Chapter 23

From Special to General Relativity

Version 0223.2.pdf, 29 April 2003.

Please send comments, suggestions, and errata via email to kip@tapir.caltech.edu and also to rdb@caltech.edu, or on paper to Kip Thorne, 130-33 Caltech, Pasadena CA 91125

23.1 Overview

We have reached the final Part of this book, in which we present an introduction to the basic concepts of general relativity and its most important applications. This subject, although a little more challenging than the material that we have covered so far, is nowhere near as formidable as its reputation. Indeed, if you have mastered the techniques developed in the first five Parts, the path to the Einstein Field Equations should be short and direct.

The General Theory of Relativity is the crowning achievement of classical physics, the last great fundamental theory created prior to the discovery of quantum mechanics; its formulation by Albert Einstein in 1915 marks the culmination of the great intellectual adventure undertaken by Newton 250 years earlier. It was created after many wrong turns and with little experimental guidance, almost by pure thought. Unlike the special theory, whose physical foundations and logical consequences were clearly appreciated by physicists soon after Einstein's 1905 formulation, the unique and distinctive character of the general theory only came to be widely appreciated long after its creation. Ultimately, in hindsight, rival classical theories of gravitation came to seem unnatural, inelegant and arbitrary by comparison.[1] Experimental tests of Einstein's theory also were slow to come; only in the last three decades have there been striking tests of high enough precision to convince most empiricists that, in all probability, and in its domain of applicability, general relativity is essentially correct. Despite this, it is still very poorly tested compared with, for example, quantum electrodynamics.

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. 23.2, we give a brief encapsulation of the special theory drawn largely from Chap. 1, emphasizing those aspects that we must generalize to deal with non-inertial frames of reference. Then in Sec. 23.3 we

[1]For a readable account, see Will 1987 Was Einstein Right.

collect, review and extend the fundamental ideas of differential geometry that have been scattered throughout the book and which we shall need as foundations for the mathematics of *spacetime curvature* (Chap. 24); most importantly, we generalize differential geometry to encompass coordinate systems and bases that are not orthogonal. Einstein's field equations are a relationship between the curvature of spacetime and the matter that generates it, akin to the Maxwell equations' relationship between the electromagnetic field and electric currents and charges. The matter is described using the *stress-energy tensor* that we introduced in Sec. 1.12. We revisit the stress-energy tensor in Sec. 23.4 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—e.g., an observer who hovers in a spaceship just above the horizon of a black hole, or a gravitational-wave experimenter in an earth-bound laboratory. As a foundation for treating such observers, in Sec. 23.5 we examine measurements made by accelerated observers in the flat spacetime of special relativity.

This chapter will leave us well prepared to develop, in Chap. 24, the basic concepts of general relativity, including spacetime curvature, the Einstein Field Equation, and the laws of physics in curved spacetime. In Chaps. 25–27 we shall explore the major applications of general relativity: to stars, black holes, gravitational waves, and cosmology. We begin in Chap 25 by studying the the spacetime curvature around and inside highly compact stars (such as neutron stars) and showing how, in the weak field limit, non-Newtonian effects are predicted in our own solar system and in binary neutron star systems and how these predictions have been verified. We also discuss the implosion of massive stars and describe the circumstances under which the implosion inevitably produces a black hole, and we explore the surprising and, initially, counter-intuitive properties of black holes. In Chap. 26 we study gravitational waves, i.e. ripples in the curvature of spacetime that propagate with the speed of light, and we explore their close analogy with the electromagnetic waves that were first predicted by Maxwell's equations. We explore the properties of these waves, their production by binary stars and merging black holes, projects to detect them, both on earth and in space, and the prospects for using them to explore observationally the dark side of the universe and the nature of ultrastrong spacetime curvature. Finally, in Chap. 26 we draw once more upon all the previous Parts of this book, combining them with general relativity to describe the universe on the largest of scales and longest of times: cosmology. It is here, more than anywhere else in classical physics, that we are conscious of reaching a frontier where the still-promised land of quantum gravity beckons.

23.2 Special Relativity Once Again

A pre-requisite to learning the theory of general relativity is to understand special relativity in geometric language. In Chap. 1, we discussed the foundations of special relativity with this in mind and it is now time to remind ourselves of what we learned.

