends on x^μ but not on $\dot{x}^\mu$. Substituting these relations back into (5.17), we obtain the geodesic equation,

$$\frac{d}{d\tau}(g_{\mu\nu}\dot{x}^\nu) - \frac{1}{2}\frac{\partial g_{\lambda\rho}}{\partial x^\mu}\dot{x}^\lambda\dot{x}^\rho = 0, \tag{5.25}$$

which determines the geodesic, the curve that extremizes the invariant interval (the spatial path of the shortest length or the timelike trajectory of maximal proper time) between two points.

Exercise 5.4 Geodesics on simple surfaces

Use the geodesic equation (5.25) to confirm the familiar results that the geodesic is (a) a straight line on a flat plane and (b) a great circle on a spherical surface.

Suggestion: *For case (b), working out the full parametrization can be complicated; just check that the great circle given by ϕ = constant and $\theta = \alpha + \beta\tau$ solves the relevant geodesic equation.*

Casting the geodesic equation into standard form We can rewrite (5.25) in a more symmetric form. Differentiating the first term and noting that the metric's dependence on the curve parameter τ is entirely through $x^\mu(\tau)$, we have

$$g_{\mu\nu}\frac{d^2x^\nu}{d\tau^2}+\frac{\partial g_{\mu\rho}}{\partial x^\lambda}\frac{dx^\lambda}{d\tau}\frac{dx^\rho}{d\tau}-\frac{1}{2}\frac{\partial g_{\lambda\rho}}{\partial x^\mu}\frac{dx^\lambda}{d\tau}\frac{dx^\rho}{d\tau}=0. \tag{5.26}$$

In the second term, we have relabeled the dummy index $\nu\to\rho$; also, its coefficient can be decomposed into a symmetric (with respect to the exchange of λ and ρ indices) and an antisymmetric term:

$$\frac{\partial g_{\mu\rho}}{\partial x^\lambda}=\frac{1}{2}\left(\frac{\partial g_{\mu\rho}}{\partial x^\lambda}+\frac{\partial g_{\mu\lambda}}{\partial x^\rho}\right)+\frac{1}{2}\left(\frac{\partial g_{\mu\rho}}{\partial x^\lambda}-\frac{\partial g_{\mu\lambda}}{\partial x^\rho}\right). \tag{5.27}$$

Since the product $(dx^\lambda/d\tau)(dx^\rho/d\tau)$ in the second term on the left-hand side of (5.26) is symmetric, the antisymmetric part will not survive their contraction (cf. Exercise 3.1). In this way, the geodesic equation (5.25) becomes

$$g_{\mu\nu}\frac{d^2x^\nu}{d\tau^2}+\frac{1}{2}\left(\frac{\partial g_{\mu\rho}}{\partial x^\lambda}+\frac{\partial g_{\mu\lambda}}{\partial x^\rho}-\frac{\partial g_{\lambda\rho}}{\partial x^\mu}\right)\frac{dx^\lambda}{d\tau}\frac{dx^\rho}{d\tau}=0. \tag{5.28}$$

We can remove the first metric $g_{\mu\nu}$ factor by contracting the whole equation with the inverse metric $g^{\mu\sigma}$. In this way, the geodesic equation can now be written in its standard form,

$$\frac{d^2x^\sigma}{d\tau^2}+\Gamma^\sigma_{\lambda\rho}\frac{dx^\lambda}{d\tau}\frac{dx^\rho}{d\tau}=0, \tag{5.29}$$

where

$$\Gamma^\sigma_{\lambda\rho}=\frac{1}{2}g^{\sigma\mu}\left(\frac{\partial g_{\mu\rho}}{\partial x^\lambda}+\frac{\partial g_{\mu\lambda}}{\partial x^\rho}-\frac{\partial g_{\lambda\rho}}{\partial x^\mu}\right). \tag{5.30}$$

The set $\Gamma^\mu_{\lambda\rho}$ defined by this particular combination of the first derivatives of the metric tensor are called the Christoffel symbols (also known as the affine or connection coefficients). Recall that the metric is directly related to the coordinate bases, $g_{\mu\nu}=\mathbf{e}_\mu\cdot\mathbf{e}_\nu$, as in (3.8). $\Gamma^\mu_{\nu\lambda}$ are nonvanishing because the bases, and hence the metric, are position-dependent. They are called *symbols,* because, despite their appearance with indices, $\Gamma^\mu_{\nu\lambda}$ are not tensor elements.[10] Namely, they do not transform as tensor components under a coordinate transformation, cf. (3.27). Clearly, (5.29) is applicable for all higher-dimensional spaces (just by varying the range of the indices). Of particular relevance to us, this geodesic equation for 4D spacetime turns out to be the equation of motion in the GR theory of gravitation.

