CONTENTS

T2 Track Two; see page xvii

CHAPTER TWENTY-FOUR

24

From Special to General Relativity

The Theory of Relativity confers an absolute meaning on a magnitude which in classical theory has only a relative significance: the velocity of light. The velocity of light is to the Theory of Relativity as the elementary quantum of action is to the Quantum Theory: it is its absolute core.

MAX PLANCK (1949)

24.1 Overview 24.1

We begin our discussion of general relativity in this chapter with a review, and elaboration of relevant material already covered in earlier chapters. In Sec. 24.2, we give a brief encapsulation of special relativity drawn largely from Chap. 2, emphasizing those aspects that underpin the transition to general relativity. Then in Sec. 24.3 we collect, review, and extend the fundamental ideas of differential geometry that have been scattered throughout the book and that we shall need as foundations for the mathematics of *spacetime curvature* (Chap. 25). Most importantly, we generalize differential geometry to encompass coordinate systems whose coordinate lines are not orthogonal and bases that are not orthonormal.

Einstein's field equation (to be studied in Chap. 25) is a relationship between the curvature of spacetime and the matter that generates it, akin to the Maxwell equations' relationship between the electromagnetic field and the electric currents and charges that generate it. The matter in Einstein's equation is described by the stress-energy tensor that we introduced in Sec. 2.13. We revisit the stress-energy tensor in Sec. 24.4 and develop a deeper understanding of its properties.

In general relativity one often wishes to describe the outcome of measurements made by observers who refuse to fall freely—for example, an observer who hovers in a spaceship just above the horizon of a black hole, or a gravitational-wave experimenter in an Earthbound laboratory. As a foundation for treating such observers, in Sec. 24.5 we examine measurements made by accelerated observers in the flat spacetime of special relativity.

24.2 Special Relativity Once Again 24.2

Our viewpoint on general relativity is unapologetically geometrical. (Other viewpoints, e.g., those of particle theorists such as Feynman and Weinberg, are quite different.) Therefore, a prerequisite for our treatment of general relativity is understanding special relativity in geometric language. In Chap. 2, we discussed the foundations of

BOX 24.1. READERS' GUIDE

- This chapter relies significantly on:
  - Chap. 2 on special relativity, which now should be regarded as Track One.
  - The discussion of connection coefficients in Sec. 11.8.
- This chapter is a foundation for the presentation of general relativity theory and cosmology in Chaps. 25–28.

special relativity with this in mind. In this section we briefly review the most important points.

We suggest that any reader who has not studied Chap. 2 read Sec. 24.2 first, to get an overview and flavor of what will be important for our development of general relativity, and then (or in parallel with reading Sec. 24.2) read those relevant sections of Chap. 2 that the reader does not already understand.

24.2.1

24.2.1 Geometric, Frame-Independent Formulation

review of the geometric, frame-independent formulation of special relativity

In Secs. 1.1.1 and 2.2.2, we learned that *every law of physics must be expressible as a geometric, frame-independent relationship among geometric, frame-independent objects.* This is equally true in Newtonian physics, in special relativity, and in general relativity. The key difference between the three is the geometric arena: in Newtonian physics, the arena is 3-dimensional Euclidean space; in special relativity, it is 4-dimensional Minkowski spacetime; in general relativity (Chap. 25), it is 4-dimensional curved spacetime (see Fig. 1 in the Introduction to Part I and the associated discussion).

Principle of Relativity—laws as geometric relations between geometric objects

In special relativity, the demand that the laws be geometric relationships among geometric objects that live in Minkowski spacetime is the *Principle of Relativity*; see Sec. 2.2.2. Examples of the geometric objects are:

examples of geometric objects: points, curves, proper time ticked by an ideal clock, vectors, tensors, scalar product

1. A point $\mathcal{P}$ in spacetime (which represents an *event*); Sec. 2.2.1.
2. A parameterized curve in spacetime, such as the world line $\mathcal{P}(\tau)$ of a particle, for which the parameter τ is the particle's *proper time* (i.e., the time measured by an ideal clock[1] that the particle carries; Fig. 24.1); Sec. 2.4.1.

1. Recall that an ideal clock is one that ticks uniformly when compared, e.g., to the period of the light emitted by some standard type of atom or molecule, and that has been made impervious to accelerations. Thus two ideal clocks momentarily at rest with respect to each other tick at the same rate independent of their relative acceleration; see Secs. 2.2.1 and 2.4.1. For greater detail, see Misner, Thorne, and Wheeler (1973, pp. 23–29, 395–399).

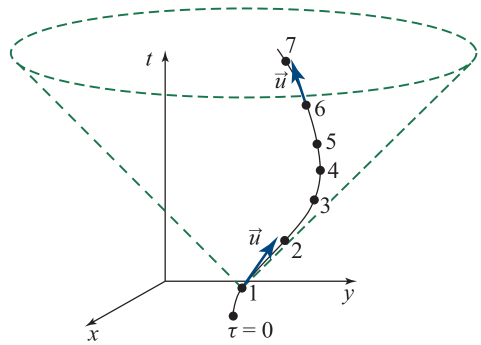

FIGURE 24.1 The world line $\mathcal{P}(\tau)$ of a particle in Minkowski spacetime and the tangent vector $\vec{u} = d\mathcal{P}/d\tau$ to this world line; $\vec{u}$ is the particle's 4-velocity. The bending of the world line is produced by some force that acts on the particle, such as the Lorentz force embodied in Eq. (24.3). Also shown is the light cone emitted from the event $\mathcal{P}(\tau = 1)$. Although the axes of an (arbitrary) inertial reference frame are shown, no reference frame is needed for the definition of the world line, its tangent vector $\vec{u}$, or the light cone. Nor is one needed for the formulation of the Lorentz force law.

3. Vectors, such as the particle's 4-velocity $\vec{u} = d\mathcal{P}/d\tau$ [the tangent vector to the curve $\mathcal{P}(\tau)$] and the particle's 4-momentum $\vec{p} = m\vec{u}$ (with m the particle's rest mass); Secs. 2.2.1 and 2.4.1.
4. Tensors, such as the electromagnetic field tensor $\boldsymbol{F}(\_\_, \_\_)$; Secs. 1.3 and 2.3.

Recall that a tensor is a linear real-valued function of vectors; when one puts vectors $\vec{A}$ and $\vec{B}$ into the two slots of $\boldsymbol{F}$, one obtains a real number (a scalar) $\boldsymbol{F}(\vec{A}, \vec{B})$ that is linear in $\vec{A}$ and in $\vec{B}$ so, for example: $\boldsymbol{F}(\vec{A}, b\vec{B} + c\vec{C}) = b\boldsymbol{F}(\vec{A}, \vec{B}) + c\boldsymbol{F}(\vec{A}, \vec{C})$. When one puts a vector $\vec{B}$ into just one of the slots of $\boldsymbol{F}$ and leaves the other empty, one obtains a tensor with one empty slot, $\boldsymbol{F}(\_\_, \vec{B})$, that is, a vector. The result of putting a vector into the slot of a vector is the scalar product: $\vec{D}(\vec{B}) = \vec{D} \cdot \vec{B} = \boldsymbol{g}(\vec{D}, \vec{B})$, where $\boldsymbol{g}(\_\_, \_\_)$ is the metric.

spacetime metric

In Secs. 2.3 and 2.4.1, we tied our definitions of the inner product and the spacetime metric to the ticking of ideal clocks: If $\Delta\vec{x}$ is the vector separation of two neighboring events $\mathcal{P}(\tau)$ and $\mathcal{P}(\tau + \Delta\tau)$ along a particle's world line, then

$$\boxed{\boldsymbol{g}(\Delta\vec{x}, \Delta\vec{x}) \equiv \Delta\vec{x} \cdot \Delta\vec{x} \equiv -(\Delta\tau)^2.} \tag{24.1}$$

This relation for any particle with any timelike world line, together with the linearity of $\boldsymbol{g}(\_\_, \_\_)$ in its two slots, is enough to determine $\boldsymbol{g}$ completely and to guarantee that it is symmetric: $\boldsymbol{g}(\vec{A}, \vec{B}) = \boldsymbol{g}(\vec{B}, \vec{A})$ for all $\vec{A}$ and $\vec{B}$. Since the particle's 4-velocity $\vec{u}$ is

$$\vec{u} = \frac{d\mathcal{P}}{d\tau} = \lim_{\Delta\tau \to 0} \frac{\mathcal{P}(\tau + \Delta\tau) - \mathcal{P}(\tau)}{\Delta\tau} \equiv \lim_{\Delta\tau \to 0} \frac{\Delta\vec{x}}{\Delta\tau}, \tag{24.2}$$

Eq. (24.1) implies that $\vec{u} \cdot \vec{u} = \boldsymbol{g}(\vec{u}, \vec{u}) = -1$ (Sec. 2.4.1).

light cone; timelike, null, and spacelike vectors

The 4-velocity $\vec{u}$ is an example of a *timelike* vector (Sec. 2.2.3); it has a negative inner product with itself (negative "squared length"). This shows up pictorially in the fact that $\vec{u}$ lies inside the *light cone* (the cone swept out by the trajectories of photons emitted from the tail of $\vec{u}$; see Fig. 24.1). Vectors $\vec{k}$ on the light cone (the tangents to the world lines of the photons) are *null* and so have vanishing squared lengths: $\vec{k} \cdot \vec{k} = \boldsymbol{g}(\vec{k}, \vec{k}) = 0$; vectors $\vec{A}$ that lie outside the light cone are *spacelike* and have positive squared lengths: $\vec{A} \cdot \vec{A} > 0$ (Sec. 2.2.3).

An example of a physical law in 4-dimensional geometric language is the Lorentz force law (Sec. 2.4.2):

Lorentz force law

$$\frac{d\vec{p}}{d\tau} = q\boldsymbol{F}(\_\_, \vec{u}). \tag{24.3}$$

Here q is the particle's charge (a scalar), and both sides of this equation are vectors, or equivalently, first-rank tensors (i.e., tensors with just one slot). As we learned in Secs. 1.5.1 and 2.5.3, it is convenient to give names to slots. When we do so, we can rewrite the Lorentz force law as

$$\frac{dp^\alpha}{d\tau} = qF^{\alpha\beta}u_\beta. \tag{24.4}$$

slot-naming index notation

Here α is the name of the slot of the vector $d\vec{p}/d\tau$, α and β are the names of the slots of $\boldsymbol{F}$, β is the name of the slot of **u**. The double use of β with one up and one down on the right-hand side of the equation represents the insertion of $\vec{u}$ into the β slot of $\boldsymbol{F}$, whereby the two β slots disappear, and we wind up with a vector whose slot is named α. As we learned in Sec. 1.5, this slot-naming index notation is isomorphic to the notation for components of vectors, tensors, and physical laws in some reference frame. However, no reference frames are needed or involved when one formulates the laws of physics in geometric, frame-independent language as above.

Those readers who do not feel completely comfortable with these concepts, statements, and notation should reread the relevant portions of Chaps. 1 and 2.

EXERCISES

Exercise 24.1 *Practice: Frame-Independent Tensors*
Let $\boldsymbol{A}$, $\boldsymbol{B}$ be second-rank tensors.

(a) Show that $\boldsymbol{A} + \boldsymbol{B}$ is also a second-rank tensor.

(b) Show that $\boldsymbol{A} \otimes \boldsymbol{B}$ is a fourth-rank tensor.

(c) Show that the contraction of $\boldsymbol{A} \otimes \boldsymbol{B}$ on its first and fourth slots is a second-rank tensor. (If necessary, consult Secs. 1.5 and 2.5 for discussions of contraction.)

(d) Write the following quantities in slot-naming index notation: the tensor $\boldsymbol{A} \otimes \boldsymbol{B}$, and the simultaneous contraction of this tensor on its first and fourth slots and on its second and third slots.

24.2.2

24.2.2 Inertial Frames and Components of Vectors, Tensors, and Physical Laws

inertial reference frame

In special relativity, a key role is played by *inertial reference frames,* Sec. 2.2.1. An inertial frame is an (imaginary) latticework of rods and clocks that moves through spacetime freely (inertially, without any force acting on it). The rods are orthogonal to one another and attached to inertial-guidance gyroscopes, so they do not rotate. These rods are used to identify the spatial, Cartesian coordinates $(x^1, x^2, x^3) = (x, y, z)$ of an event $\mathcal{P}$ [which we also denote by lowercased Latin indices $x^j(\mathcal{P})$, with j running over 1, 2, 3]. The latticework's clocks are ideal and are synchronized with one another by the Einstein light-pulse process. They are used to identify the temporal coordinate $x^0 = t$ of an event $\mathcal{P}$: $x^0(\mathcal{P})$ is the time measured by that latticework clock whose world line passes through $\mathcal{P}$, at the moment of passage. The spacetime coordinates of $\mathcal{P}$ are denoted by lowercased Greek indices x^α, with α running over 0, 1, 2, 3. An inertial frame's spacetime coordinates $x^\alpha(\mathcal{P})$ are called *Lorentz coordinates* or *inertial coordinates.*

Lorentz (inertial) coordinates

In the real universe, spacetime curvature is small in regions well removed from concentrations of matter (e.g., in intergalactic space), so special relativity is highly accurate there. In such a region, frames of reference (rod-clock latticeworks) that are nonaccelerating and nonrotating with respect to cosmologically distant galaxies (and hence with respect to a local frame in which the cosmic microwave radiation looks isotropic) constitute good approximations to inertial reference frames.

orthonormal basis vectors of an inertial frame

Associated with an inertial frame's Lorentz coordinates are basis vectors $\vec{e}_\alpha$ that point along the frame's coordinate axes (and thus are orthogonal to one another) and have unit length (making them orthonormal); see Sec. 2.5. This orthonormality is embodied in the inner products

$$\boxed{\vec{e}_\alpha \cdot \vec{e}_\beta = \eta_{\alpha\beta},} \tag{24.5}$$

where by definition:

$$\boxed{\eta_{00} = -1, \quad \eta_{11} = \eta_{22} = \eta_{33} = +1, \quad \eta_{\alpha\beta} = 0 \quad \text{if } \alpha \neq \beta.} \tag{24.6}$$

geometrized units

Here and throughout Part VII (as in Chap. 2), we set the speed of light to unity (i.e., we use the geometrized units introduced in Sec. 1.10), so spatial lengths (e.g., along the x-axis) and time intervals (e.g., along the t-axis) are measured in the same units, seconds or meters, with $1\ \text{s} = 2.99792458 \times 10^8$ m.