23.2.1 Geometric, Frame-Independent Formulation

In Chap. 1 we learned that *every law of physics must be expressible as a geometric, frame-independent relationship between geometric, frame-independent objects.* This is equally true

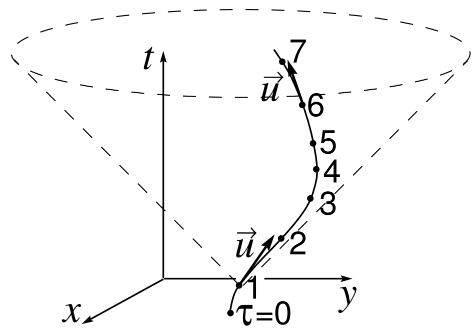

Fig. 23.1: The world line $\mathcal{P}(\tau)$ of a particle in Minkowski spacetime and the tangent vector $\vec{u} = d\mathcal{P}/d\tau$ to this world line; $\vec{u}$ is the particle's 4-velocity. The bending of the world line is produced by some force that acts on the particle, e.g. by the Lorentz force embodied in Eq. (23.3). 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 or its tangent vector $\vec{u}$ or the light cone, or for the formulation of the Lorentz force law.

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. 24) it is 4-dimensional curved spacetime; see Fig. 1.1 and associated discussion.

In special relativity, the demand that the laws be geometric relationships between geometric objects in Minkowski spacetime is called the *Principle of Relativity*; see Sec. 1.2.

Examples of the geometric objects are: (i) a point $\mathcal{P}$ in spacetime (which represents an *event*); (ii) a parametrized 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[2] that the particle carries (Fig. 23.1); (iii) 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); and (iv) tensors such as the electromagnetic field tensor $\mathbf{F}(\_,\_)$. A tensor, as we recall, is a linear real-valued function of vectors; when one puts vectors $\vec{A}$ and $\vec{B}$ into the slots of $\mathbf{F}$, one obtains a real number (a scalar) $\mathbf{F}(\vec{A},\vec{B})$ that is linear in $\vec{A}$ and in $\vec{B}$ so for example $\mathbf{F}(\vec{A}, b\vec{B} + c\vec{C}) = b\mathbf{F}(\vec{A},\vec{B}) + c\mathbf{F}(\vec{A},\vec{C})$. When one puts a vector $\vec{B}$ into just one of the slots of $\mathbf{F}$ and leaves the other empty, one obtains a tensor with one empty slot, $\mathbf{F}(\_,\vec{B})$, i.e. a vector. The result of putting a vector into the slot of a vector is the scalar product, $\vec{D}(\vec{B}) = \vec{D}\cdot\vec{B} = \mathbf{g}(\vec{D},\vec{B})$, where $\mathbf{g}(\_,\_)$ is the metric.

In Secs. 1.2 and 1.3 we tied our definitions of the inner product and the 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

$$\mathbf{g}(\Delta\vec{x},\Delta\vec{x}) \equiv \Delta\vec{x}\cdot\Delta\vec{x} \equiv -(\Delta\tau)^2 \ . \tag{23.1}$$

[2]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 so two ideal clocks momentarily at rest with respect to each other tick at the same rate independent of their relative acceleration; cf. Secs. 1.2 and 1.4, and for greater detail, pp. 23–29 and 395–299 of MTW.

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

$$\vec{u} = \frac{d\mathcal{P}}{d\tau} = \lim_{\Delta\tau\to 0} \frac{\mathcal{P}(\tau+\Delta\tau) - \mathcal{P}(\tau)}{\Delta\tau} \equiv \lim_{\Delta\tau\to 0} \frac{\Delta\vec{x}}{\Delta\tau} \ , \tag{23.2}$$

Eq. (23.1) implies that $\vec{u}\cdot\vec{u} = \mathbf{g}(\vec{u},\vec{u}) = -1$.

The 4-velocity $\vec{u}$ is an example of a *timelike* vector; 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. 23.1). Vectors $\vec{k}$ on the light cone (the tangents to the world lines of the photons) are *null* and so have vanishing squared lengths, $\vec{k}\cdot\vec{k} = \mathbf{g}(\vec{k},\vec{k}) = 0$; and vectors $\vec{A}$ that lie outside the light cone are *spacelike* and have positive squared lengths, $\vec{A}\cdot\vec{A} > 0$.

An example of a physical law in 4-dimensional geometric language is the Lorentz force law

$$\frac{d\vec{p}}{d\tau} = q\mathbf{F}(\_\,,\vec{u}) \ , \tag{23.3}$$

where q is the particle's charge and both sides of this equation are vectors, i.e. first-rank tensors, i.e. tensors with just one slot. As we learned in Sec. 1.5, it is convenient to give names to slots. When we do so, we can rewrite the Lorentz force law as

$$\frac{dp^\alpha}{d\tau} = qF^{\alpha\beta}u_\beta \ . \tag{23.4}$$

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 $\mathbf{F}$, β is the name of the slot of $\mathbf{u}$, and the double use of β with one up and one down on the right side of the equation represents the insertion of $\vec{u}$ into the β slot of $\mathbf{F}$, 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 Chap. 1.