[10] As we shall demonstrate in Chapter 11, in particular (11.84), the geodesic equation (5.29) is a proper tensor equation, even though $\Gamma^\mu_{\lambda\rho}$ and the first derivative term have extra terms in their transformation. But these extra terms mutually cancel.

Box 5.2 Geodesic as a straight line

In the above, we have introduced the geodesic as the shortest line between two fixed endpoints in a curved space by a variational calculation. Here we will provide another view of (5.29) as the equation describing a straight line in coordinate systems with position-dependent bases.

A curved space must necessarily have position-dependent coordinates. In a flat space, it is possible to have constant coordinate bases (Cartesian coordinates), but we can still have curvilinear coordinates such as polar coordinates. Thus this second interpretation of the geodesic equation can be illustrated by the case of a straight line in a flat 2D plane. In the Cartesian system $\{x^a\} = (x, y)$, a straight line is the curve $x^a(\tau)$ with the tangent $(dx^a/d\tau)$ unchanged along the curve (see, e.g., Fig. 11.2a):

$$\frac{d^2x^a}{d\tau^2} = 0. \tag{5.31}$$

We now transform this equation to another system such as polar coordinates $\{x'^a\} = (r, \theta)$. The coordinate derivatives (with respect to the curve parameter) transform in the same way as the coordinate differentials of (5.10):

$$\frac{dx^a}{d\tau} = \frac{\partial x^a}{\partial x'^b}\frac{dx'^b}{d\tau}. \tag{5.32}$$

In this way, the left-hand side of (5.31) can be written as

$$\frac{d}{d\tau}\left(\frac{dx^a}{d\tau}\right) = \frac{d}{d\tau}\left(\frac{\partial x^a}{\partial x'^b}\frac{dx'^b}{d\tau}\right) = \frac{\partial x^a}{\partial x'^b}\frac{d^2x'^b}{d\tau^2} + \frac{d}{d\tau}\left(\frac{\partial x^a}{\partial x'^b}\right)\frac{dx'^b}{d\tau}. \tag{5.33}$$

The last term contains

$$\frac{d}{d\tau}\left(\frac{\partial x^a}{\partial x'^b}\right) = \frac{\partial^2 x^a}{\partial x'^b \partial x'^c}\frac{dx'^c}{d\tau} \tag{5.34}$$

with the first factor on the right-hand side being the position derivatives of the transformation matrix,

$$\frac{\partial^2 x^a}{\partial x'^b \partial x'^c} = \frac{\partial}{\partial x'^b}\left(\frac{\partial x^a}{\partial x'^c}\right), \tag{5.35}$$

which are nonvanishing because the transformation is position-dependent when the bases are position-dependent. The straight-line equation thus takes the form

$$\frac{\partial x^a}{\partial x'^b}\frac{d^2x'^b}{d\tau^2} + \frac{\partial^2 x^a}{\partial x'^b \partial x'^c}\frac{dx'^b}{d\tau}\frac{dx'^c}{d\tau} = 0. \tag{5.36}$$

In order to compare this straight-line equation with the geodesic equation (5.29), we multiply it by the *inverse* coordinate transformation $\partial x'^b/\partial x^c$, with the property $(\partial x'^b/\partial x^c)(\partial x^a/\partial x'^b) = \delta^a_c$. After relabeling some indices, we can write (5.36) as

$$\frac{d^2x'^a}{d\tau^2} + \left[\frac{\partial x'^a}{\partial x^d}\frac{\partial^2 x^d}{\partial x'^b \partial x'^c}\right]\frac{dx'^b}{d\tau}\frac{dx'^c}{d\tau} = 0. \tag{5.37}$$

continued

Box 5.2 *continued*

This is the geodesic equation (5.29) when we identify the square bracket as the Christoffel symbols (5.30): [. . .] = Γ^a_{bc}. We will not work out their explicit correspondence, but only remark that both expressions are related to the moving bases—in Γ^a_{bc}, we have the derivatives of the metric, and in [. . .], we have derivatives of the coordinate transformation. Although our discussion has been carried out in a flat space, the key ingredient is the position dependence of the coordinates. This property is a necessary feature of a curved space, while it is optional in a flat space. Thus this demonstration that the geodesic equation (5.29) describes a straight line is also valid in a curved space. The proper proof, involving the concepts of covariant differentiation and parallel transport, will be presented in Section 11.1.