In Sec. 2.5 (see also Sec. 1.5), we used the basis vectors of an inertial frame to build a component representation of tensor analysis. The fact that the inner products of timelike vectors with each other are negative (e.g., $\vec{e}_0 \cdot \vec{e}_0 = -1$), while those of spacelike vectors are positive (e.g., $\vec{e}_1 \cdot \vec{e}_1 = +1$), forced us to introduce two types of components: *covariant* (indices down) and *contravariant* (indices up). The covariant components of a tensor are computable by inserting the basis vectors into the tensor's slots: $u_\alpha = \vec{u}(\vec{e}_\alpha) \equiv \vec{u} \cdot \vec{e}_\alpha$; $F_{\alpha\beta} = \boldsymbol{F}(\vec{e}_\alpha, \vec{e}_\beta)$. For example, in our Lorentz basis the covariant components of the metric are $g_{\alpha\beta} = \boldsymbol{g}(\vec{e}_\alpha, \vec{e}_\beta) = \vec{e}_\alpha \cdot \vec{e}_\beta = \eta_{\alpha\beta}$. The contravariant components of a tensor were related to the covariant components via "index lowering" with the aid of the metric, $F_{\alpha\beta} = g_{\alpha\mu} g_{\beta\nu} F^{\mu\nu}$, which simply said that one reverses the sign when lowering a time index and makes no change of sign when lowering a space index. This lowering rule implied that the contravariant components of the metric in a Lorentz basis are the same numerically as the covariant

covariant and contravariant components of vectors and tensors

components, $g^{\alpha\beta} = \eta_{\alpha\beta}$, and that they can be used to raise indices (i.e., to perform the trivial sign flip for temporal indices): $F^{\mu\nu} = g^{\mu\alpha} g^{\nu\beta} F_{\alpha\beta}$. As we saw in Sec. 2.5, tensors can be expressed in terms of their contravariant components as $\vec{p} = p^\alpha \vec{e}_\alpha$, and $\boldsymbol{F} = F^{\alpha\beta} \vec{e}_\alpha \otimes \vec{e}_\beta$, where $\otimes$ represents the tensor product [Eqs. (1.5)].

component equations are same as slot-naming-index-notation equations

We also learned in Chap. 2 that any frame-independent geometric relation among tensors can be rewritten as a relation among those tensors' components in any chosen Lorentz frame. When one does so, the resulting component equation takes precisely the same form as the slot-naming-index-notation version of the geometric relation (Sec. 1.5.1). For example, the component version of the Lorentz force law says $dp^\alpha/d\tau = qF^{\alpha\beta} u_\beta$, which is identical to Eq. (24.4). The only difference is the interpretation of the symbols. In the component equation $F^{\alpha\beta}$ are the components of $\boldsymbol{F}$ and the repeated β in $F^{\alpha\beta} u_\beta$ is to be summed from 0 to 3. In the geometric relation $F^{\alpha\beta}$ means $\boldsymbol{F}(\_\_, \_\_)$, with the first slot named α and the second β, and the repeated β in $F^{\alpha\beta} u_\beta$ implies the insertion of $\vec{u}$ into the second slot of $\boldsymbol{F}$ to produce a single-slotted tensor (i.e., a vector) whose slot is named α.

components of 4-velocity in an inertial frame

3 + 1 split

As we saw in Sec. 2.6, a particle's 4-velocity $\vec{u}$ (defined originally without the aid of any reference frame; Fig. 24.1) has components, in any inertial frame, given by $u^0 = \gamma$, $u^j = \gamma v^j$, where $v^j = dx^j/dt$ is the particle's ordinary velocity and $\gamma \equiv 1\Big/\sqrt{1 - \delta_{ij} v^i v^j}$. Similarly, the particle's energy $E \equiv p^0$ is $m\gamma$, and its spatial momentum is $p^j = m\gamma v^j$ (i.e., in 3-dimensional geometric notation: $\mathbf{p} = m\gamma \mathbf{v}$). This is an example of the manner in which a choice of Lorentz frame produces a "3+1" split of the physics: a split of 4-dimensional spacetime into 3-dimensional space (with Cartesian coordinates x^j) plus 1-dimensional time $t = x^0$; a split of the particle's 4-momentum $\vec{p}$ into its 3-dimensional spatial momentum $\mathbf{p}$ and its 1-dimensional energy $\mathcal{E} = p^0$; and similarly a split of the electromagnetic field tensor $\boldsymbol{F}$ into the 3-dimensional electric field $\mathbf{E}$ and 3-dimensional magnetic field $\mathbf{B}$ (cf. Secs. 2.6 and 2.11).

Principle of Relativity restated: laws take same form in every inertial frame

The Principle of Relativity (all laws expressible as geometric relations between geometric objects in Minkowski spacetime), when translated into 3+1 language, says that, when the laws of physics are expressed in terms of components in a specific Lorentz frame, the form of those laws must be independent of one's choice of frame. When translated into operational terms, it says that, if two observers in two different Lorentz frames are given identical written instructions for a self-contained physics experiment, then their two experiments must yield the same results to within their experimental accuracies (Sec. 2.2.2).

Lorentz transformations

Principle of Relativity restated: laws are Lorentz invariant

The components of tensors in one Lorentz frame are related to those in another by a Lorentz transformation (Sec. 2.7), so the Principle of Relativity can be restated as saying that, when expressed in terms of Lorentz-frame components, *the laws of physics must be Lorentz-invariant* (unchanged by Lorentz transformations). This is the version of the Principle of Relativity that one meets in most elementary treatments of special relativity. However, as the above discussion shows, it is a mere shadow of the true Principle of Relativity—the shadow cast into Lorentz frames when one performs a 3+1 split. The ultimate, fundamental version of the Principle of Relativity is the one that needs no frames at all for its expression: *all the laws of physics are expressible as geometric relations among geometric objects that reside in Minkowski spacetime*.

ultimate version of Principle of Relativity

24.2.3 Light Speed, the Interval, and Spacetime Diagrams

24.2.3

One set of physical laws that must be the same in all inertial frames is Maxwell's equations. Let us discuss the implications of Maxwell's equations and the Principle of Relativity for the speed of light c. (For a more detailed discussion, see Sec. 2.2.2.) According to Maxwell, c can be determined by performing nonradiative laboratory experiments; it is not necessary to measure the time it takes light to travel along some path; see Box 2.2. The Principle of Relativity requires that such experiments must give the same result for c, independent of the reference frame in which the measurement apparatus resides, so the speed of light must be independent of reference frame. It is this frame independence that enables us to introduce geometrized units with $c = 1$.

light speed is the same in all inertial frames

Another example of frame independence (Lorentz invariance) is provided by the *interval between two events* (Sec. 2.2.3). The components $g_{\alpha\beta} = \eta_{\alpha\beta}$ of the metric imply that, if $\Delta\vec{x}$ is the vector separating the two events and Δx^α are its components in some Lorentz coordinate system, then the squared length of $\Delta\vec{x}$ [also called the *interval* and denoted $(\Delta s)^2$] is given by

$$\boxed{\begin{aligned}(\Delta s)^2 \equiv \Delta\vec{x}\cdot\Delta\vec{x} = \boldsymbol{g}(\Delta\vec{x}, \Delta\vec{x}) = g_{\alpha\beta}\Delta x^\alpha \Delta x^\beta \\ = -(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2.\end{aligned}} \tag{24.7}$$

interval between two events

Since $\Delta\vec{x}$ is a geometric, frame-independent object, so must be the interval. This implies that the equation $(\Delta s)^2 = -(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$ by which one computes the interval between the two chosen events in one Lorentz frame must give the same numerical result when used in any other frame (i.e., this expression must be Lorentz invariant). This *invariance of the interval* is the starting point for most introductions to special relativity—and, indeed, we used it as a starting point in Sec. 2.2.

invariance of the interval

Spacetime diagrams play a major role in our development of general relativity. Accordingly, it is important that the reader feel very comfortable with them. We recommend reviewing Fig. 2.7 and Ex. 2.14.

spacetime diagrams

EXERCISES

Exercise 24.2 *Example: Invariance of a Null Interval*
You have measured the intervals between a number of adjacent events in spacetime and thereby have deduced the metric $\boldsymbol{g}$. Your friend claims that the metric is some other frame-independent tensor $\tilde{\boldsymbol{g}}$ that differs from $\boldsymbol{g}$. Suppose that your correct metric $\boldsymbol{g}$ and his wrong one $\tilde{\boldsymbol{g}}$ agree on the forms of the light cones in spacetime (i.e., they agree as to which intervals are null, which are spacelike, and which are timelike), but they give different answers for the value of the interval in the spacelike and timelike cases: $\boldsymbol{g}(\Delta\vec{x}, \Delta\vec{x}) \neq \tilde{\boldsymbol{g}}(\Delta\vec{x}, \Delta\vec{x})$. Prove that $\tilde{\boldsymbol{g}}$ and $\boldsymbol{g}$ differ solely by

a scalar multiplicative factor, $\tilde{\boldsymbol{g}} = a\boldsymbol{g}$ for some scalar a. We say that $\tilde{\boldsymbol{g}}$ and $\boldsymbol{g}$ are *conformal to each other.* [Hint: Pick some Lorentz frame and perform computations there, then lift yourself back up to a frame-independent viewpoint.]

Exercise 24.3 *Problem: Causality*
If two events occur at the same spatial point but not simultaneously in one inertial frame, prove that the temporal order of these events is the same in all inertial frames. Prove also that in all other frames the temporal interval Δt between the two events is larger than in the first frame, and that there are no limits on the events' spatial or temporal separation in the other frames. Give *two* proofs of these results, one algebraic and the other via spacetime diagrams.

24.3

24.3 Differential Geometry in General Bases and in Curved Manifolds

The differential geometry (tensor-analysis) formalism reviewed in the last section is inadequate for general relativity in several ways.

First, in general relativity we need to use bases $\vec{e}_\alpha$ that are not orthonormal (i.e., for which $\vec{e}_\alpha \cdot \vec{e}_\beta \neq \eta_{\alpha\beta}$). For example, near a spinning black hole there is much power in using a time basis vector $\vec{e}_t$ that is tied in a simple way to the metric's time-translation symmetry and a spatial basis vector $\vec{e}_\phi$ that is tied to its rotational symmetry. This time basis vector has an inner product with itself $\vec{e}_t \cdot \vec{e}_t = g_{tt}$ that is influenced by the slowing of time near the hole (so $g_{tt} \neq -1$); and $\vec{e}_\phi$ is not orthogonal to $\vec{e}_t$ ($\vec{e}_t \cdot \vec{e}_\phi = g_{t\phi} \neq 0$), as a result of the dragging of inertial frames by the hole's spin. In this section, we generalize our formalism to treat such nonorthonormal bases.

Second, in the curved spacetime of general relativity (and in any other curved space, e.g., the 2-dimensional surface of Earth), the definition of a vector as an arrow connecting two points (Secs. 1.2 and 2.2.1) is suspect, as it is not obvious on what route the arrow should travel nor that the linear algebra of tensor analysis should be valid for such arrows. In this section, we refine the concept of a vector to deal with this problem. In the process we introduce the concept of a *tangent space* in which the linear algebra of tensors takes place—a different tangent space for tensors that live at different points in the space.

Third, once we have been forced to think of a tensor as residing in a specific tangent space at a specific point in the space, the question arises: how can one transport tensors from the tangent space at one point to the tangent space at an adjacent point? Since the notion of a gradient of a vector depends on comparing the vector at two different points and thus depends on the details of transport, we have to rework the notion of a gradient and the gradient's connection coefficients.

Fourth, when doing an integral, one must add contributions that live at different points in the space, so we must also rework the notion of integration.

We tackle each of these four issues in turn in the following four subsections.

24.3.1 Nonorthonormal Bases

24.3.1

Consider an n-dimensional *manifold,* that is, a space that, in the neighborhood of any point, has the same topological and smoothness properties as n-dimensional Euclidean space, though it might not have a locally Euclidean or locally Lorentz metric and perhaps has no metric at all. If the manifold has a metric (e.g., 4-dimensional spacetime, 3-dimensional Euclidean space, and the 2-dimensional surface of a sphere) it is called "Riemannian." In this chapter, all manifolds we consider will be Riemannian.

manifold

At some point $\mathcal{P}$ in our chosen n-dimensional manifold with metric, introduce a set of basis vectors $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ and denote them generally as $\vec{e}_\alpha$. We seek to generalize the formalism of Sec. 24.2 in such a way that the index-manipulation rules for components of tensors are unchanged. For example, we still want it to be true that covariant components of any tensor are computable by inserting the basis vectors into the tensor's slots, $F_{\alpha\beta} = \boldsymbol{F}(\vec{e}_\alpha, \vec{e}_\beta)$, and that the tensor itself can be reconstructed from its contravariant components: $\boldsymbol{F} = F^{\mu\nu}\vec{e}_\mu \otimes \vec{e}_\nu$. We also require that the two sets of components are computable from each other via raising and lowering with the metric components: $F_{\alpha\beta} = g_{\alpha\mu}g_{\beta\nu}F^{\mu\nu}$. The only thing we do not want to preserve is the orthonormal values of the metric components: we must allow the basis to be nonorthonormal and thus $\vec{e}_\alpha \cdot \vec{e}_\beta = g_{\alpha\beta}$ to have arbitrary values (except that the metric should be nondegenerate, so no linear combination of the $\vec{e}_\alpha$s vanishes, which means that the matrix $||g_{\alpha\beta}||$ should have nonzero determinant).

tensors in a nonorthonormal basis

We can easily achieve our goal by introducing a second set of basis vectors, denoted $\{\vec{e}^1, \vec{e}^2, \ldots, \vec{e}^n\}$, which is *dual* to our first set in the sense that

dual sets of basis vectors

$$\boxed{\vec{e}^\mu \cdot \vec{e}_\beta \equiv \boldsymbol{g}(\vec{e}^\mu, \vec{e}_\beta) = \delta^\mu{}_\beta.} \tag{24.8}$$

Here $\delta^\alpha{}_\beta$ is the Kronecker delta. This duality relation actually constitutes a *definition* of the e^μ once the $\vec{e}_\alpha$ have been chosen. To see this, regard $\vec{e}^\mu$ as a tensor of rank one. This tensor is defined as soon as its value on each and every vector has been determined. Expression (24.8) gives the value $\vec{e}^\mu(\vec{e}_\beta) = \vec{e}^\mu \cdot \vec{e}_\beta$ of $\vec{e}^\mu$ on each of the four basis vectors $\vec{e}_\beta$; and since every other vector can be expanded in terms of the $\vec{e}_\beta$s and $\vec{e}^\mu(\_\_)$ is a linear function, Eq. (24.8) thereby determines the value of $\vec{e}^\mu$ on every other vector.

The duality relation (24.8) says that $\vec{e}^1$ is always perpendicular to all the $\vec{e}_\alpha$s except $\vec{e}_1$, and its scalar product with $\vec{e}_1$ is unity—and similarly for the other basis vectors. This interpretation is illustrated for 3-dimensional Euclidean space in Fig. 24.2. In Minkowski spacetime, if the $\vec{e}_\alpha$ are an orthonormal Lorentz basis, then duality dictates that $\vec{e}^0 = -\vec{e}_0$, and $\vec{e}^j = +\vec{e}_j$.

The duality relation (24.8) leads immediately to the same index-manipulation formalism as we have been using, if one defines the contravariant, covariant, and mixed components of tensors in the obvious manner:

covariant, contravariant, and mixed components of a tensor

$$\boxed{F^{\mu\nu} = \boldsymbol{F}(\vec{e}^\mu, \vec{e}^\nu), \quad F_{\alpha\beta} = \boldsymbol{F}(\vec{e}_\alpha, \vec{e}_\beta), \quad F^\mu{}_\beta = \boldsymbol{F}(\vec{e}^\mu, \vec{e}_\beta);} \tag{24.9}$$

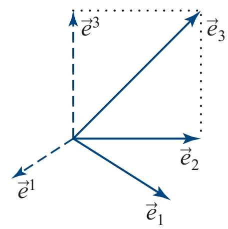

FIGURE 24.2 Nonorthonormal basis vectors $\vec{e}_j$ in Euclidean 3-space and two members $\vec{e}^1$ and $\vec{e}^3$ of the dual basis. The vectors $\vec{e}_1$ and $\vec{e}_2$ lie in the horizontal plane, so $\vec{e}^3$ is orthogonal to that plane (i.e., it points vertically upward), and its inner product with $\vec{e}_3$ is unity. Similarly, the vectors $\vec{e}_2$ and $\vec{e}_3$ span a vertical plane, so $\vec{e}^1$ is orthogonal to that plane (i.e., it points horizontally), and its inner product with $\vec{e}_1$ is unity.

see Ex. 24.4. Among the consequences of this duality are the following:

covariant and contravariant components of the metric

1. The matrix of contravariant components of the metric is inverse to that of the covariant components, $||g^{\mu\nu}|| = ||g_{\alpha\beta}||^{-1}$, so that

$$\boxed{g^{\mu\beta} g_{\beta\nu} = \delta^\mu{}_\nu.} \tag{24.10}$$

This relation guarantees that when one raises an index on a tensor $F_{\alpha\beta}$ with $g^{\mu\beta}$ and then lowers it back down with $g_{\beta\mu}$, one recovers one's original covariant components $F_{\alpha\beta}$ unaltered.

reconstructing a tensor from its components

2. One can reconstruct a tensor from its components by lining up the indices in a manner that accords with the rules of index manipulation:

$$\boxed{\mathbf{F} = F^{\mu\nu}\vec{e}_\mu \otimes \vec{e}_\nu = F_{\alpha\beta}\vec{e}^\alpha \otimes \vec{e}^\beta = F^\mu{}_\beta \vec{e}_\mu \otimes \vec{e}^\beta.} \tag{24.11}$$

component equations are same as slot-naming-index-notation equations

3. The component versions of tensorial equations are identical in mathematical symbology to the slot-naming-index-notation versions:

$$\boxed{\mathbf{F}(\vec{p}, \vec{q}) = F^{\alpha\beta} p_\alpha p_\beta.} \tag{24.12}$$

Associated with any coordinate system $x^\alpha(\mathcal{P})$ there is a *coordinate basis* whose basis vectors are defined by

coordinate basis

$$\boxed{\vec{e}_\alpha \equiv \frac{\partial \mathcal{P}}{\partial x^\alpha}.} \tag{24.13}$$

Since the derivative is taken holding the other coordinates fixed, the basis vector $\vec{e}_\alpha$ points along the α coordinate axis (the axis on which x^α changes and all the other coordinates are held fixed).