5.1.3 Local Euclidean frames and the flatness theorem

A different choice of coordinates leads to a different metric, which is generally position-dependent. What distinguishes a flat space from a curved one is that for a flat space it is possible to find a coordinate system for which the metric is a constant, like Cartesian coordinates in Euclidean space with $[g] = [1]$ or in the Minkowski space of SR with $[g] = \mathrm{diag}(-1, 1, 1, 1) \equiv [\eta]$.

While it is clear that flat and curved spaces are different geometric entities, they are closely related. We are familiar from our experience with curved surfaces that any curved space can be approximated locally by a flat plane. This is the content of the flatness theorem.

In a curved spacetime with a general coordinate system x^μ and a metric value $g_{\mu\nu}$ at a given point P, it is always possible to find a coordinate transformation $x^\mu \to \bar{x}^\mu$ and $g_{\mu\nu} \to \bar{g}_{\mu\nu}$ such that the metric is flat at P (which can be taken to be the origin of the transformed coordinates, $P \to 0$), with $\bar{g}_{\mu\nu} = \eta_{\mu\nu}$ and $\partial\bar{g}_{\mu\nu}/\partial\bar{x}^\lambda = 0$; thus

$$\bar{g}_{\mu\nu}(\bar{x}) = \eta_{\mu\nu} + \gamma_{\mu\nu\lambda\rho}(0)\bar{x}^\lambda\bar{x}^\rho + \cdots . \tag{5.38}$$

Namely, the metric in the neighborhood of the origin (P) will differ from $\eta_{\mu\nu}$ only by the second- and higher-order derivatives. This is simply a Taylor series expansion of the metric at the origin; there is the constant $\bar{g}_{\mu\nu}(0)$ plus second-order derivative terms $\gamma_{\mu\nu\lambda\rho}(0)\bar{x}^\lambda\bar{x}^\rho$. That $\bar{g}_{\mu\nu}(0) = \eta_{\mu\nu}$ should not be a surprise. For a metric value at one point, it is always possible to find an orthogonal system so that $\bar{g}_{\mu\nu}(0) = 0$ for $\mu \neq \nu$. The diagonal elements can be scaled to unity so that the new coordinate bases all have unit length. Thus the metric is an identity matrix or whatever is the correct orthogonal flat metric with the appropriate signature. The nontrivial content of (5.38) is the absence of the first derivative.

In short, only in a flat manifold is it possible to have a constant metric for the entire space. However, in a curved space, it is still possible to have local Euclidean frames $\{\bar{x}^\mu\}$. The flatness theorem informs us that the general spacetime metric

$g_{\mu\nu}(x)$ is characterized at a point (P) not so much by the value $g_{\mu\nu}|_P$, since that can always be chosen to be its flat-space value, $\bar{g}_{\mu\nu}|_P = \eta_{\mu\nu}$, or by its first derivative, which can always be chosen to vanish, $\partial\bar{g}_{\mu\nu}/\partial\bar{x}^\lambda|_P = 0$, but by the second derivative of the metric, $\partial^2\bar{g}_{\mu\nu}/\partial x^\lambda\partial x^\rho$, which characterizes the curved space. It is related to the curvature of the space, to be discussed in Chapter 6.

5.2 From the equivalence principle to a metric theory of gravity

How did Einstein get the idea for a geometric theory of gravitation? What does one mean by a geometric theory?

A geometric physics theory

By a geometric theory or a geometric description of any physical phenomenon, we mean a theory that attributes the results of physical measurements directly to the underlying geometry of space and time. This can be illustrated by the description of a spherical surface (Fig. 5.2) that we discussed earlier in this chapter. The length measurements on the surface of a globe are different in different directions: the east/west distances between pairs of points separated by the same azimuthal angle $\Delta\phi$ are smaller for pairs farther from the equator, while the north/south lengths for a given $\Delta\theta$ are all the same. We could, in principle, interpret such results in two equivalent ways:

1. Without considering that the 2D space is curved, we could say that the physics (i.e., dynamics) is such that the measuring ruler changes its scale at different points or when pointing in different directions—much in the same manner as the Lorentz–FitzGerald length contraction of SR was originally interpreted.
2. Alternatively, we could use a standard ruler with a fixed scale, and attribute the varying length measurements to the underlying geometry of a curved spherical surface per Fig. 5.2. This geometry is expressed mathematically by a position-dependent metric tensor $g_{ab}(x) \neq \delta_{ab}$.

EP physics and a warped spacetime

In Chapter 4, we deduced several physical consequences from the empirical principle of gravity–inertia equivalence. In a geometric theory, these gravitational phenomena are attributed to the underlying curved spacetime, which has a metric as defined in (5.5):

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu. \tag{5.39}$$

For SR, we have the geometry of a flat spacetime with a position-independent metric $g_{\mu\nu} = \eta_{\mu\nu} \equiv \mathrm{diag}(-1, 1, 1, 1)$. The study of EP physics led Einstein to propose that gravity represents the structure of a curved spacetime, with $g_{\mu\nu} \neq \eta_{\mu\nu}$, and that gravitational phenomena are just the effects of curved spacetime on a test object. For instance, gravitational time dilation just reflects the bending of spacetime in the time direction.

Gravitational time dilation due to $g_{00} \neq -1$ For gravitational time dilation, instead of working with a complicated scheme of clocks running at different rates, this physical phenomenon can be interpreted geometrically as showing the presence of a nontrivial metric. Namely, we can interpret (4.40) in terms of a nontrivial metric. Recall our discussion relating the metric elements to the defined coordinates and distance measurements, cf. (5.6); the time–time element of the metric can be fixed by

$$g_{00} = \frac{ds_0^2}{(dx^0)^2}, \tag{5.40}$$

where $dx^0 = c\,dt$ is the coordinate time interval and $ds_0^2 = [ds^2]_{dx^i=0}$ is the measured interval in the time direction. The coordinate time is measured by the clock at a location far from the source of gravity ($\Phi = 0$), while $ds_0^2 = -c^2\,d\tau^2$ is directly related to the proper time $\tau(x)$ measured by stationary clocks located at x. Thus (5.40) is simply the relation $d\tau^2 = -g_{00}\,dt^2$, which, by (4.40), implies

$$g_{00} = -\left(1 + \frac{\Phi(x)}{c^2}\right)^2 \simeq -\left(1 + \frac{2\Phi(x)}{c^2}\right). \tag{5.41}$$

The metric element g_{00} deviates from its flat-spacetime value of $\eta_{00} = -1$ because of the presence of gravity. Thus the geometric interpretation of gravitational time dilation is that gravity warps spacetime (in this case in the time direction), changing the spacetime metric element from a constant $\eta_{00} = -1$ to an x-dependent function $g_{00}(x)$.

5.2.1 Curved spacetime as gravitational field

We provide further arguments for identifying warped spacetime as the gravitational field. Adopting a geometric interpretation of EP physics, we find that the resultant spacetime geometry has characteristics of a warped manifold such as a position-dependent metric and deviations from Euclidean geometric relations. Moreover, at every location, we can always transform gravity away to obtain a locally inertial spacetime, just as one can always find a locally flat region in a curved space (per the flatness theorem).

Position-dependent metric

The metric tensor in a curved space is necessarily position-dependent as in (5.41). In Einstein's geometric theory of gravitation, the metric function completely describes the gravitational field. The metric $g_{\mu\nu}(x)$ plays the role of the relativistic gravitational potential, just as $\Phi(x)$ is the Newtonian gravitational potential.

Non-Euclidean relations

In a curved space, Euclidean relations no longer hold. For example, the ratio of a circle's circumference to its radius may differ from 2π. The EP also implies non-Euclidean relations among geometric measurements. We illustrate this with a simple example. Consider a cylindrical room in high-speed rotation around its axis. The centripetal acceleration of the reference frame, according to the EP, is equivalent to a centrifugal gravitational field. (This is one way to produce artificial gravity.) For such a rotating frame, one finds that, because of SR (longitudinal) length contraction, the circular circumference of the cylinder is no longer equal to 2π times the radius, which does not change because the frame's velocity is transverse to the radial direction (see Fig. 5.3). Thus Euclidean geometry fails in the presence of gravity. We reiterate this connection: the rotating frame, according to the EP, is a frame with gravity; the rotating frame, according to SR length contraction, has a relation between its radius and circumference that is not Euclidean. Hence we say that in the presence of gravity the measuring rods map out a non-Euclidean geometry.