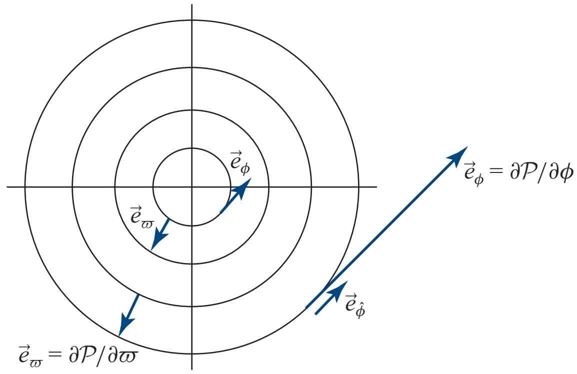

FIGURE 24.3 A circular coordinate system $\{\varpi, \phi\}$ and its coordinate basis vectors $\vec{e}_\varpi = \partial\mathcal{P}/\partial\varpi$, $\vec{e}_\phi = \partial\mathcal{P}/\partial\phi$ at several locations in the coordinate system. Also shown is the orthonormal basis vector $\vec{e}_{\hat\phi}$.

orthogonal curvilinear coordinates

In an orthogonal curvilinear coordinate system [e.g., circular polar coordinates (ϖ, ϕ) in Euclidean 2-space; Fig. 24.3], this coordinate basis is quite different from the coordinate system's orthonormal basis. For example, $\vec{e}_\phi = (\partial\mathcal{P}/\partial\phi)_\varpi$ is a very long vector at large radii and a very short one at small radii; the corresponding unit-length vector is $\vec{e}_{\hat\phi} = (1/\varpi)\vec{e}_\phi = (1/\varpi)\partial/\partial\phi$ (i.e., the derivative with respect to physical distance along the ϕ direction). By contrast, $\vec{e}_\varpi = (\partial\mathcal{P}/\partial\varpi)_\phi$ already has unit length, so the corresponding orthonormal basis vector is simply $\vec{e}_{\hat\varpi} = \vec{e}_\varpi$. The metric components in the coordinate basis are readily seen to be $g_{\phi\phi} = \varpi^2$, $g_{\varpi\varpi} = 1$, and $g_{\varpi\phi} = g_{\phi\varpi} = 0$, which are in accord with the equation for the squared distance (interval) between adjacent points: $ds^2 = g_{ij}dx^i dx^j = d\varpi^2 + \varpi^2 d\phi^2$. Of course, the metric components in the orthonormal basis are $g_{\hat i \hat j} = \delta_{ij}$.

Henceforth, we use hats to identify orthonormal bases; bases whose indices do not have hats will typically (though not always) be coordinate bases.

We can construct the basis $\{\vec{e}^\mu\}$ that is dual to the coordinate basis $\{\vec{e}_\alpha\} = \{\partial\mathcal{P}/\partial x^\alpha\}$ by taking the gradients of the coordinates, viewed as scalar fields $x^\alpha(\mathcal{P})$:

the basis dual to a coordinate basis

$$\boxed{\vec{e}^\mu = \vec\nabla x^\mu.} \tag{24.14}$$

It is straightforward to verify the duality relation (24.8) for these two bases:

$$\vec{e}^\mu \cdot \vec{e}_\alpha = \vec{e}_\alpha \cdot \vec\nabla x^\mu = \nabla_{\vec{e}_\alpha} x^\mu = \nabla_{\partial\mathcal{P}/\partial x^\alpha} x^\mu = \frac{\partial x^\mu}{\partial x^\alpha} = \delta^\mu{}_\alpha. \tag{24.15}$$

In any coordinate system, the expansion of the metric in terms of the dual basis, $\boldsymbol{g} = g_{\alpha\beta}\vec{e}^\alpha \otimes \vec{e}^\beta = g_{\alpha\beta}\vec\nabla x^\alpha \otimes \vec\nabla x^\beta$, is intimately related to the line element $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$. Consider an infinitesimal vectorial displacement $d\vec{x} = dx^\alpha(\partial/\partial x^\alpha)$. Insert this displacement into the metric's two slots to obtain the interval ds^2 along

$d\vec{x}$. The result is $ds^2 = g_{\alpha\beta}\nabla x^\alpha \otimes \nabla x^\beta(d\vec{x}, d\vec{x}) = g_{\alpha\beta}(d\vec{x}\cdot\nabla x^\alpha)(d\vec{x}\cdot\nabla x^\beta) = g_{\alpha\beta}dx^\alpha dx^\beta$:

the line element for the invariant interval along a displacement vector

$$\boxed{ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta.} \tag{24.16}$$

Here the second equality follows from the definition of the tensor product $\otimes$, and the third from the fact that for any scalar field ψ, $d\vec{x}\cdot\nabla\psi$ is the change $d\psi$ along $d\vec{x}$.

Any two bases $\{\vec{e}_\alpha\}$ and $\{\vec{e}_{\bar{\mu}}\}$ can be expanded in terms of each other:

transformation matrices linking two bases

$$\boxed{\vec{e}_\alpha = \vec{e}_{\bar{\mu}} L^{\bar{\mu}}{}_\alpha, \quad \vec{e}_{\bar{\mu}} = \vec{e}_\alpha L^\alpha{}_{\bar{\mu}}.} \tag{24.17}$$

(By convention the first index on L is always placed up, and the second is always placed down.) The quantities $||L^{\bar{\mu}}{}_\alpha||$ and $||L^\alpha{}_{\bar{\mu}}||$ are transformation matrices, and since they operate in opposite directions, they must be the inverse of each other:

$$\boxed{L^{\bar{\mu}}{}_\alpha L^\alpha{}_{\bar{\nu}} = \delta^{\bar{\mu}}{}_{\bar{\nu}}, \quad L^\alpha{}_{\bar{\mu}} L^{\bar{\mu}}{}_\beta = \delta^\alpha{}_\beta.} \tag{24.18}$$

These $||L^{\bar{\mu}}{}_\alpha||$ are the generalizations of Lorentz transformations to arbitrary bases [cf. Eqs. (2.34) and (2.35a)]. As in the Lorentz-transformation case, the transformation laws (24.17) for the basis vectors imply corresponding transformation laws for components of vectors and tensors—laws that entail lining up indices in the obvious manner:

transformation of tensor components between bases

$$\boxed{\begin{gathered} A_{\bar{\mu}} = L^\alpha{}_{\bar{\mu}} A_\alpha, \quad T^{\bar{\mu}\bar{\nu}}{}_{\bar{\rho}} = L^{\bar{\mu}}{}_\alpha L^{\bar{\nu}}{}_\beta L^\gamma{}_{\bar{\rho}} T^{\alpha\beta}{}_\gamma, \\ \text{and similarly in the opposite direction.} \end{gathered}} \tag{24.19}$$

For coordinate bases, these $L^{\bar{\mu}}{}_\alpha$ are simply the partial derivatives of one set of coordinates with respect to the other:

transformation matrices between coordinate bases

$$\boxed{L^{\bar{\mu}}{}_\alpha = \frac{\partial x^{\bar{\mu}}}{\partial x^\alpha}, \quad L^\alpha{}_{\bar{\mu}} = \frac{\partial x^\alpha}{\partial x^{\bar{\mu}}},} \tag{24.20}$$

as one can easily deduce via

$$\vec{e}_\alpha = \frac{\partial \mathcal{P}}{\partial x^\alpha} = \frac{\partial x^\mu}{\partial x^\alpha}\frac{\partial \mathcal{P}}{\partial x^\mu} = \vec{e}_\mu \frac{\partial x^\mu}{\partial x^\alpha}. \tag{24.21}$$

In many physics textbooks a tensor is *defined* as a set of components $F_{\alpha\beta}$ that obey the transformation laws

$$F_{\alpha\beta} = F_{\mu\nu}\frac{\partial x^\mu}{\partial x^\alpha}\frac{\partial x^\nu}{\partial x^\beta}. \tag{24.22}$$

This definition (valid only in a coordinate basis) is in accord with Eqs. (24.19) and (24.20), though it hides the true and very simple nature of a tensor as a linear function of frame-independent vectors.

EXERCISES

Exercise 24.4 *Derivation: Index-Manipulation Rules from Duality*

For an arbitrary basis $\{\vec{e}_\alpha\}$ and its dual basis $\{\vec{e}^\mu\}$, use (i) the duality relation (24.8), (ii) the definition (24.9) of components of a tensor, and (iii) the relation $\vec{A} \cdot \vec{B} = \boldsymbol{g}(\vec{A}, \vec{B})$ between the metric and the inner product to deduce the following results.

(a) The relations

$$\vec{e}^\mu = g^{\mu\alpha}\vec{e}_\alpha, \quad \vec{e}_\alpha = g_{\alpha\mu}\vec{e}^\mu. \tag{24.23}$$

(b) The fact that indices on the components of tensors can be raised and lowered using the components of the metric:

$$F^{\mu\nu} = g^{\mu\alpha}F_\alpha{}^\nu, \quad p_\alpha = g_{\alpha\beta}p^\beta. \tag{24.24}$$

(c) The fact that a tensor can be reconstructed from its components in the manner of Eq. (24.11).

Exercise 24.5 *Practice: Transformation Matrices for Circular Polar Bases*

Consider the circular polar coordinate system $\{\varpi, \phi\}$ and its coordinate bases and orthonormal bases as shown in Fig. 24.3 and discussed in the associated text. These coordinates are related to Cartesian coordinates $\{x, y\}$ by the usual relations: $x = \varpi \cos\phi$, $y = \varpi \sin\phi$.

(a) Evaluate the components ($L^x{}_\varpi$, etc.) of the transformation matrix that links the two coordinate bases $\{\vec{e}_x, \vec{e}_y\}$ and $\{\vec{e}_\varpi, \vec{e}_\phi\}$. Also evaluate the components ($L^\varpi{}_x$, etc.) of the inverse transformation matrix.

(b) Similarly, evaluate the components of the transformation matrix and its inverse linking the bases $\{\vec{e}_x, \vec{e}_y\}$ and $\{\vec{e}_{\hat\varpi}, \vec{e}_{\hat\phi}\}$.

(c) Consider the vector $\vec{A} \equiv \vec{e}_x + 2\vec{e}_y$. What are its components in the other two bases?

24.3.2 Vectors as Directional Derivatives; Tangent Space; Commutators

24.3.2

As discussed in the introduction to Sec. 24.3, the notion of a vector as an arrow connecting two points is problematic in a curved manifold and must be refined. As a first step in the refinement, let us consider the tangent vector $\vec{A}$ to a curve $\mathcal{P}(\zeta)$ at some point $\mathcal{P}_o \equiv \mathcal{P}(\zeta = 0)$. We have defined that tangent vector by the limiting process:

tangent vector to a curve

$$\boxed{\vec{A} \equiv \frac{d\mathcal{P}}{d\zeta} \equiv \lim_{\Delta\zeta \to 0} \frac{\mathcal{P}(\Delta\zeta) - \mathcal{P}(0)}{\Delta\zeta}} \tag{24.25}$$

[Eq. (24.2)]. In this definition the difference $\mathcal{P}(\zeta) - \mathcal{P}(0)$ means the tiny arrow reaching from $\mathcal{P}(0) \equiv \mathcal{P}_o$ to $\mathcal{P}(\Delta\zeta)$. In the limit as $\Delta\zeta$ becomes vanishingly small, these two points get arbitrarily close together. In such an arbitrarily small region of the manifold, the effects of the manifold's curvature become arbitrarily small and

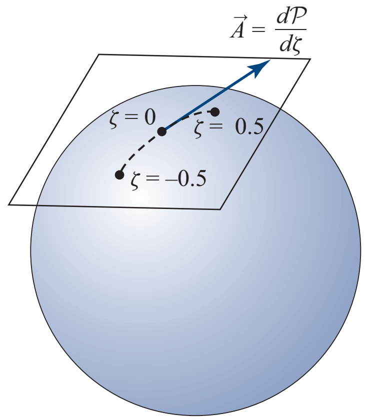

FIGURE 24.4 A curve $\mathcal{P}(\zeta)$ on the surface of a sphere and the curve's tangent vector $\vec{A} = d\mathcal{P}/d\zeta$ at $\mathcal{P}(\zeta = 0) \equiv \mathcal{P}_o$. The tangent vector lives in the tangent space at $\mathcal{P}_o$ (i.e., in the flat plane that is tangent to the sphere there, as seen in the flat Euclidean 3-space in which the sphere's surface is embedded).

negligible (just think of an arbitrarily tiny region on the surface of a sphere), so the notion of the arrow should become sensible. However, before the limit is completed, we are required to divide by $\Delta\zeta$, which makes our arbitrarily tiny arrow big again. What meaning can we give to this?

One way to think about it is to imagine embedding the curved manifold in a higher-dimensional flat space (e.g., embed the surface of a sphere in a flat 3-dimensional Euclidean space, as shown in Fig. 24.4). Then the tiny arrow $\mathcal{P}(\Delta\zeta) - \mathcal{P}(0)$ can be thought of equally well as lying on the sphere, or as lying in a surface that is tangent to the sphere and is flat, as measured in the flat embedding space. We can give meaning to $[\mathcal{P}(\Delta\zeta) - \mathcal{P}(0)]/\Delta\zeta$ if we regard this expression as a formula for lengthening an arrow-type vector in the flat tangent surface; correspondingly, we must regard the resulting tangent vector $\vec{A}$ as an arrow living in the tangent surface.

tangent space at a point

The (conceptual) flat tangent surface at the point $\mathcal{P}_o$ is called the *tangent space* to the curved manifold at that point. It has the same number of dimensions n as the manifold itself (two in the case of the surface of the sphere in Fig. 24.4). Vectors at $\mathcal{P}_o$ are arrows residing in that point's tangent space, tensors at $\mathcal{P}_o$ are linear functions of these vectors, and all the linear algebra of vectors and tensors that reside at $\mathcal{P}_o$ occurs in this tangent space. For example, the inner product of two vectors $\vec{A}$ and $\vec{B}$ at $\mathcal{P}_o$ (two arrows living in the tangent space there) is computed via the standard relation $\vec{A} \cdot \vec{B} = \boldsymbol{g}(\vec{A}, \vec{B})$ using the metric $\boldsymbol{g}$ that also resides in the tangent space. (Scalars reside in both the manifold and the tangent space.)

This pictorial way of thinking about the tangent space and vectors and tensors that reside in it is far too heuristic to satisfy most mathematicians. Therefore, mathematicians have insisted on making it much more precise at the price of greater abstraction. Mathematicians define the tangent vector to the curve $\mathcal{P}(\zeta)$ to be the derivative $d/d\zeta$

that differentiates scalar fields along the curve. This derivative operator is well defined by the rules of ordinary differentiation: if $\psi(\mathcal{P})$ is a scalar field in the manifold, then $\psi[\mathcal{P}(\zeta)]$ is a function of the real variable ζ, and its derivative $(d/d\zeta)\psi[\mathcal{P}(\zeta)]$ evaluated at $\zeta = 0$ is the ordinary derivative of elementary calculus. Since the derivative operator $d/d\zeta$ differentiates in the manifold along the direction in which the curve is moving, it is often called the *directional derivative* along $\mathcal{P}(\zeta)$. Mathematicians notice that all the directional derivatives at a point $\mathcal{P}_o$ of the manifold form a vector space (they can be multiplied by scalars and added and subtracted to get new vectors), and so the mathematicians define this vector space to be the tangent space at $\mathcal{P}_o$.

directional derivative

This mathematical procedure turns out to be isomorphic to the physicists' more heuristic way of thinking about the tangent space. In physicists' language, if one introduces a coordinate system in a region of the manifold containing $\mathcal{P}_o$ and constructs the corresponding coordinate basis $\vec{e}_\alpha = \partial\mathcal{P}/\partial x^\alpha$, then one can expand any vector in the tangent space as $\vec{A} = A^\alpha \partial\mathcal{P}/\partial x^\alpha$. One can also construct, in physicists' language, the directional derivative along $\vec{A}$; it is $\partial_{\vec{A}} \equiv A^\alpha \partial/\partial x^\alpha$. Evidently, the components A^α of the physicist's vector $\vec{A}$ (an arrow) are identical to the coefficients A^α in the coordinate-expansion of the directional derivative $\partial_{\vec{A}}$. Therefore a one-to-one correspondence exists between the directional derivatives $\partial_{\vec{A}}$ at $\mathcal{P}_o$ and the vectors $\vec{A}$ there, and a complete isomorphism holds between the tangent-space manipulations that a mathematician performs treating the directional derivatives as vectors, and those that a physicist performs treating the arrows as vectors.