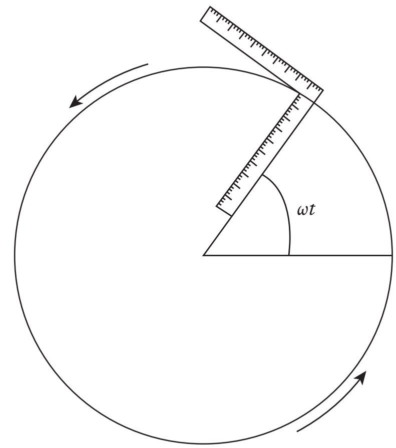

Figure 5.3 *Rotating cylinder with length contraction in the tangential direction but not in the radial direction, resulting in a non-Euclidean relation between circumference and radius.*

Local flat metric and local inertial frame

In a curved space, a small region can always be described approximately as a flat space per the flatness theorem, cf. (5.38). Now, if we identify the curvature of our spacetime as the gravitational field, the corresponding flatness theorem must be satisfied. Indeed, the EP informs us that gravity can always be transformed away locally. In the absence of gravity, spacetime is flat. Thus Einstein put forward this elegant theory that identifies gravity as the structure of spacetime, thereby incorporating the EP in a fundamental way.

A 2D illustration of geometry as gravity The possibility of using a curved space to represent a gravitational field can be illustrated with the following example involving a 2D curved surface. Two masses on a spherical surface start out at the equator and move along two longitudinal great circles (which are geodesics, as shown in Exercise 5.4). As they move along, the distance between them decreases (Fig. 5.4). We can attribute this convergence to some attractive force between them or simply to the curvature of space. We will discuss such tidal effects in more detail in Section 6.2.2.

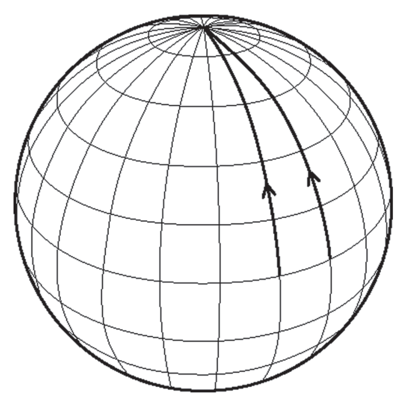

Figure 5.4 *The convergence of two particle trajectories can be explained by either an attractive force or the underlying geometry of a spherical surface.*

Example 5.1 Curved spacetime and gravitational redshift

In Chapter 4, we showed that the strong EP implies a gravitational redshift (in a static gravitational field) of light frequency ω, cf. (4.20),

$$\frac{\Delta\omega}{\omega} = -\frac{\Delta\Phi}{c^2}. \tag{5.42}$$

continued

Example 5.1 *continued*

From this result, we heuristically deduced that in the presence of nonzero gravitational potential, the metric must deviate from its flat-space value. Namely, from the gravitational redshift, we deduced a curved spacetime. Now we shall show the converse: that a curved spacetime implies redshift. In this chapter, we have seen that Einstein's theory based on a curved spacetime yields the result (5.41) in the Newtonian limit. This can be stated as a relation between the proper time τ and the coordinate time t:

$$d\tau = \sqrt{-g_{00}}\, dt, \qquad \text{with} \qquad g_{00} = -\left(1 + 2\frac{\Phi}{c^2}\right). \tag{5.43}$$

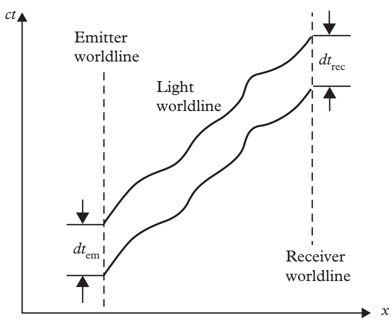

Figure 5.5 *Worldlines for two light wavefronts propagating from emitter to receiver in a static curved spacetime.*

Here we wish to show how the gravitational frequency shift result (5.42) emerges in this curved-spacetime description. In Fig. 5.5, the two curvy lines are the lightlike worldlines of two wavefronts emitted at an interval dt_{em} apart. They curve because the spacetime is warped by gravity. (In flat spacetime, they would be two straight 45° lines.) Because the gravitational field is static (hence spacetime curvature is time-independent), this dt_{em} time interval between the two wavefronts is maintained throughout the trip from emission to reception:

$$dt_{\mathrm{em}} = dt_{\mathrm{rec}}. \tag{5.44}$$

On the other hand, because the frequency is inversely proportional to the proper time interval $\omega = 1/d\tau$, we can use (5.43) and (5.44) to derive the redshift:

$$\begin{aligned}\frac{\omega_{\mathrm{rec}}}{\omega_{\mathrm{em}}} &= \frac{d\tau_{\mathrm{em}}}{d\tau_{\mathrm{rec}}} = \frac{\sqrt{-(g_{00})_{\mathrm{em}}}\, dt_{\mathrm{em}}}{\sqrt{-(g_{00})_{\mathrm{rec}}}\, dt_{\mathrm{rec}}} = \left(\frac{1 + 2(\Phi_{\mathrm{em}}/c^2)}{1 + 2(\Phi_{\mathrm{rec}}/c^2)}\right)^{1/2} \\ &= 1 + \frac{\Phi_{\mathrm{em}} - \Phi_{\mathrm{rec}}}{c^2} + O\left(\Phi^2/c^4\right),\end{aligned} \tag{5.45}$$

which is the claimed result of (5.42):

$$\frac{\omega_{\mathrm{rec}} - \omega_{\mathrm{em}}}{\omega_{\mathrm{em}}} = -\frac{\Phi_{\mathrm{rec}} - \Phi_{\mathrm{em}}}{c^2}. \tag{5.46}$$

5.2.2 GR as a field theory of gravitation

Recall that a field-theoretical description of the interaction between a source and a test particle involves two steps:

$$\boxed{\text{Source particle}} \xrightarrow[\text{Field equation}]{} \boxed{\text{Field}} \xrightarrow[\text{Equation of motion}]{} \boxed{\text{Test particle}}$$

Instead of acting directly on the test particle through some instantaneous action-at-a-distance force, the source particle creates a field everywhere, which then acts on the test particle locally. The first step is governed by the field equation, which, given the source distribution, dictates the field everywhere. In the case of electromagnetism, this is Maxwell's equations. In the second step, the equation of motion determines a test particle's motion from the field at its location. The electromagnetic equation of motion follows directly from the Lorentz force law.

Based on his study of EP phenomenology, Einstein made the conceptual leap (a logical deduction, but a startling leap nevertheless) that curved spacetime is the gravitational field. A source mass gives rise to a curved spacetime, which in turn influences the motion of a test mass:

$$\boxed{\text{Source}} \xrightarrow[\text{Einstein field equation}]{} \boxed{\text{Curved spacetime}} \xrightarrow[\text{Geodesic equation}]{} \boxed{\text{Test particle}}$$

While spacetime in SR, as in all pre-GR physics, is fixed, it is dynamic in the general theory of relativity and is determined by the matter/energy distribution. GR fulfills Einstein's conviction that "space is not a thing." Spacetime is merely an expression of the ever-changing relations among physical processes. Thus the metric,[11] which describes the geometry, is ever-changing. Furthermore, the laws of physics should not depend on reference frames. Physics equations should be tensor equations under general coordinate transformations. This principle of general covariance is a key feature of GR.

[11] It is important to note that the gravitational field is not a scalar, nor is it a four-component vector, but rather a symmetric tensor $g_{\mu\nu} = g_{\nu\mu}$ with ten independent components; in contrast, the antisymmetric electromagnetic field tensor $F_{\mu\nu} = -F_{\nu\mu}$ has six components.

Next we shall study the GR equation of motion, the geodesic equation, which describes the motion of a test particle in a curved spacetime. The more difficult topic of the GR field equation, the Einstein equation, is deferred to Chapter 6, after we have briefly discussed the Riemann curvature tensor.

5.3 Geodesic equation as the GR equation of motion

In GR, the mass/energy source determines the metric function through the field equation. Namely, the metric $g_{\mu\nu}(x)$ is the solution of the GR field equation. From $g_{\mu\nu}(x)$, one can write down the geodesic equation, which fixes the trajectory of the test particle. In this approach, gravity is regarded as the structure of spacetime rather than as a force (which would bring about acceleration through Newton's second law). That is, a test body will move freely in such a curved spacetime; it is natural to expect[12] it to follow the shortest and straightest possible trajectory, the geodesic curve. Thus the particle's coordinate acceleration comes from the geodesic equation (5.29) rather than Newton's second law.