"Why not abandon the fuzzy concept of a vector as an arrow, and *redefine the vector $\vec{A}$ to be the same as the directional derivative $\partial_{\vec{A}}$*?" mathematicians have demanded of physicists. Slowly, over the past century, physicists have come to see the merit in this approach. (i) It does, indeed, make the concept of a vector more rigorous than before. (ii) It simplifies a number of other concepts in mathematical physics (e.g., the commutator of two vector fields; see below). (iii) It facilitates communication with mathematicians. (iv) It provides a formalism that is useful for calculation. With these motivations in mind, and because one always gains conceptual and computational power by having multiple viewpoints at one's fingertips (see Feynman, 1966, p. 160), we henceforth shall regard vectors both as arrows living in a tangent space and as directional derivatives. Correspondingly, we assert the equalities:

tangent vector as directional derivative along a curve

$$\boxed{\frac{\partial\mathcal{P}}{\partial x^\alpha} = \frac{\partial}{\partial x^\alpha}\ , \quad \vec{A} = \partial_{\vec{A}},} \tag{24.26}$$

and often expand vectors in a coordinate basis using the notation

$$\boxed{\vec{A} = A^\alpha \frac{\partial}{\partial x^\alpha}.} \tag{24.27}$$

This directional-derivative viewpoint on vectors makes natural the concept of the *commutator* of two vector fields $\vec{A}$ and $\vec{B}$: $[\vec{A}, \vec{B}]$ is the vector that, when viewed

commutator of two vector fields

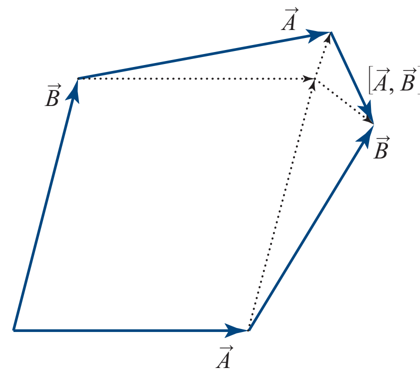

FIGURE 24.5 The commutator $[\vec{A}, \vec{B}]$ of two vector fields. The vectors are assumed to be so small that the curvature of the manifold is negligible in the region of the diagram, so all the vectors can be drawn lying in the manifold itself rather than in their respective tangent spaces. In evaluating the two terms in the commutator (24.28), a locally orthonormal coordinate basis is used, so $A^\alpha \partial B^\beta/\partial x^\alpha$ is the amount by which the vector $\vec{B}$ changes when one travels along $\vec{A}$ (i.e., it is the rightward-and-downward pointing dashed arrow in the upper right), and $B^\alpha \partial A^\beta/\partial x^\alpha$ is the amount by which $\vec{A}$ changes when one travels along $\vec{B}$ (i.e., it is the rightward-and-upward pointing dashed arrow). According to Eq. (24.28), the difference of these two dashed arrows is the commutator $[\vec{A}, \vec{B}]$. As the diagram shows, this commutator closes the quadrilateral whose legs are $\vec{A}$ and $\vec{B}$. If the commutator vanishes, then there is no gap in the quadrilateral, which means that in the region covered by this diagram, one can construct a coordinate system in which $\vec{A}$ and $\vec{B}$ are coordinate basis vectors.

as a differential operator, is given by $[\partial_{\vec{A}}, \partial_{\vec{B}}]$—where the latter quantity is the same commutator as one meets elsewhere in physics (e.g., in quantum mechanics). Using this definition, we can compute the components of the commutator in a coordinate basis:

$$\boxed{[\vec{A}, \vec{B}] \equiv \left[A^\alpha \frac{\partial}{\partial x^\alpha}, B^\beta \frac{\partial}{\partial x^\beta}\right] = \left(A^\alpha \frac{\partial B^\beta}{\partial x^\alpha} - B^\alpha \frac{\partial A^\beta}{\partial x^\alpha}\right) \frac{\partial}{\partial x^\beta}.} \tag{24.28}$$

This is an operator equation where the final derivative is presumed to operate on a scalar field, just as in quantum mechanics. From this equation we can read off the components of the commutator in any coordinate basis; they are $A^\alpha B^\beta{}_{,\alpha} - B^\alpha A^\beta{}_{,\alpha}$, where the comma denotes partial differentiation. Figure 24.5 uses this equation to deduce the geometric meaning of the commutator: it is the fifth leg needed to close a quadrilateral whose other four legs are constructed from the vector fields $\vec{A}$ and $\vec{B}$. In other words, it is "the change in $\vec{B}$ relative to $\vec{A}$," and as such it is a type of derivative of $\vec{B}$ along $\vec{A}$, called the *Lie derivative*: $\mathcal{L}_{\vec{A}}\vec{B} \equiv [\vec{A}, \vec{B}]$ (cf. footnote 2 in Chap. 14).

coordinate bases have vanishing commutators

The commutator is useful as a tool for distinguishing between coordinate bases and noncoordinate bases (also called nonholonomic bases). In a coordinate basis, the basis vectors are just the coordinate system's partial derivatives, $\vec{e}_\alpha = \partial/\partial x^\alpha$, and since partial derivatives commute, it must be that $[\vec{e}_\alpha, \vec{e}_\beta] = 0$. Conversely (as Fig. 24.5 shows), if one has a basis with vanishing commutators $[\vec{e}_\alpha, \vec{e}_\beta] = 0$, then it is possible to construct a coordinate system for which this is the coordinate basis. In a noncoordinate basis, at least one of the commutators $[\vec{e}_\alpha, \vec{e}_\beta]$ will be nonzero.

24.3.3 Differentiation of Vectors and Tensors; Connection Coefficients

24.3.3

In a curved manifold, the differentiation of vectors and tensors is rather subtle. To elucidate the problem, let us recall how we defined such differentiation in Minkowski spacetime or Euclidean space (Sec. 1.7). Converting to the notation used in Eq. (24.25), we began by defining the directional derivative of a tensor field $\boldsymbol{F}(\mathcal{P})$ along the tangent vector $\vec{A} = d/d\zeta$ to a curve $\mathcal{P}(\zeta)$:

directional derivative of a tensor field

$$\nabla_{\vec{A}}\boldsymbol{F} \equiv \lim_{\Delta\zeta\to 0} \frac{\boldsymbol{F}[\mathcal{P}(\Delta\zeta)] - \boldsymbol{F}[\mathcal{P}(0)]}{\Delta\zeta}. \tag{24.29}$$

This definition is problematic, because $\boldsymbol{F}[\mathcal{P}(\Delta\zeta)]$ lives in a different tangent space than does $\boldsymbol{F}[\mathcal{P}(0)]$. To make the definition meaningful, we must identify some connection between the two tangent spaces, when their points $\mathcal{P}(\Delta\zeta)$ and $\mathcal{P}(0)$ are arbitrarily close together. That connection is equivalent to identifying a rule for transporting $\boldsymbol{F}$ from one tangent space to the other.

In flat space or flat spacetime, and when $\boldsymbol{F}$ is a vector $\vec{F}$, that transport rule is obvious: keep $\vec{F}$ parallel to itself and keep its length fixed during the transport. In other words, keep constant its components in an orthonormal coordinate system (Cartesian coordinates in Euclidean space, Lorentz coordinates in Minkowski spacetime). This is called the *law of parallel transport*. For a tensor $\boldsymbol{F}$, the parallel transport law is the same: keep its components fixed in an orthonormal coordinate basis.

Now, just as the curvature of Earth's surface prevents one from placing a Cartesian coordinate system on it, so nonzero curvature of any other manifold prevents one from introducing orthonormal coordinates; see Sec. 25.3. However, in an arbitrarily small region on Earth's surface, one can introduce coordinates that are arbitrarily close to Cartesian (as surveyors well know); the fractional deviations from Cartesian need be no larger than $O(L^2/R^2)$, where L is the size of the region and R is Earth's radius (see Sec. 25.3). Similarly, in curved spacetime, in an arbitrarily small region, one can introduce coordinates that are arbitrarily close to Lorentz, differing only by amounts quadratic in the size of the region—and similarly for a *local* orthonormal coordinate basis in any curved manifold.

When defining $\nabla_{\vec{A}}\boldsymbol{F}$, one is sensitive only to first-order changes of quantities, not second, so the parallel transport used in defining it in a flat manifold, based on constancy of components in an orthonormal coordinate basis, must also work in a *local* orthonormal coordinate basis of any curved manifold: In Eq. (24.29), one must transport $\boldsymbol{F}$ from $\mathcal{P}(\Delta\zeta)$ to $\mathcal{P}(0)$, holding its components fixed in a locally orthonormal coordinate basis (parallel transport), and then take the difference in the tangent space at $\mathcal{P}_o = \mathcal{P}(0)$, divide by $\Delta\zeta$, and let $\Delta\zeta \to 0$. The result is a tensor at $\mathcal{P}_o$: the directional derivative $\nabla_{\vec{A}}\boldsymbol{F}$ of $\boldsymbol{F}$.

gradient of a tensor field

Having made the directional derivative meaningful, one can proceed as in Secs. 1.7 and 2.10: define the gradient of $\boldsymbol{F}$ by $\nabla_{\vec{A}}\boldsymbol{F} = \vec{\nabla}\boldsymbol{F}(\_\_, \_\_, \vec{A})$ [i.e., put $\vec{A}$ in the last—differentiation—slot of $\vec{\nabla}\boldsymbol{F}$; Eq. (1.15b)].

As in Chap. 2, in any basis we denote the components of $\vec{\nabla}\boldsymbol{F}$ by $F_{\alpha\beta;\gamma}$. And as in Sec. 11.8 (elasticity theory), we can compute these components in any basis with the aid of that basis's *connection coefficients*.

In Sec. 11.8, we restricted ourselves to an orthonormal basis in Euclidean space and thus had no need to distinguish between covariant and contravariant indices; all indices were written as subscripts. Now, dealing with nonorthonormal bases in spacetime, we must distinguish covariant and contravariant indices. Accordingly, by analogy with Eq. (11.68), we define the connection coefficients $\Gamma^\mu{}_{\alpha\beta}$ as

connection coefficients for a basis and its dual

$$\boxed{\nabla_\beta \vec{e}_\alpha \equiv \nabla_{\vec{e}_\beta}\vec{e}_\alpha = \Gamma^\mu{}_{\alpha\beta}\vec{e}_\mu .} \tag{24.30}$$

The duality between bases $\vec{e}^{\,\nu} \cdot \vec{e}_\alpha = \delta^\nu{}_\alpha$ then implies

$$\boxed{\nabla_\beta \vec{e}^{\,\mu} \equiv \nabla_{\vec{e}_\beta}\vec{e}^{\,\mu} = -\Gamma^\mu{}_{\alpha\beta}\vec{e}^{\,\alpha} .} \tag{24.31}$$

Note the sign flip, which is required to keep $\nabla_\beta(\vec{e}^{\,\mu} \cdot \vec{e}_\alpha) = 0$, and note that the differentiation index always goes last on Γ. Duality also implies that Eqs. (24.30) and (24.31) can be rewritten as

$$\Gamma^\mu{}_{\alpha\beta} = \vec{e}^{\,\mu} \cdot \nabla_\beta \vec{e}_\alpha = -\vec{e}_\alpha \cdot \nabla_\beta \vec{e}^{\,\mu} . \tag{24.32}$$

With the aid of these connection coefficients, we can evaluate the components $A_{\alpha;\beta}$ of the gradient of a vector field in any basis. We just compute

$$\begin{aligned} A^\mu{}_{;\beta}\vec{e}_\mu &= \nabla_\beta \vec{A} = \nabla_\beta(A^\mu \vec{e}_\mu) = (\nabla_\beta A^\mu)\vec{e}_\mu + A^\mu \nabla_\beta \vec{e}_\mu \\ &= A^\mu{}_{,\beta}\vec{e}_\mu + A^\mu \Gamma^\alpha{}_{\mu\beta}\vec{e}_\alpha \\ &= (A^\mu{}_{,\beta} + A^\alpha \Gamma^\mu{}_{\alpha\beta})\vec{e}_\mu . \end{aligned} \tag{24.33}$$

In going from the first line to the second, we have used the notation

$$A^\mu{}_{,\beta} \equiv \partial_{\vec{e}_\beta} A^\mu ; \tag{24.34}$$

that is, *the comma denotes the result of letting a basis vector act as a differential operator on the component of the vector*. In going from the second line of (24.33) to the third, we have renamed some summed-over indices. By comparing the first and last expressions in Eq. (24.33), we conclude that

components of the gradient of a vector field

$$\boxed{A^\mu{}_{;\beta} = A^\mu{}_{,\beta} + A^\alpha \Gamma^\mu{}_{\alpha\beta} .} \tag{24.35}$$

The first term in this equation describes the changes in $\vec{A}$ associated with changes of its component A^μ; the second term *corrects for* artificial changes of A^μ that are induced by turning and length changes of the basis vector $\vec{e}_\mu$. We shall use the short-hand terminology that the second term "corrects the index μ."

By a similar computation, we conclude that in any basis the covariant components of the gradient are

$$\boxed{A_{\alpha;\beta} = A_{\alpha,\beta} - \Gamma^{\mu}{}_{\alpha\beta} A_{\mu},} \tag{24.36}$$

where again $A_{\alpha,\beta} \equiv \partial_{\vec{e}_\beta} A_\alpha$. Notice that, when the index being corrected is down [α in Eq. (24.36)], the connection coefficient has a minus sign; when it is up [μ in Eq. (24.35)], the connection coefficient has a plus sign. This is in accord with the signs in Eqs. (24.30) and (24.31).

These considerations should make obvious the following equations for the components of the gradient of a second rank tensor field:

components of the gradient of a tensor field

$$\begin{aligned}
F^{\alpha\beta}{}_{;\gamma} &= F^{\alpha\beta}{}_{,\gamma} + \Gamma^{\alpha}{}_{\mu\gamma} F^{\mu\beta} + \Gamma^{\beta}{}_{\mu\gamma} F^{\alpha\mu},\\
F_{\alpha\beta;\gamma} &= F_{\alpha\beta,\gamma} - \Gamma^{\mu}{}_{\alpha\gamma} F_{\mu\beta} - \Gamma^{\mu}{}_{\beta\gamma} F_{\alpha\mu},\\
F^{\alpha}{}_{\beta;\gamma} &= F^{\alpha}{}_{\beta,\gamma} + \Gamma^{\alpha}{}_{\mu\gamma} F^{\mu}{}_{\beta} - \Gamma^{\mu}{}_{\beta\gamma} F^{\alpha}{}_{\mu}.
\end{aligned} \tag{24.37}$$

Notice that each index of $\boldsymbol{F}$ must be corrected, the correction has a sign dictated by whether the index is up or down, the differentiation index always goes last on the Γ, and all other indices can be deduced by requiring that the free indices in each term be the same and all other indices be summed.

If we have been given a basis, then how can we compute the connection coefficients? We can try to do so by drawing pictures and examining how the basis vectors change from point to point—a method that is fruitful in spherical and cylindrical coordinates in Euclidean space (Sec. 11.8). However, in other situations this method is fraught with peril, so we need a firm mathematical prescription. It turns out that the following prescription works (see Ex. 24.7 for a proof).

1. Evaluate the commutation coefficients $c_{\alpha\beta}{}^{\rho}$ of the basis, which are defined by the two equivalent relations:

commutation coefficients for a basis

$$\boxed{[\vec{e}_\alpha, \vec{e}_\beta] \equiv c_{\alpha\beta}{}^{\rho} \vec{e}_\rho, \quad c_{\alpha\beta}{}^{\rho} \equiv \vec{e}^{\,\rho} \cdot [\vec{e}_\alpha, \vec{e}_\beta].} \tag{24.38a}$$

   (Note that in a coordinate basis the commutation coefficients will vanish. Warning: Commutation coefficients also appear in the theory of Lie groups; there it is conventional to use a different ordering of indices than here: $c_{\alpha\beta}{}^{\rho}{}_{\text{here}} = c^{\rho}{}_{\alpha\beta\,\text{Lie groups}}$.)