[12] The correctness of this heuristic choice will be justified by a formal derivation of the geodesic equation in Section 11.3.1.

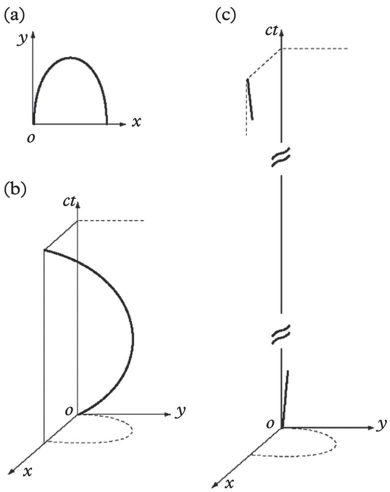

Figure 5.6 *(a) Particle trajectory in the (x, y) plane. (b) Particle worldline with projection onto the (x, y) plane as shown in (a), plotted with a greatly compressed time axis. (c) Spacetime diagram with the time axis restored to its proper scale.*

[13] One can picture the spacetime being curved by the (gravitational) change of the time intervals when moving away from the origin.

Box 5.3 The geodesic is the worldline of a test particle

It may appear somewhat surprising to hear that a test particle will follow a "straight line" in the presence of a gravitational field. After all, our experience is just the opposite: when we throw an object in the earth's gravitational field, it follows a parabolic trajectory. Was Einstein saying that the parabolic trajectory is actually straight? All such paradoxes result from confusing 4D spacetime with ordinary 3D space. The GR equation of motion tells us that a test particle will follow a geodesic line in spacetime (whose invariant interval has a negative part from the time coordinate) rather than a geodesic line in the 3D space (which minimizes ordinary length). A geodesic in spacetime (a worldline) generally does not imply a straight trajectory in its spatial subspace. A simple illustration using a spacetime diagram should make this clear.

Consider an object thrown to a height of 10 m over a distance of 10 m. Its spatial trajectory is displayed in Fig. 5.6(a). When we represent the corresponding worldline in the spacetime diagram, we must also plot the time axis *ct*; see Fig. 5.6(b). The object takes 1.4 s to reach the highest point and another 1.4 s to come down. But a 2.8 s time interval will be represented by almost one million kilometers of *ct* in the spacetime diagram (more than the round-trip distance to the moon), leading to a very nearly straight worldline as depicted in Fig. 5.6(c).

The main point here is not so much the straightness of the worldline, which reflects the practically flat spacetime in the very weak terrestrial gravitational field (recall that $\Phi_{\oplus}/c^2 \simeq 10^{-10}$). Rather, the point is that one must not confuse the trajectory in regular 3D space with the geodesic curve in spacetime.[13]

The interval extremized by the geodesic in spacetime is not simply the spatial length (cf. Box 5.3). In fact, the invariant interval of a particle's worldline is its proper time. We shall demonstrate that this geodesic equation is the relativistic generalization of the Newtonian equation of motion (4.9). To do so, we must define more precisely the sense in which Einstein's theory is an extension of Newtonian gravity; it is much more than an extension to higher-speed motion.

5.3.1 The Newtonian limit

To support our claim that the geodesic equation is the GR equation of motion, we shall now show that the geodesic equation (5.29) does reduce to the Newtonian equation of motion (4.9) in the Newtonian limit of

a test particle moving with nonrelativistic velocity $v \ll c$,
in a weak and static gravitational field.

We now take such a limit of the GR equation of motion (5.29):

- Nonrelativistic speed, $dx^i/dt \ll c$: This inequality, $dx^i \ll c\,dt$, implies that

$$\frac{dx^i}{d\tau} \ll c\frac{dt}{d\tau} = \frac{dx^0}{d\tau}. \tag{5.47}$$

  Keeping only the dominant term $(dx^0/d\tau)(dx^0/d\tau)$ in the double sum over indices λ and ρ in the geodesic equation (5.29), we have

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{00}\frac{dx^0}{d\tau}\frac{dx^0}{d\tau} = 0. \tag{5.48}$$

- Static field, $\partial g_{\mu\nu}/\partial x^0 = 0$: Because all time derivatives vanish, the Christoffel symbols of (5.30) take a simpler form:

$$\Gamma^\mu_{00} = -\frac{1}{2}g^{\mu\nu}\frac{\partial g_{00}}{\partial x^\nu}. \tag{5.49}$$