2. Lower the last index on the commutation coefficients using the metric components in the basis:

$$\boxed{c_{\alpha\beta\gamma} \equiv c_{\alpha\beta}{}^{\rho} g_{\rho\gamma}.} \tag{24.38b}$$

3. Compute the quantities

formulas for computing connection coefficients

$$\boxed{\Gamma_{\alpha\beta\gamma} \equiv \frac{1}{2}(g_{\alpha\beta,\gamma} + g_{\alpha\gamma,\beta} - g_{\beta\gamma,\alpha} + c_{\alpha\beta\gamma} + c_{\alpha\gamma\beta} - c_{\beta\gamma\alpha}).} \tag{24.38c}$$

Here the commas denote differentiation with respect to the basis vectors as though the metric components were scalar fields [as in Eq. (24.34)]. Notice that the pattern of indices is the same on the gs and on the cs. It is a peculiar pattern—one of the few aspects of index gymnastics that cannot be reconstructed by merely lining up indices. In a coordinate basis the c terms will vanish, so $\Gamma_{\alpha\beta\gamma}$ will be symmetric in its last two indices. In an orthonormal basis $g_{\mu\nu}$ are constant, so the g terms will vanish, and $\Gamma_{\alpha\beta\gamma}$ will be antisymmetric in its first two indices. And in a Cartesian or Lorentz coordinate basis, which is both coordinate and orthonormal, both the c terms and the g terms will vanish, so $\Gamma_{\alpha\beta\gamma}$ will vanish.

4. Raise the first index on $\Gamma_{\alpha\beta\gamma}$ to obtain the connection coefficients

$$\boxed{\Gamma^{\mu}{}_{\beta\gamma} = g^{\mu\alpha}\Gamma_{\alpha\beta\gamma}.} \tag{24.38d}$$

In a coordinate basis, the $\Gamma^{\mu}{}_{\beta\gamma}$ are sometimes called *Christoffel symbols*, though we will use the name connection coefficients independent of the nature of the basis.

The first three steps in the above prescription for computing the connection coefficients follow from two key properties of the gradient $\vec{\nabla}$. First, the gradient of the metric tensor vanishes:

vanishing gradient of the metric tensor

$$\boxed{\vec{\nabla}\boldsymbol{g} = 0.} \tag{24.39}$$

Second, for any two vector fields $\vec{A}$ and $\vec{B}$, the gradient is related to the commutator by

relation of gradient to commutator

$$\boxed{\nabla_{\vec{A}}\vec{B} - \nabla_{\vec{B}}\vec{A} = [\vec{A}, \vec{B}].} \tag{24.40}$$

For a derivation of these relations and then a derivation of the prescription 1–4, see Exs. 24.6 and 24.7.

The gradient operator $\vec{\nabla}$ is an example of a geometric object that is not a tensor. The connection coefficients $\Gamma^{\mu}{}_{\beta\gamma} = \vec{e}^{\mu} \cdot \left(\nabla_{\vec{e}_\gamma}\vec{e}_\beta\right)$ can be regarded as the components of $\vec{\nabla}$; because it is not a tensor, these components do not obey the tensorial transformation law (24.19) when switching from one basis to another. Their transformation law is far more complicated and is rarely used. Normally one computes them from scratch in the new basis, using the above prescription or some other, equivalent prescription (cf. Misner, Thorne, and Wheeler, 1973, Chap. 14). For most curved spacetimes that one meets in general relativity, these computations are long and tedious and therefore are normally carried out on computers using symbolic manipulation software, such as Maple, Matlab, or Mathematica, or such programs as GR-Tensor and MathTensor that run under Maple or Mathematica. Such software is easily found on the Internet using a search engine. A particularly simple Mathematica program for use with coordinate bases is presented and discussed in Appendix C of Hartle (2003) and is available on that book's website: http://web.physics.ucsb.edu/~gravitybook/.

EXERCISES

Exercise 24.6 *Derivation: Properties of the Gradient* $\vec{\nabla}$

(a) Derive Eq. (24.39). [Hint: At a point $\mathcal{P}$ where $\vec{\nabla}\boldsymbol{g}$ is to be evaluated, introduce a locally orthonormal coordinate basis (i.e., locally Cartesian or locally Lorentz). When computing in this basis, the effects of curvature show up only to second order in distance from $\mathcal{P}$. Show that in this basis, the components of $\vec{\nabla}\boldsymbol{g}$ vanish, and from this infer that $\vec{\nabla}\boldsymbol{g}$, viewed as a frame-independent third-rank tensor, vanishes.]

(b) Derive Eq. (24.40). [Hint: Again work in a locally orthonormal coordinate basis.]

Exercise 24.7 *Derivation and Example: Prescription for Computing Connection Coefficients*

Derive the prescription 1–4 [Eqs. (24.38)] for computing the connection coefficients in any basis. [Hints: (i) In the chosen basis, from $\vec{\nabla}\boldsymbol{g} = 0$ infer that $\Gamma_{\alpha\beta\gamma} + \Gamma_{\beta\alpha\gamma} = g_{\alpha\beta,\gamma}$. Notice that this determines the part of $\Gamma_{\alpha\beta\gamma}$ that is symmetric in its first two indices. Show that the number of independent components of $\Gamma_{\alpha\beta\gamma}$ thereby determined is $\frac{1}{2}n^2(n+1)$, where n is the manifold's dimension. (ii) From Eq. (24.40) infer that $\Gamma_{\gamma\beta\alpha} - \Gamma_{\gamma\alpha\beta} = c_{\alpha\beta\gamma}$, which fixes the part of Γ antisymmetric in the last two indices. Show that the number of independent components thereby determined is $\frac{1}{2}n^2(n-1)$. (iii) Infer that the number of independent components determined by (i) and (ii) together is n^3, which is the entirety of $\Gamma_{\alpha\beta\gamma}$. By somewhat complicated algebra, deduce Eq. (24.38c) for $\Gamma_{\alpha\beta\gamma}$. (The algebra is sketched in Misner, Thorne, and Wheeler, 1973, Ex. 8.15.) (iv) Then infer the final answer, Eq. (24.38d), for $\Gamma^{\mu}{}_{\beta\gamma}$.]

Exercise 24.8 *Practice: Commutation and Connection Coefficients for Circular Polar Bases*

Consider the circular polar coordinates $\{\varpi, \phi\}$ of Fig. 24.3 and their associated bases.

(a) Evaluate the commutation coefficients $c_{\alpha\beta}{}^{\rho}$ for the coordinate basis $\{\vec{e}_\varpi, \vec{e}_\phi\}$, and also for the orthonormal basis $\{\vec{e}_{\hat\varpi}, \vec{e}_{\hat\phi}\}$.

(b) Compute by hand the connection coefficients for the coordinate basis and also for the orthonormal basis, using Eqs. (24.38). [Note: The answer for the orthonormal basis was worked out pictorially in our study of elasticity theory; Fig. 11.15 and Eq. (11.70).]

(c) Repeat this computation using symbolic manipulation software on a computer.

Exercise 24.9 *Practice: Connection Coefficients for Spherical Polar Coordinates*

(a) Consider spherical polar coordinates in 3-dimensional space, and verify that the nonzero connection coefficients, assuming an orthonormal basis, are given by Eq. (11.71).

(b) Repeat the exercise in part (a) assuming a coordinate basis with

$$\mathbf{e}_r \equiv \frac{\partial}{\partial r}, \quad \mathbf{e}_\theta \equiv \frac{\partial}{\partial \theta}, \quad \mathbf{e}_\phi \equiv \frac{\partial}{\partial \phi}. \tag{24.41}$$

(c) Repeat both computations in parts (a) and (b) using symbolic manipulation software on a computer.

Exercise 24.10 *Practice: Index Gymnastics—Geometric Optics*
This exercise gives the reader practice in formal manipulations that involve the gradient operator. In the geometric-optics (eikonal) approximation of Sec. 7.3, for electromagnetic waves in Lorenz gauge, one can write the 4-vector potential in the form $\vec{A} = \vec{\mathcal{A}} e^{i\varphi}$, where $\vec{\mathcal{A}}$ is a slowly varying amplitude and φ is a rapidly varying phase. By the techniques of Sec. 7.3, one can deduce from the vacuum Maxwell equations that the wave vector, defined by $\vec{k} \equiv \vec{\nabla}\varphi$, is null: $\vec{k} \cdot \vec{k} = 0$.

(a) Rewrite all the equations in the above paragraph in slot-naming index notation.

(b) Using index manipulations, show that the wave vector $\vec{k}$ (which is a vector field, because the wave's phase φ is a scalar field) satisfies the geodesic equation $\nabla_{\vec{k}}\vec{k} = 0$ (cf. Sec. 24.5.2). The geodesics, to which $\vec{k}$ is the tangent vector, are the rays discussed in Sec. 7.3, along which the waves propagate.

24.3.4

24.3.4 Integration

Our desire to use general bases and work in curved manifolds gives rise to two new issues in the definition of integrals.

The first issue is that the volume elements used in integration involve the Levi-Civita tensor [Eqs. (2.43), (2.52), and (2.55)], so we need to know the components of the Levi-Civita tensor in a general basis. It turns out (see, e.g., Misner, Thorne, and Wheeler, 1973, Ex. 8.3) that the covariant components differ from those in an orthonormal basis by a factor $\sqrt{|g|}$ and the contravariant by $1/\sqrt{|g|}$, where

$$\boxed{g \equiv \det ||g_{\alpha\beta}||} \tag{24.42}$$

is the determinant of the matrix whose entries are the covariant components of the metric. More specifically, let us denote by $[\alpha\beta \ldots \nu]$ the value of $\epsilon_{\alpha\beta\ldots\nu}$ in an orthonormal basis of our n-dimensional space [Eq. (2.43)]:

$$[12\ldots n] = +1,$$
$$[\alpha\beta\ldots\nu] = \begin{cases} +1 & \text{if } \alpha, \beta, \ldots, \nu \text{ is an even permutation of } 1, 2, \ldots, n \\ -1 & \text{if } \alpha, \beta, \ldots, \nu \text{ is an odd permutation of } 1, 2, \ldots, n \\ 0 & \text{if } \alpha, \beta, \ldots, \nu \text{ are not all different.} \end{cases} \tag{24.43}$$

(In spacetime the indices must run from 0 to 3 rather than 1 to $n = 4$.) Then in a general right-handed basis the components of the Levi-Civita tensor are

$$\boxed{\epsilon_{\alpha\beta\ldots\nu} = \sqrt{|g|}\,[\alpha\beta\ldots\nu], \quad \epsilon^{\alpha\beta\ldots\nu} = \pm\frac{1}{\sqrt{|g|}}\,[\alpha\beta\ldots\nu],} \tag{24.44}$$

components of Levi-Civita tensor in an arbitrary basis

where the $\pm$ is plus in Euclidean space and minus in spacetime. In a left-handed basis the sign is reversed.

As an example of these formulas, consider a spherical polar coordinate system (r, θ, ϕ) in 3-dimensional Euclidean space, and use the three infinitesimal vectors $dx^j(\partial/\partial x^j)$ to construct the volume element $d\Sigma$ [cf. Eq. (1.26)]:

$$dV = \boldsymbol{\epsilon}\left(dr\frac{\partial}{\partial r}, d\theta\frac{\partial}{\partial\theta}, d\phi\frac{\partial}{\partial\phi}\right) = \epsilon_{r\theta\phi}drd\theta d\phi = \sqrt{g}\,drd\theta d\phi = r^2\sin\theta drd\theta d\phi. \tag{24.45}$$

Here the second equality follows from linearity of $\boldsymbol{\epsilon}$ and the formula for computing its components by inserting basis vectors into its slots; the third equality follows from our formula (24.44) for the components. The fourth equality entails the determinant of the metric coefficients, which in spherical coordinates are $g_{rr} = 1$, $g_{\theta\theta} = r^2$, and $g_{\phi\phi} = r^2\sin^2\theta$; all other g_{jk} vanish, so $g = r^4\sin^2\theta$. The resulting volume element $r^2\sin\theta drd\theta d\phi$ should be familiar and obvious.

The second new integration issue we must face is that such integrals as

$$\int_{\partial\mathcal{V}} T^{\alpha\beta} d\Sigma_\beta \tag{24.46}$$

[cf. Eqs. (2.55), (2.56)] involve constructing a vector $T^{\alpha\beta}d\Sigma_\beta$ in each infinitesimal region $d\Sigma_\beta$ of the surface of integration $\partial\mathcal{V}$ and then adding up the contributions from all the infinitesimal regions. A major difficulty arises because each contribution lives in a different tangent space. To add them together, we must first transport them all to the same tangent space at some single location in the manifold. How is that transport to be performed? The obvious answer is "by the same parallel transport technique that we used in defining the gradient." However, when defining the gradient, we only needed to perform the parallel transport over an infinitesimal distance, and now we must perform it over long distances. When the manifold is curved, long-distance parallel transport gives a result that depends on the route of the transport, and in general there is no way to identify any preferred route (see, e.g., Misner, Thorne, and Wheeler, 1973, Sec. 11.4).

integrals in a curved manifold are well defined only if infinitesimal contributions are scalars

As a result, *integrals such as Eq. (24.46) are ill-defined in a curved manifold. The only integrals that are well defined in a curved manifold are those such as* $\int_{\partial\mathcal{V}} S^\alpha d\Sigma_\alpha$*, whose infinitesimal contributions* $S^\alpha d\Sigma_\alpha$ *are scalars* (i.e., integrals whose value is a scalar). This fact will have profound consequences in curved spacetime for the laws of conservation of energy, momentum, and angular momentum (Secs. 25.7 and 25.9.4).

EXERCISES

Exercise 24.11 *Practice: Integration—Gauss's Theorem*

In 3-dimensional Euclidean space Maxwell's equation $\nabla \cdot \mathbf{E} = \rho_e/\epsilon_0$ can be combined with Gauss's theorem to show that the electric flux through the surface $\partial\mathcal{V}$ of a sphere is equal to the charge in the sphere's interior $\mathcal{V}$ divided by ϵ_0:

$$\int_{\partial\mathcal{V}} \mathbf{E} \cdot d\mathbf{\Sigma} = \int_{\mathcal{V}} (\rho_e/\epsilon_0)\, dV. \tag{24.47}$$

Introduce spherical polar coordinates so the sphere's surface is at some radius $r = R$. Consider a surface element on the sphere's surface with vectorial legs $d\phi\,\partial/\partial\phi$ and $d\theta\,\partial/\partial\theta$. Evaluate the components $d\Sigma_j$ of the surface integration element $d\mathbf{\Sigma} = \boldsymbol{\epsilon}(\ldots, d\theta\,\partial/\partial\theta, d\phi\,\partial/\partial\phi)$. (Here $\boldsymbol{\epsilon}$ is the Levi-Civita tensor.) Similarly, evaluate dV in terms of vectorial legs in the sphere's interior. Then use these results for $d\Sigma_j$ and dV to convert Eq. (24.47) into an explicit form in terms of integrals over r, θ, and ϕ. The final answer should be obvious, but the above steps in deriving it are informative.

24.4

24.4 The Stress-Energy Tensor Revisited

In Sec. 2.13.1, we defined the stress-energy tensor $\boldsymbol{T}$ of any matter or field as a symmetric, second-rank tensor that describes the flow of 4-momentum through spacetime. More specifically, the total 4-momentum $\vec{P}$ that flows through some small 3-volume $\vec{\Sigma}$ (defined in Sec. 2.12.1), going from the negative side of $\vec{\Sigma}$ to its positive side, is

stress-energy tensor

$$\boldsymbol{T}(\_\_, \vec{\Sigma}) = (\text{total 4-momentum } \vec{P} \text{ that flows through } \vec{\Sigma}); \quad T^{\alpha\beta}\Sigma_\beta = P^\alpha \tag{24.48}$$

[Eq. (2.66)]. Of course, this stress-energy tensor depends on the location $\mathcal{P}$ of the 3-volume in spacetime [i.e., it is a tensor field $\boldsymbol{T}(\mathcal{P})$].

From this geometric, frame-independent definition of the stress-energy tensor, we were able to read off the physical meaning of its components in any inertial reference frame [Eqs. (2.67)]: T^{00} is the total energy density, including rest mass-energy; $T^{j0} = T^{0j}$ is the j-component of momentum density, or equivalently, the j-component of energy flux; and T^{jk} are the components of the stress tensor, or equivalently, of the momentum flux.

In Sec. 2.13.2, we formulated the law of conservation of 4-momentum in a local form and a global form. The local form,

local form of 4-momentum conservation

$$\vec{\nabla} \cdot \boldsymbol{T} = 0, \tag{24.49}$$

says that, in any chosen Lorentz frame, the time derivative of the energy density plus the divergence of the energy flux vanishes, $\partial T^{00}/\partial t + \partial T^{0j}/\partial x^j = 0$, and similarly for the momentum, $\partial T^{j0}/\partial t + \partial T^{jk}/\partial x^k = 0$. The global form, $\int_{\partial\mathcal{V}} T^{\alpha\beta} d\Sigma_\beta = 0$ [Eq. (2.71)], says that all the 4-momentum that enters a closed 4-volume $\mathcal{V}$ in spacetime through its boundary $\partial\mathcal{V}$ in the past must ultimately exit through $\partial\mathcal{V}$ in the future (Fig. 2.11). Unfortunately, this global form requires transporting vectorial contributions $T^{\alpha\beta} d\Sigma_\beta$ to a common location and adding them, which cannot be done in a route-independent way in curved spacetime (see the end of Sec. 24.3.4). Therefore (as we shall discuss in greater detail in Secs. 25.7 and 25.9.4), the global conservation law becomes problematic in curved spacetime.

The stress-energy tensor and local 4-momentum conservation play major roles in our development of general relativity. Almost all of our examples will entail perfect fluids.

Recall [Eq. (2.74a)] that in the local rest frame of a perfect fluid, there is no energy flux or momentum density, $T^{j0} = T^{0j} = 0$, but there is a total energy density (including rest mass) ρ and an isotropic pressure P:

$$T^{00} = \rho, \quad T^{jk} = P\delta^{jk}. \tag{24.50}$$

From this special form of $T^{\alpha\beta}$ in the fluid's local rest frame, one can derive a geometric, frame-independent expression for the fluid's stress-energy tensor $\boldsymbol{T}$ in terms of its 4-velocity $\vec{u}$, the metric tensor $\boldsymbol{g}$, and the rest-frame energy density ρ and pressure P:

stress-energy tensor for a perfect fluid

$$\boxed{\boldsymbol{T} = (\rho + P)\vec{u} \otimes \vec{u} + P\boldsymbol{g}; \quad T^{\alpha\beta} = (\rho + P)u^\alpha u^\beta + P g^{\alpha\beta}} \tag{24.51}$$

[Eq. (2.74b)]; see Ex. 2.26. This expression for the stress-energy tensor of a perfect fluid is an example of a geometric, frame-independent description of physics.

The equations of relativistic fluid dynamics for a perfect fluid are obtained by inserting the stress-energy tensor (24.51) into the law of 4-momentum conservation $\vec{\nabla} \cdot \boldsymbol{T} = 0$, and augmenting with the law of rest-mass conservation. We explored this in brief in Ex. 2.26, and in much greater detail in Sec. 13.8. Applications that we have explored are the relativistic Bernoulli equation and ultrarelativistic jets (Sec. 13.8.2) and relativistic shocks (Ex. 17.9). In Sec. 13.8.3, we explored in detail the slightly subtle way in which a fluid's nonrelativistic energy density, energy flux, and stress tensor arise from the relativistic perfect-fluid stress-energy tensor (24.51).

These issues for a perfect fluid are so important that readers are encouraged to review them (except possibly the applications) in preparation for our foray into general relativity.

Four other examples of the stress-energy tensor are those for the electromagnetic field (Ex. 2.28), for a kinetic-theory swarm of relativistic particles (Secs. 3.4.2 and 3.5.3), for a point particle (Box 24.2), and for a relativistic fluid with viscosity and diffusive heat conduction (Ex. 24.13). However, we shall not do much with any of these during our study of general relativity, except viscosity and heat conduction in Sec. 28.5.

T2

BOX 24.2. STRESS-ENERGY TENSOR FOR A POINT PARTICLE T2

For a point particle that moves through spacetime along a world line $\mathcal{P}(\zeta)$ [where ζ is the affine parameter such that the particle's 4-momentum is $\vec{p} = d/d\zeta$, Eq. (2.14)], the stress-energy tensor vanishes everywhere except on the world line itself. Correspondingly, $\boldsymbol{T}$ must be expressed in terms of a Dirac delta function. The relevant delta function is a scalar function of two points in spacetime, $\delta(\mathcal{Q}, \mathcal{P})$, with the property that when one integrates over the point $\mathcal{P}$, using the 4-dimensional volume element $d\Sigma$ (which in any inertial frame just reduces to $d\Sigma = dtdxdydz$), one obtains

$$\boxed{\int_{\mathcal{V}} f(\mathcal{P})\delta(\mathcal{Q}, \mathcal{P})d\Sigma = f(\mathcal{Q}).} \qquad (1)$$

Here $f(\mathcal{P})$ is an arbitrary scalar field, and the region $\mathcal{V}$ of 4-dimensional integration must include the point $\mathcal{Q}$. One can easily verify that in terms of Lorentz coordinates this delta function can be expressed as

$$\boxed{\delta(\mathcal{Q}, \mathcal{P}) = \delta(t_{\mathcal{Q}} - t_{\mathcal{P}})\delta(x_{\mathcal{Q}} - x_{\mathcal{P}})\delta(y_{\mathcal{Q}} - y_{\mathcal{P}})\delta(z_{\mathcal{Q}} - z_{\mathcal{P}}),} \qquad (2)$$

where the deltas on the right-hand side are ordinary 1-dimensional Dirac delta functions. [Proof: Simply insert Eq. (2) into Eq. (1), replace $d\Sigma$ by $dt_{\mathcal{Q}}dx_{\mathcal{Q}}dy_{\mathcal{Q}}dz_{\mathcal{Q}}$, and perform the four integrations.]

The general definition (24.48) of the stress-energy tensor $\boldsymbol{T}$ implies that the integral of a point particle's stress-energy tensor over any 3-surface $\mathcal{S}$ that slices through the particle's world line just once, at an event $\mathcal{P}(\zeta_o)$, must be equal to the particle's 4-momentum at the intersection point:

$$\int_{\mathcal{S}} T^{\alpha\beta} d\Sigma_\beta = p^\alpha(\zeta_o). \qquad (3)$$

It is a straightforward but sophisticated exercise (Ex. 24.12) to verify that the following frame-independent expression has this property:

$$\boxed{\boldsymbol{T}(\mathcal{Q}) = \int_{-\infty}^{+\infty} \vec{p}(\zeta) \otimes \vec{p}(\zeta)\, \delta[\mathcal{Q}, \mathcal{P}(\zeta)]\, d\zeta.} \qquad (4)$$

Here the integral is along the world line $\mathcal{P}(\zeta)$ of the particle, and $\mathcal{Q}$ is the point at which $\boldsymbol{T}$ is being evaluated. Therefore, Eq. (4) is the point-particle stress-energy tensor.

EXERCISES

T2

Exercise 24.12 *Derivation: Stress-Energy Tensor for a Point Particle* T2
Show that the point-particle stress-energy tensor (4) of Box 24.2 satisfies that box's Eq. (3), as claimed.

Exercise 24.13 *Example: Stress-Energy Tensor for a Viscous Fluid with Diffusive Heat Conduction*
This exercise serves two roles: It develops the relativistic stress-energy tensor for a viscous fluid with diffusive heat conduction, and in the process it allows the reader to gain practice in index gymnastics.

In our study of elasticity theory, we introduced the concept of the irreducible tensorial parts of a second-rank tensor in Euclidean space (Box 11.2). Consider a relativistic fluid flowing through spacetime with a 4-velocity $\vec{u}(\mathcal{P})$. The fluid's gradient $\vec{\nabla}\vec{u}$ ($u_{\alpha;\beta}$ in slot-naming index notation) is a second-rank tensor in spacetime. With the aid of the 4-velocity itself, we can break it down into irreducible tensorial parts as follows:

$$u_{\alpha;\beta} = -a_\alpha u_\beta + \frac{1}{3}\theta P_{\alpha\beta} + \sigma_{\alpha\beta} + \omega_{\alpha\beta}. \tag{24.52}$$

Here: (i)

$$P_{\alpha\beta} \equiv g_{\alpha\beta} + u_\alpha u_\beta \tag{24.53}$$

is a tensor that projects vectors into the 3-space orthogonal to $\vec{u}$ (it can also be regarded as that 3-space's metric; see Ex. 2.10); (ii) $\sigma_{\alpha\beta}$ is symmetric, trace-free, and orthogonal to the 4-velocity; and (iii) $\omega_{\alpha\beta}$ is antisymmetric and orthogonal to the 4-velocity.

(a) Show that the rate of change of $\vec{u}$ along itself, $\nabla_{\vec{u}}\vec{u}$ (i.e., the fluid 4-acceleration) is equal to the vector $\vec{a}$ that appears in the decomposition (24.52). Show, further, that $\vec{a} \cdot \vec{u} = 0$.

(b) Show that the divergence of the 4-velocity, $\boldsymbol{\nabla} \cdot \vec{u}$, is equal to the scalar field θ that appears in the decomposition (24.52). As we shall see in part (d), this is the fluid's rate of expansion.

(c) The quantities $\sigma_{\alpha\beta}$ and $\omega_{\alpha\beta}$ are the relativistic versions of a Newtonian fluid's shear and rotation tensors, which we introduced in Sec. 13.7.1. Derive equations for these tensors in terms of $u_{\alpha;\beta}$ and $P_{\mu\nu}$.

(d) Show that, as viewed in a Lorentz reference frame where the fluid is moving with speed small compared to the speed of light, to first order in the fluid's ordinary velocity $v^j = dx^j/dt$, the following statements are true: (i) $u^0 = 1$, $u^j = v^j$; (ii) θ is the nonrelativistic rate of expansion of the fluid, $\theta = \boldsymbol{\nabla} \cdot \mathbf{v} \equiv v^j{}_{,j}$ [Eq. (13.67a)]; (iii) σ_{jk} is the fluid's nonrelativistic shear [Eq. (13.67b)]; and (iv) ω_{jk} is the fluid's nonrelativistic rotation tensor [denoted r_{ij} in Eq. (13.67c)].

(e) At some event $\mathcal{P}$ where we want to know the influence of viscosity on the fluid's stress-energy tensor, introduce the fluid's local rest frame. Explain why, in that frame, the only contributions of viscosity to the components of the stress-energy tensor are $T^{jk}_{\text{visc}} = -\zeta\theta g^{jk} - 2\mu\sigma^{jk}$, where ζ and μ are the coefficients of bulk and shear viscosity, respectively; the contributions to T^{00} and $T^{j0} = T^{0j}$ vanish. [Hint: See Eq. (13.73) and associated discussions.]

(f) From nonrelativistic fluid mechanics, infer that, in the fluid's rest frame at $\mathcal{P}$, the only contributions of diffusive heat conductivity to the stress-energy tensor are $T^{0j}_{\text{cond}} = T^{j0}_{\text{cond}} = -\kappa\,\partial T/\partial x^j$, where κ is the fluid's thermal conductivity and T is its temperature. [Hint: See Eq. (13.74) and associated discussion.] Actually, this expression is not fully correct. If the fluid is accelerating, there is a correction term: $\partial T/\partial x^j$ gets replaced by $\partial T/\partial x^j + a^j T$, where a^j is the acceleration. After reading Sec. 24.5 and especially Ex. 24.16, explain this correction.

(g) Using the results of parts (e) and (f), deduce the following geometric, frame-invariant form of the fluid's stress-energy tensor:

$$T_{\alpha\beta} = (\rho + P)u_\alpha u_\beta + P g_{\alpha\beta} - \zeta\theta g_{\alpha\beta} - 2\mu\sigma_{\alpha\beta} - 2\kappa u_{(\alpha} P_{\beta)}{}^{\mu}(T_{;\mu} + a_\mu T). \quad (24.54)$$

Here the subscript parentheses in the last term mean to symmetrize in the α and β slots.

From the divergence of this stress-energy tensor, plus the first law of thermodynamics and the law of rest-mass conservation, one can derive the full theory of relativistic fluid mechanics for a fluid with viscosity and heat flow (see, e.g., Misner, Thorne, and Wheeler, 1973, Ex. 22.7). This particular formulation of the theory, including Eq. (24.54), is due to Carl Eckart (1940). Landau and Lifshitz (1959) have given a slightly different formulation. For discussion of the differences, and of causal difficulties with both formulations and the difficulties' repair, see, for example, the reviews by Israel and Stewart (1980), Andersson and Comer (2007, Sec. 14), and López-Monsalvo (2011, Sec. 4).

24.5

24.5 The Proper Reference Frame of an Accelerated Observer

Physics experiments and astronomical measurements almost always use an apparatus that accelerates and rotates. For example, if the apparatus is in an Earthbound laboratory and is attached to the laboratory floor and walls, then it accelerates upward (relative to freely falling particles) with the negative of the "acceleration of gravity," and it rotates (relative to inertial gyroscopes) because of the rotation of Earth. It is useful, in studying such an apparatus, to regard it as attached to an accelerating, rotating reference frame. As preparation for studying such reference frames in the presence of gravity, we study them in flat spacetime. For a somewhat more sophisticated treatment, see Misner, Thorne, and Wheeler (1973, pp. 163–176, 327–332).

Consider an observer with 4-velocity $\vec{U}$, who moves along an accelerated world line through flat spacetime (Fig. 24.6) so she has a nonzero 4-acceleration:

$$\vec{a} = \vec{\nabla}_{\vec{U}}\vec{U}. \quad (24.55)$$

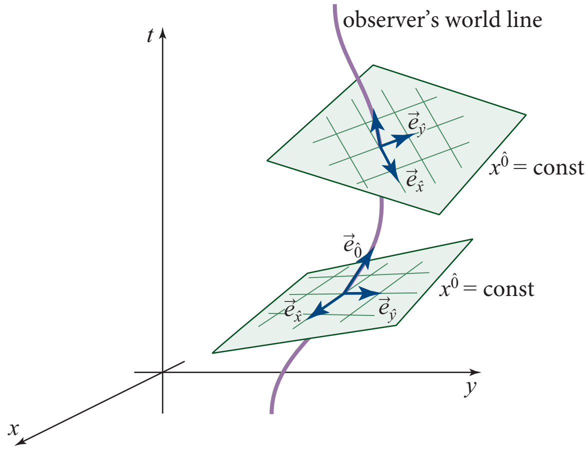

FIGURE 24.6 The proper reference frame of an accelerated observer. The spatial basis vectors $\vec{e}_{\hat{x}}$, $\vec{e}_{\hat{y}}$, and $\vec{e}_{\hat{z}}$ are orthogonal to the observer's world line and rotate, relative to local gyroscopes, as they move along the world line. The flat 3-planes spanned by these basis vectors are surfaces of constant coordinate time $x^{\hat{0}} \equiv$ (proper time as measured by the observer's clock at the event where the 3-plane intersects the observer's world line); in other words, they are the observer's slices of simultaneity and "3-space." In each of these flat 3-planes the spatial coordinates $\{\hat{x}, \hat{y}, \hat{z}\}$ are Cartesian, with $\partial/\partial\hat{x} = \vec{e}_{\hat{x}}$, $\partial/\partial\hat{y} = \vec{e}_{\hat{y}}$, and $\partial/\partial\hat{z} = \vec{e}_{\hat{z}}$.

proper reference frame of an accelerated observer

Have that observer construct, in the vicinity of her world line, a coordinate system $\{x^{\hat{\alpha}}\}$ (called her *proper reference frame*) with these properties: (i) The spatial origin is centered on her world line at all times (i.e., her world line is given by $x^{\hat{j}} = 0$). (ii) Along her world line, the time coordinate $x^{\hat{0}}$ is the same as the proper time ticked by an ideal clock that she carries. (iii) In the immediate vicinity of her world line, the spatial coordinates $x^{\hat{j}}$ measure physical distance along the axes of a little Cartesian latticework that she carries (and that she regards as purely spatial, which means it lies in the 3-plane orthogonal to her world line). These properties dictate that, in the immediate vicinity of her world line, the metric has the form $ds^2 = \eta_{\hat{\alpha}\hat{\beta}} dx^{\hat{\alpha}} dx^{\hat{\beta}}$, where $\eta_{\hat{\alpha}\hat{\beta}}$ are the Lorentz-basis metric coefficients, Eq. (24.6); in other words, all along her world line the coordinate basis vectors are orthonormal:

$$g_{\hat{\alpha}\hat{\beta}} = \frac{\partial}{\partial x^{\hat{\alpha}}} \cdot \frac{\partial}{\partial x^{\hat{\beta}}} = \eta_{\hat{\alpha}\hat{\beta}} \quad \text{at } x^{\hat{j}} = 0. \tag{24.56}$$

Moreover, properties (i) and (ii) dictate that along the observer's world line, the basis vector $\vec{e}_{\hat{0}} \equiv \partial/\partial x^{\hat{0}}$ differentiates with respect to her proper time, and thus is identically equal to her 4-velocity $\vec{U}$:

$$\vec{e}_{\hat{0}} = \frac{\partial}{\partial x^{\hat{0}}} = \vec{U}. \tag{24.57}$$

rotating and nonrotating proper reference frames

There remains freedom as to how the observer's latticework is oriented spatially. The observer can lock it to the gyroscopes of an *inertial-guidance system* that she carries (Box 24.3), in which case we say that it is "nonrotating"; or she can rotate it relative to such gyroscopes. For generality, we assume that the latticework rotates.