- Weak field, $h_{\mu\nu} \ll 1$: We assume that the metric is not too different from the flat-spacetime metric $\eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{5.50}$$

  where $h_{\mu\nu}(x)$ is a small correction field. $\eta_{\mu\nu}$ is constant, so $\partial g_{\mu\nu}/\partial x^\lambda = \partial h_{\mu\nu}/\partial x^\lambda$. The Christoffel symbols, being derivatives of the metric, are of order $h_{\mu\nu}$. To leading order, (5.49) is

$$\Gamma^\mu_{00} = -\frac{1}{2}\eta^{\mu\nu}\frac{\partial h_{00}}{\partial x^\nu},$$

  which, because $\eta_{\nu\mu}$ is diagonal, has (for a static h_{00}) the following components:

$$\Gamma^0_{00} = \frac{1}{2}\frac{\partial h_{00}}{\partial x^0} = 0 \quad \text{and} \quad \Gamma^i_{00} = -\frac{1}{2}\frac{\partial h_{00}}{\partial x^i}. \tag{5.51}$$

  We can now evaluate (5.48) by using (5.51): the $\mu = 0$ equation leads to

$$\frac{dx^0}{d\tau} = \text{constant}, \tag{5.52}$$

  and the three $\mu = i$ equations are

$$\frac{d^2x^i}{d\tau^2} + \Gamma^i_{00}\frac{dx^0}{d\tau}\frac{dx^0}{d\tau} = \left(\frac{d^2x^i}{c^2\,dt^2} + \Gamma^i_{00}\right)\left(\frac{dx^0}{d\tau}\right)^2 = 0, \tag{5.53}$$

where we have used the chain rule $dx^i/d\tau = (dx^i/dx^0)(dx^0/d\tau)$ and the condition (5.52) to conclude that $d^2x^i/d\tau^2 = (d^2x^i/dx^{0\,2})(dx^0/d\tau)^2$. Equation (5.53), together with (5.51), implies

$$\frac{d^2x^i}{c^2\,dt^2} - \frac{1}{2}\frac{\partial h_{00}}{\partial x^i} = 0, \tag{5.54}$$

which can be compared with the Newtonian equation of motion (4.9). Thus $h_{00} = -2\Phi/c^2$, and, using the definition (5.50), we recover (5.41), first obtained heuristically in the previous discussion.

We can indeed regard the metric tensor as the relativistic generalization of the gravitational potential. The expression (5.50) also provides us with a criterion to characterize a field as weak:

$$\left[|h_{00}| \ll |\eta_{00}|\right] \quad \Rightarrow \quad \left[\frac{|\Phi|}{c^2} \ll 1\right]. \tag{5.55}$$

Consider the gravitational potential at the earth's surface. It is equal in magnitude to the gravitational acceleration times the earth's radius, $|\Phi_\oplus| = g \times R_\oplus = O(10^7\,\mathrm{m^2/s^2})$, or $|\Phi_\oplus|/c^2 = O(10^{-10})$. Thus any gravitational field less than ten billion g's (acting over distances comparable to the earth's radius) may be considered weak.

Review questions

1. What is an intrinsic geometric description (vs. an extrinsic description)? Describe the intrinsic geometric operations that fix the metric elements.
2. What is the relation of the geodesic equation to the length-extremization condition?
3. What is the fundamental difference between coordinate transformations in a curved space and those in flat space (e.g., Lorentz transformations in flat Minkowski space)?
4. What is a local Euclidean frame of reference? What is the flatness theorem?
5. What does one mean by a "geometric theory of physics"? Use distance measurements on the surface of a globe to illustrate your answer.
6. How can the phenomenon of gravitational time dilation be described in geometric terms? Use this to argue that the spacetime metric can be regarded as the relativistic gravitational potential.
7. Use the simple example of a rotating cylinder to illustrate how EP physics can imply a non-Euclidean geometric relation.
8. What significant conclusion did Einstein draw from the analogy between the facts that a curved space is locally flat and that gravity can be transformed away locally?
9. Give the heuristic argument that the GR equation of motion is the geodesic equation.
10. What is the Newtonian limit? In this limit, what relation can one infer between the Newtonian gravitational potential and a certain metric tensor component of the spacetime. Use this relation to derive the gravitational redshift.