BOX 24.3. INERTIAL GUIDANCE SYSTEMS

Aircraft and rockets often carry inertial guidance systems, which consist of an accelerometer and a set of gyroscopes.

The accelerometer measures the system's 4-acceleration $\vec{a}$ (in relativistic language). Equivalently, it measures the system's Newtonian 3-acceleration **a** relative to inertial coordinates in which the system is momentarily at rest. As we see in Eq. (24.58), these quantities are two different ways of thinking about the same thing.

Each gyroscope is constrained to remain at rest in the aircraft or rocket by a force that is applied at its center of mass. Such a force exerts no torque around the center of mass, so the gyroscope maintains its direction (does not precess) relative to an inertial frame in which it is momentarily at rest.

As the accelerating aircraft or rocket turns, its walls rotate with some angular velocity $\vec{\Omega}$ relative to these inertial-guidance gyroscopes. This is the angular velocity discussed in the text between Eqs. (24.57) and (24.58).

From the time-evolving 4-acceleration $\vec{a}(\tau)$ and angular velocity $\vec{\Omega}(\tau)$, a computer can calculate the aircraft's (or rocket's) world line and its changing orientation.

Its angular velocity, as measured by the observer (by comparing the latticework's orientation with inertial-guidance gyroscopes), is a 3-dimensional spatial vector $\boldsymbol{\Omega}$ in the 3-plane orthogonal to her world line; and as viewed in 4-dimensional spacetime, it is a 4-vector $\vec{\Omega}$ whose components in the observer's reference frame are $\Omega^{\hat{j}} \neq 0$ and $\Omega^{\hat{0}} = 0$. Similarly, the latticework's acceleration, as measured by an inertial-guidance accelerometer attached to it (Box 24.3), is a 3-dimensional spatial vector **a** that can be thought of as a 4-vector with components in the observer's frame:

$$a^{\hat{0}} = 0, \quad a^{\hat{j}} = (\hat{j}\text{-component of the measured } \mathbf{a}). \tag{24.58}$$

This 4-vector is the observer's 4-acceleration, as one can verify by computing the 4-acceleration in an inertial frame in which the observer is momentarily at rest.

constructing coordinates of proper reference frame

Geometrically, the coordinates of the proper reference frame are constructed as follows. Begin with the basis vectors $\vec{e}_{\hat{\alpha}}$ along the observer's world line (Fig. 24.6)—basis vectors that satisfy Eqs. (24.56) and (24.57), and that rotate with angular velocity $\vec{\Omega}$ relative to gyroscopes. Through the observer's world line at time $x^{\hat{0}}$ construct the flat 3-plane spanned by the spatial basis vectors $\vec{e}_{\hat{j}}$. Because $\vec{e}_{\hat{j}} \cdot \vec{e}_{\hat{0}} = 0$, this 3-plane is orthogonal to the world line. All events in this 3-plane are given the same value of coordinate time $x^{\hat{0}}$ as the event where it intersects the world line; thus the 3-plane is a surface of constant coordinate time $x^{\hat{0}}$. The spatial coordinates in this flat 3-plane are ordinary, Cartesian coordinates $x^{\hat{j}}$ with $\vec{e}_{\hat{j}} = \partial/\partial x^{\hat{j}}$.

24.5.1 Relation to Inertial Coordinates; Metric in Proper Reference Frame; Transport Law for Rotating Vectors

24.5.1

It is instructive to examine the coordinate transformation between these proper-reference-frame coordinates $x^{\hat{\alpha}}$ and the coordinates x^{μ} of an inertial reference frame. We pick a very special inertial frame for this purpose. Choose an event on the observer's world line, near which the coordinate transformation is to be constructed; adjust the origin of the observer's proper time, so this event is $x^{\hat{0}} = 0$ (and of course $x^{\hat{j}} = 0$); and choose the inertial frame to be one that, arbitrarily near this event, coincides with the observer's proper reference frame. If we were doing Newtonian physics, then the coordinate transformation from the proper reference frame to the inertial frame would have the form (accurate through terms quadratic in $x^{\hat{\alpha}}$):

$$x^i = x^{\hat{i}} + \frac{1}{2}a^{\hat{i}}(x^{\hat{0}})^2 + \epsilon^{\hat{i}}_{\hat{j}\hat{k}}\Omega^{\hat{j}}x^{\hat{k}}x^{\hat{0}}, \quad x^0 = x^{\hat{0}}. \tag{24.59}$$

Here the term $\frac{1}{2}a^{\hat{i}}(x^{\hat{0}})^2$ is the standard expression for the vectorial displacement produced after time $x^{\hat{0}}$ by the acceleration $a^{\hat{i}}$; and the term $\epsilon^{\hat{i}}_{\hat{j}\hat{k}}\Omega^{\hat{j}}x^{\hat{k}}x^{\hat{0}}$ is the standard expression for the displacement produced by the rotation rate (rotational angular velocity) $\Omega^{\hat{j}}$ during a short time $x^{\hat{0}}$. In relativity theory there is only one departure from these familiar expressions (up through quadratic order): after time $x^{\hat{0}}$ the acceleration has produced a velocity $v^{\hat{j}} = a^{\hat{j}}x^{\hat{0}}$ of the proper reference frame relative to the inertial frame; correspondingly, there is a Lorentz-boost correction to the transformation of time: $x^0 = x^{\hat{0}} + v^{\hat{j}}x^{\hat{j}} = x^{\hat{0}}(1 + a_{\hat{j}}x^{\hat{j}})$ [cf. Eq. (2.37c)], accurate only to quadratic order. Thus, the full transformation to quadratic order is

inertial coordinates related to those of the proper reference frame of an accelerated, rotating observer

$$x^i = x^{\hat{i}} + \frac{1}{2}a^{\hat{i}}(x^{\hat{0}})^2 + \epsilon^{\hat{i}}{}_{\hat{j}\hat{k}}\Omega^{\hat{j}}x^{\hat{k}}x^{\hat{0}},$$
$$x^0 = x^{\hat{0}}(1 + a_{\hat{j}}x^{\hat{j}}). \tag{24.60a}$$

From this transformation and the form of the metric, $ds^2 = -(dx^0)^2 + \delta_{ij}dx^i dx^j$ in the inertial frame, we easily can evaluate the form of the metric, accurate to linear order in $\mathbf{x}$, in the proper reference frame:

metric in proper reference frame of an accelerated, rotating observer

$$\boxed{ds^2 = -(1 + 2\mathbf{a}\cdot\mathbf{x})(dx^{\hat{0}})^2 + 2(\boldsymbol{\Omega}\times\mathbf{x})\cdot d\mathbf{x}\, dx^{\hat{0}} + \delta_{jk}dx^{\hat{j}}dx^{\hat{k}}} \tag{24.60b}$$

(Ex. 24.14a). Here the notation is that of 3-dimensional vector analysis, with $\mathbf{x}$ the 3-vector whose components are $x^{\hat{j}}$, $d\mathbf{x}$ that with components $dx^{\hat{j}}$, $\mathbf{a}$ that with components $a^{\hat{j}}$, and $\boldsymbol{\Omega}$ that with components $\Omega^{\hat{j}}$.

Because the transformation (24.60a) was constructed near an arbitrary event on the observer's world line, the metric (24.60b) is valid near any and every event on the world line (i.e., it is valid all along the world line). In fact, it is the leading order in an expansion in powers of the spatial separation $x^{\hat{j}}$ from the world line. For higher-order terms in this expansion see, for example, Ni and Zimmermann (1978).

Notice that precisely on the observer's world line, the metric coefficients $g_{\hat{\alpha}\hat{\beta}}$ [the coefficients of $dx^{\hat{\alpha}}dx^{\hat{\beta}}$ in Eq. (24.60b)] are $g_{\hat{\alpha}\hat{\beta}} = \eta_{\hat{\alpha}\hat{\beta}}$, in accord with Eq. (24.56). However, as one moves farther away from the observer's world line, the effects of the acceleration $a^{\hat{j}}$ and rotation $\Omega^{\hat{j}}$ cause the metric coefficients to deviate more and more strongly from $\eta_{\hat{\alpha}\hat{\beta}}$.

From the metric coefficients of Eq. (24.60b), one can compute the connection coefficients $\Gamma^{\hat{\alpha}}{}_{\hat{\beta}\hat{\gamma}}$ on the observer's world line, and from these connection coefficients, one can infer the rates of change of the basis vectors along the world line: $\nabla_{\vec{U}}\vec{e}_{\hat{\alpha}} = \nabla_{\hat{0}}\vec{e}_{\hat{\alpha}} = \Gamma^{\hat{\mu}}{}_{\hat{\alpha}\hat{0}}\vec{e}_{\hat{\mu}}$. The result is (Ex. 24.14b):

equations for transport of proper reference frame's basis vectors along observer's world line

$$\nabla_{\vec{U}}\vec{e}_{\hat{0}} \equiv \nabla_{\vec{U}}\vec{U} = \vec{a}, \tag{24.61a}$$

$$\nabla_{\vec{U}}\vec{e}_{\hat{j}} = (\vec{a}\cdot\vec{e}_{\hat{j}})\vec{U} + \boldsymbol{\epsilon}(\vec{U}, \vec{\Omega}, \vec{e}_{\hat{j}}, \_\_). \tag{24.61b}$$

Equation (24.61b) is the general "law of transport" for constant-length vectors that are orthogonal to the observer's world line and that the observer thus sees as purely spatial. For the spin vector $\vec{S}$ of an inertial-guidance gyroscope (Box 24.3), the transport law is Eq. (24.61b) with $\vec{e}_{\hat{j}}$ replaced by $\vec{S}$ and with $\vec{\Omega} = 0$:

Fermi-Walker transport for the spin of an inertial-guidance gyroscope

$$\boxed{\nabla_{\vec{U}}\vec{S} = \vec{U}(\vec{a}\cdot\vec{S}).} \tag{24.62}$$

This is called *Fermi-Walker transport*. The term on the right-hand side of this transport law is required to keep the spin vector always orthogonal to the observer's 4-velocity: $\nabla_{\vec{U}}(\vec{S}\cdot\vec{U}) = 0$. For any other vector $\vec{A}$ that rotates relative to inertial-guidance gyroscopes, the transport law has, in addition to this "keep-it-orthogonal-to $\vec{U}$" term, a second term, which is the 4-vector form of $d\mathbf{A}/dt = \boldsymbol{\Omega}\times\mathbf{A}$:

transport law for a vector that is orthogonal to observer's 4-velocity and rotates relative to gyroscopes

$$\nabla_{\vec{U}}\vec{A} = \vec{U}(\vec{a}\cdot\vec{A}) + \boldsymbol{\epsilon}(\vec{U},\ \vec{\Omega},\ \vec{A},\ \_\_). \tag{24.63}$$

Equation (24.61b) is this general transport law with $\vec{A}$ replaced by $\vec{e}_{\hat{j}}$.

24.5.2

24.5.2 Geodesic Equation for a Freely Falling Particle

Consider a particle with 4-velocity $\vec{u}$ that moves freely through the neighborhood of an accelerated observer. As seen in an inertial reference frame, the particle travels through spacetime on a straight line, also called a *geodesic* of flat spacetime. Correspondingly, a geometric, frame-independent version of its *geodesic law of motion* is

geodesic law of motion for freely falling particle

$$\boxed{\nabla_{\vec{u}}\vec{u} = 0} \tag{24.64}$$

(i.e., the particle parallel transports its 4-velocity $\vec{u}$ along $\vec{u}$). It is instructive to examine the component form of this geodesic equation in the proper reference frame of the observer. Since the components of $\vec{u}$ in this frame are $u^{\alpha} = dx^{\alpha}/d\tau$, where τ is the particle's proper time (not the observer's proper time), the components $u^{\hat{\alpha}}{}_{;\hat{\mu}}u^{\hat{\mu}} = 0$ of the geodesic equation (24.64) are

$$u^{\hat{\alpha}}{}_{,\hat{\mu}}u^{\hat{\mu}} + \Gamma^{\hat{\alpha}}{}_{\hat{\mu}\hat{\nu}}u^{\hat{\mu}}u^{\hat{\nu}} = \left(\frac{\partial}{\partial x^{\hat{\mu}}}\frac{dx^{\hat{\alpha}}}{d\tau}\right)\frac{dx^{\hat{\mu}}}{d\tau} + \Gamma^{\hat{\alpha}}{}_{\hat{\mu}\hat{\nu}}u^{\hat{\mu}}u^{\hat{\nu}} = 0; \tag{24.65}$$

or equivalently,

$$\boxed{\frac{d^2x^{\hat{\alpha}}}{d\tau^2} + \Gamma^{\hat{\alpha}}{}_{\hat{\mu}\hat{\nu}}\frac{dx^{\hat{\mu}}}{d\tau}\frac{dx^{\hat{\nu}}}{d\tau} = 0.} \tag{24.66}$$

Suppose, for simplicity, that the particle is moving slowly relative to the observer, so its ordinary velocity $v^{\hat{j}} = dx^{\hat{j}}/dx^{\hat{0}}$ is nearly equal to $u^{\hat{j}} = dx^{\hat{j}}/d\tau$ and is small compared to unity (the speed of light), and $u^{\hat{0}} = dx^{\hat{0}}/d\tau$ is nearly unity. Then to first order in the ordinary velocity $v^{\hat{j}}$, the spatial part of the geodesic equation (24.66) becomes

$$\frac{d^2x^{\hat{i}}}{(dx^{\hat{0}})^2} = -\Gamma^{\hat{i}}{}_{\hat{0}\hat{0}} - (\Gamma^{\hat{i}}{}_{\hat{j}\hat{0}} + \Gamma^{\hat{i}}{}_{\hat{0}\hat{j}})v^{\hat{j}}. \tag{24.67}$$

By computing the connection coefficients from the metric coefficients of Eq. (24.60b) (Ex. 24.14), we bring this low-velocity geodesic law of motion into the form

geodesic equation for slowly moving particle in proper reference frame of accelerated, rotating observer

$$\frac{d^2x^{\hat{i}}}{(dx^{\hat{0}})^2} = -a^{\hat{i}} - 2\epsilon^{\hat{i}}{}_{\hat{j}\hat{k}}\Omega^{\hat{j}}v^{\hat{k}}, \quad \text{that is,} \quad \frac{d^2\mathbf{x}}{(dx^{\hat{0}})^2} = -\mathbf{a} - 2\boldsymbol{\Omega}\times\mathbf{v}. \tag{24.68}$$

This is the standard nonrelativistic form of the law of motion for a free particle as seen in a rotating, accelerating reference frame. The first term on the right-hand side is the inertial acceleration due to the failure of the frame to fall freely, and the second term is the Coriolis acceleration due to the frame's rotation. There would also be a centrifugal acceleration if we had kept terms of higher order in distance away from the observer's world line, but this acceleration has been lost due to our linearizing the metric (24.60b) in that distance.

This analysis shows how the elegant formalism of tensor analysis gives rise to familiar physics. In the next few chapters we will see it give rise to less familiar, general relativistic phenomena.

EXERCISES

Exercise 24.14 *Derivation: Proper Reference Frame*

(a) Show that the coordinate transformation (24.60a) brings the metric $ds^2 = \eta_{\alpha\beta}dx^\alpha dx^\beta$ into the form of Eq. (24.60b), accurate to linear order in separation $x^{\hat{j}}$ from the origin of coordinates.

(b) Compute the connection coefficients for the coordinate basis of Eq. (24.60b) at an arbitrary event on the observer's world line. Do so first by hand calculations, and then verify your results using symbolic-manipulation software on a computer.

(c) Using the connection coefficients from part (b), show that the rate of change of the basis vectors $\mathbf{e}_{\hat{\alpha}}$ along the observer's world line is given by Eq. (24.61).

(d) Using the connection coefficients from part (b), show that the low-velocity limit of the geodesic equation [Eq. (24.67)] is given by Eq. (24.68).

24.5.3

24.5.3 Uniformly Accelerated Observer

transformation between inertial coordinates and uniformly accelerated coordinates

As an important example (cf. Ex. 2.16), consider an observer whose accelerated world line, written in some inertial (Lorentz) coordinate system $\{t, x, y, z\}$, is

$$t = (1/\kappa)\sinh(\kappa\tau), \quad x = (1/\kappa)\cosh(\kappa\tau), \quad y = z = 0. \tag{24.69}$$

Here τ is proper time along the world line, and κ is the magnitude of the observer's 4-acceleration: $\kappa = |\vec{a}|$ (which is constant along the world line; see Ex. 24.15, where the reader can derive the various claims made in this subsection and the next).

The world line (24.69) is depicted in Fig. 24.7 as a thick, solid hyperbola that asymptotes to the past light cone at early times and to the future light cone at late times. The dots along the world line mark events that have proper times $\tau = -1.2, -0.9, -0.6, -0.3, 0.0, +0.3, +0.6, +0.9, +1.2$ (in units of $1/\kappa$). At each of these dots, the 3-plane orthogonal to the world line is represented by a dashed line (with the 2 dimensions out of the plane of the paper suppressed from the diagram). This 3-plane is labeled by its coordinate time $x^{\hat{0}}$, which is equal to the proper time of the dot. The basis vector $\vec{e}_{\hat{1}}$ is chosen to point along the observer's 4-acceleration, so $\vec{a} = \kappa\vec{e}_{\hat{1}}$. The coordinate $x^{\hat{1}}$ measures proper distance along the straight line that starts out tangent to $\vec{e}_{\hat{1}}$. The other two basis vectors $\vec{e}_{\hat{2}}$ and $\vec{e}_{\hat{3}}$ point out of the plane of the figure and are parallel transported along the world line: $\nabla_{\vec{U}}\vec{e}_{\hat{2}} = \nabla_{\vec{U}}\vec{e}_{\hat{3}} = 0$. In addition, $x^{\hat{2}}$ and $x^{\hat{3}}$ are measured along straight lines, in the orthogonal 3-plane, that start out tangent to these vectors. This construction implies that the resulting proper reference frame has vanishing rotation, $\vec{\Omega} = 0$ (Ex. 24.15), and that $x^{\hat{2}} = y$ and $x^{\hat{3}} = z$, where y and z are coordinates in the $\{t, x, y, z\}$ Lorentz frame that we used to define the world line [Eqs. (24.69)].

Usually, when constructing an observer's proper reference frame, one confines attention to the immediate vicinity of her world line. However, in this special case it is instructive to extend the construction (the orthogonal 3-planes and their resulting spacetime coordinates) outward arbitrarily far. By doing so, we discover that the 3-planes all cross at location $x^{\hat{1}} = -1/\kappa$, which means the coordinate system $\{x^{\hat{\alpha}}\}$ becomes singular there. This singularity shows up in a vanishing $g_{\hat{0}\hat{0}}(x^{\hat{1}} = -1/\kappa)$ for the spacetime metric, written in that coordinate system:

singularity of uniformly accelerated coordinates

spacetime metric in uniformly accelerated coordinates

$$ds^2 = -(1 + \kappa x^{\hat{1}})^2 (dx^{\hat{0}})^2 + (dx^{\hat{1}})^2 + (dx^{\hat{2}})^2 + (dx^{\hat{3}})^2. \tag{24.70}$$

[Note that for $|x^{\hat{1}}| \ll 1/\kappa$ this metric agrees with the general proper-reference-frame metric (24.60b).] From Fig. 24.7, it should be clear that this coordinate system can only cover smoothly one quadrant of Minkowski spacetime: the quadrant $x > |t|$.

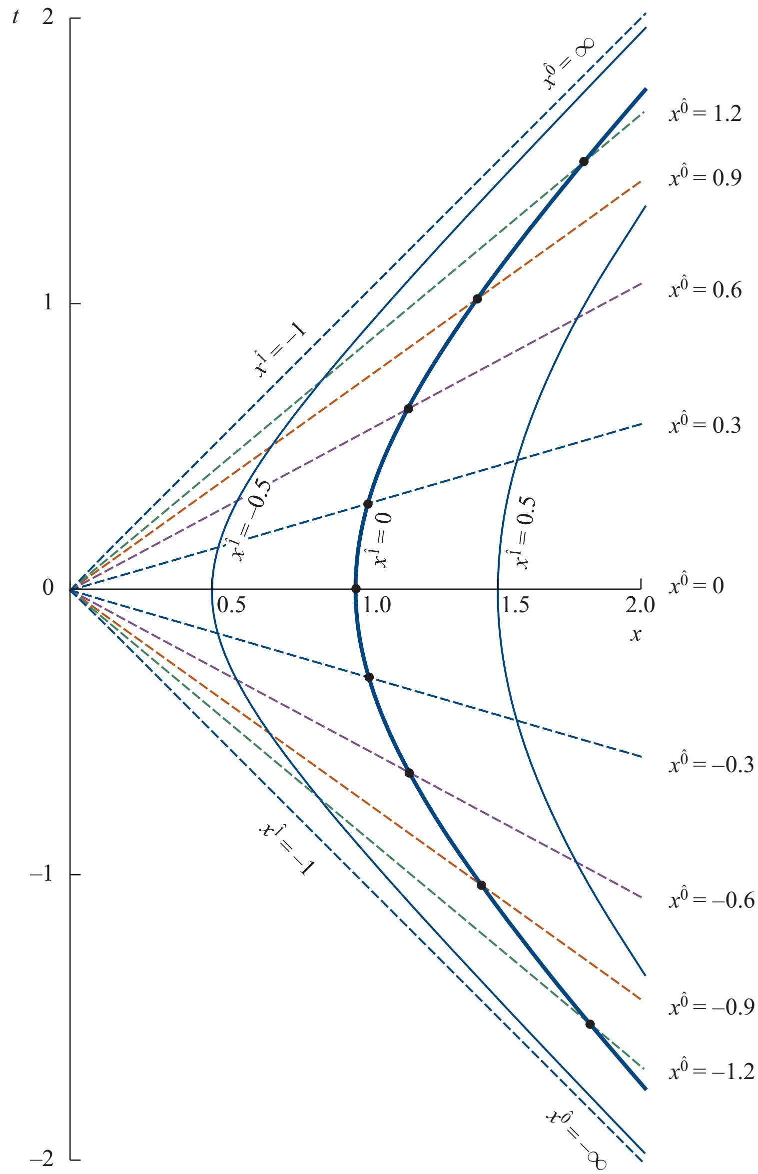

FIGURE 24.7 The proper reference frame of a uniformly accelerated observer. All lengths and times are measured in units of $1/\kappa$. We show only 2 dimensions of the reference frame—those in the 2-plane of the observer's curved world line.

24.5.4 Rindler Coordinates for Minkowski Spacetime

24.5.4

The spacetime metric (24.70) in our observer's proper reference frame resembles the metric in the vicinity of a black hole, as expressed in coordinates of observers who accelerate so as to avoid falling into the hole. In preparation for discussing this in

Rindler coordinates

Chap. 26, we shift the origin of our proper-reference-frame coordinates to the singular point and rename them. Specifically, we introduce so-called *Rindler coordinates*:

$$t' = x^{\hat{0}}, \quad x' = x^{\hat{1}} + 1/\kappa, \quad y' = x^{\hat{2}}, \quad z' = x^{\hat{3}}. \tag{24.71}$$

It turns out (Ex. 24.15) that these coordinates are related to the Lorentz coordinates that we began with, in Eqs. (24.69), by

$$t = x' \sinh(\kappa t'), \quad x = x' \cosh(\kappa t'), \quad y = y', \quad z = z'. \tag{24.72}$$

The metric in this Rindler coordinate system, of course, is the same as (24.70) with displacement of the origin:

spacetime metric in Rindler coordinates

$$ds^2 = -(\kappa x')^2 dt'^2 + dx'^2 + dy'^2 + dz'^2. \tag{24.73}$$

horizon of Rindler coordinates

The world lines of constant $\{x', y', z'\}$ have uniform acceleration: $\vec{a} = (1/x')\vec{e}_{x'}$. Thus we can think of these coordinates as the reference frame of a family of uniformly accelerated observers, each of whom accelerates away from their *horizon* $x' = 0$ with acceleration equal to 1/(her distance x' above the horizon). (We use the name "horizon" for $x' = 0$, because it represents the edge of the region of spacetime that these observers are able to observe.) The local 3-planes orthogonal to these observers' world lines all mesh to form global 3-planes of constant t'. This is a major factor in making the metric (24.73) so simple.

EXERCISES

Exercise 24.15 *Derivation: Uniformly Accelerated Observer and Rindler Coordinates*
In this exercise you will derive the various claims made in Secs. 24.5.3 and 24.5.4.

(a) Show that the parameter τ along the world line (24.69) is proper time and that the 4-acceleration has magnitude $|\vec{a}| = 1/\kappa$.

(b) Show that the unit vectors $\vec{e}_{\hat{j}}$ introduced in Sec. 24.5.3 all obey the Fermi-Walker transport law (24.62) and therefore, by virtue of Eq. (24.61b), the proper reference frame built from them has vanishing rotation rate: $\vec{\Omega} = 0$.

(c) Show that the coordinates $x^{\hat{2}}$ and $x^{\hat{3}}$ introduced in Sec. 24.5.3 are equal to the y and z coordinates of the inertial frame used to define the observer's world line [Eqs. (24.69)].

(d) Show that the proper-reference-frame coordinates constructed in Sec. 24.5.3 are related to the original $\{t, x, y, z\}$ coordinates by

$$t = (x^{\hat{1}} + 1/\kappa)\sinh(\kappa x^{\hat{0}}), \quad x = (x^{\hat{1}} + 1/\kappa)\cosh(\kappa x^{\hat{0}}), \quad y = x^{\hat{2}}, \quad z = x^{\hat{3}}; \tag{24.74}$$

and from this, deduce the form (24.70) of the Minkowski spacetime metric in the observer's proper reference frame.

(e) Show that, when converted to Rindler coordinates by moving the spatial origin, the coordinate transformation (24.74) becomes (24.72), and the metric (24.70) becomes (24.73).

(f) Show that observers at rest in the Rindler coordinate system (i.e., who move along world lines of constant $\{x', y', z'\}$) have 4-acceleration $\vec{a} = (1/x')\vec{e}_{x'}$.

Exercise 24.16 *Example: Gravitational Redshift*
Inside a laboratory on Earth's surface the effects of spacetime curvature are so small that current technology cannot measure them. Therefore, experiments performed in the laboratory can be analyzed using special relativity. (This fact is embodied in Einstein's equivalence principle; end of Sec. 25.2.)

(a) Explain why the spacetime metric in the proper reference frame of the laboratory's floor has the form

$$ds^2 = (1 + 2gz)(dx^{\hat{0}})^2 + dx^2 + dy^2 + dz^2, \tag{24.75}$$

plus terms due to the slow rotation of the laboratory walls, which we neglect in this exercise. Here g is the acceleration of gravity measured on the floor.

(b) An electromagnetic wave is emitted from the floor, where it is measured to have wavelength λ_o, and is received at the ceiling. Using the metric (24.75), show that, as measured in the proper reference frame of an observer on the ceiling, the received wave has wavelength $\lambda_r = \lambda_o(1 + gh)$, where h is the height of the ceiling above the floor (i.e., the light is *gravitationally redshifted* by $\Delta\lambda/\lambda_o = gh$). [Hint: Show that all crests of the wave must travel along world lines that have the same shape, $z = F(x^{\hat{0}} - x_e^{\hat{0}})$, where F is some function, and $x_e^{\hat{0}}$ is the coordinate time at which the crest is emitted from the floor. You can compute the shape function F if you wish, but it is not needed to derive the gravitational redshift; only its universality is needed.]

The first high-precision experiments to test this prediction were by Robert Pound and his student Glen Rebka and postdoc Joseph Snider, in a tower at Harvard University in the 1950s and 1960s. They achieved 1% accuracy. We discuss this gravitational redshift in Sec. 27.2.1.

Exercise 24.17 *Example: Rigidly Rotating Disk*
Consider a thin disk with radius R at $z = 0$ in a Lorentz reference frame. The disk rotates rigidly with angular velocity Ω. In the early years of special relativity there was much confusion over the geometry of the disk: In the inertial frame it has physical radius (proper distance from center to edge) R and physical circumference $\mathcal{C} = 2\pi R$. But Lorentz contraction dictates that, as measured on the disk, the circumference should be $\sqrt{1 - v^2}\,\mathcal{C}$ (with $v = \Omega R$), and the physical radius, R, should be unchanged. This seemed weird. How could an obviously flat disk in flat spacetime have a curved,

non-Euclidean geometry, with physical circumference divided by physical radius smaller than 2π? In this exercise you will explore this issue.

(a) Consider a family of observers who ride on the edge of the disk. Construct a circular curve, orthogonal to their world lines, that travels around the disk (at $\sqrt{x^2 + y^2} = R$). This curve can be thought of as lying in a 3-surface of constant time $x^{\hat{0}}$ of the observers' proper reference frames. Show that it spirals upward in a Lorentz-frame spacetime diagram, so it cannot close on itself after traveling around the disk. Thus the 3-planes, orthogonal to the observers' world lines at the edge of the disk, cannot mesh globally to form global 3-planes (by contrast with the case of the uniformly accelerated observers in Sec. 24.5.4 and Ex. 24.15).

(b) Next, consider a 2-dimensional family of observers who ride on the surface of the rotating disk. Show that at each radius $\sqrt{x^2 + y^2} =$ const, the constant-radius curve that is orthogonal to their world lines spirals upward in spacetime with a different slope. Show this means that even locally, the 3-planes orthogonal to each of their world lines cannot mesh to form larger 3-planes—thus there does not reside in spacetime any 3-surface orthogonal to these observers' world lines. There is no 3-surface that has the claimed non-Euclidean geometry.

Bibliographic Note

For a very readable presentation of most of this chapter's material, from much the same point of view, see Hartle (2003, Chap. 20). For an equally elementary introduction from a somewhat different viewpoint, see Schutz (2009, Chaps. 1–4). A far more detailed and somewhat more sophisticated introduction, largely but not entirely from our viewpoint, will be found in Misner, Thorne, and Wheeler (1973, Chaps. 1–6). More sophisticated treatments from rather different viewpoints than ours are given in Wald (1984, Chaps. 1, 2, and Sec. 3.1), and Carroll (2004, Chaps. 1, 2). A treasure trove of exercises on this material, with solutions, is in Lightman et al. (1975, Chaps. 6–8). See also the bibliography for Chap. 2.

For a detailed and sophisticated discussion of accelerated observers and the measurements they make, see Gourgoulhon (2013).

NAME INDEX

Page numbers for entries in boxes are followed by "b," those for epigraphs at the beginning of a chapter by "e," those for figures by "f," and those for notes by "n."

SUBJECT INDEX

Second and third level entries are not ordered alphabetically. Instead, the most important or general entries come first, followed by less important or less general ones, with specific applications last.

Page numbers for entries in boxes are followed by "b," those for epigraphs at the beginning of a chapter by "e," those for figures by "f," for notes by "n," and for tables by "t."