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

In special relativity a key role is played by *inertial reference frames.* 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 each other 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 lower case Latin indices $x^j(\mathcal{P})$ with j running over 1,2,3]. The latticework's clocks are ideal and are synchronized with each other via the Einstein light-pulse process (Sec. 1.2). They are used to identify the temporal coordinate $x^0 = t$ of an event $\mathcal{P}$; i.e. $x^0(\mathcal{P})$ 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 lower case 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*.

In the real universe, spacetime curvature is very 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 non-accelerating and non-rotating with respect to cosmologically distant galaxies (and thence with respect to a local frame in which the cosmic microwave radiation looks isotropic) constitute good approximations to inertial reference frames.

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 each other) and have unit length (making them orthonormal). This orthonormality is embodied in the inner products $\vec{e}_\alpha \cdot \vec{e}_\beta = \eta_{\alpha\beta}$, where by definition

$$\eta_{00} = -1 \; , \quad \eta_{11} = \eta_{22} = \eta_{33} = +1 \; , \quad \eta_{\alpha\beta} = 0 \text{ if } \alpha \neq \beta \; . \tag{23.5}$$

Here and throughout Part VI (as in Chap. 1), we set the speed of light to unity [i.e. we use the *geometrized units* discussed in Eqs. (1.4) and (1.5)], 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 s $= 2.9979245 \times 10^8$ m.

In 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 were computable by inserting the basis vectors into the tensor's slots, $u_\alpha = \vec{u}(\vec{e}_\alpha) \equiv \vec{u} \cdot \vec{e}_\alpha$; $F_{\alpha\beta} = \mathbf{F}(\vec{e}_\alpha, \vec{e}_\beta)$. For example, in our Lorentz basis the covariant components of the metric are $g_{\alpha\beta} = \mathbf{g}(\vec{e}_\alpha, \vec{e}_\beta) = \vec{e}_\alpha \cdot \vec{e}_\beta = \eta_{\alpha\beta}$. 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 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. 1.5, tensors can be expressed in terms of their contravariant components as $\vec{p} = p^\alpha \vec{e}_\alpha$, and $\mathbf{F} = F^{\alpha\beta} \vec{e}_\alpha \otimes \vec{e}_\beta$, where $\otimes$ represents the tensor product [Eq. (1.18)].

We also learned in Chap. 1 that any frame independent geometric relation between tensors can be rewritten as a relation between 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. For example, the component version of the Lorentz force law says $dp^\alpha/d\tau = q F^{\alpha\beta} u_\beta$, which is identical to Eq. (23.4). The only difference is the interpretation of the symbols. In the component equation $F^{\alpha\beta}$ are the components of $\mathbf{F}$ 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 $\mathbf{F}(\_\,,\_)$ 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 $\mathbf{F}$ to produce a single-slotted tensor, i.e. a vector whose slot is named α.

As we saw in Sec. 1.6, a particle's 4-velocity $\vec{u}$ (defined originally without the aid of any reference frame; Fig. 23.1) 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 $\gamma \equiv 1/\sqrt{1 - \delta_{ij} v^i v^j}$. Similarly, the particle's energy is $E \equiv p^0$ 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 $E = p^0$; and similarly a split of the electromagnetic field tensor $\mathbf{F}$ into the 3-dimensional electric field $\mathbf{E}$ and 3-dimensional magnetic field $\mathbf{B}$; cf. Secs. 1.6] and 1.10.

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.* The components of tensors in one Lorentz frame are related to those in another by a Lorentz transformation (Sec. 1.7), 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 onto 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 between geometric objects that reside in Minkowski spacetime.*

If the above discussion is not completely clear, the reader should study the relevant portions of Chap. 1.

23.2.3 Light Speed, the Interval, and Spacetime Diagrams

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 for the speed of light c, momentarily abandoning geometrized units and returning to mks/SI units. According to Maxwell, c can be determined by performing non-radiative laboratory experiments; it is not necessary to measure the time it takes light to travel along some path. For example, measure the electrostatic force between two charges; that force is $\propto \epsilon_0^{-1}$, the electric permitivity of free space. Then allow one of these charges to travel down a wire and by its motion generate a magnetic field. Let the other charge move through this field and measure the magnetic force on it; that force is $\propto \mu_0$, the magnetic permitivity of free space. The ratio of these two forces can be computed and is $\propto 1/\mu_0\epsilon_0 = c^2$. By combining the results of the two experiments, we therefore can deduce the speed of light c; this is completely analogous to deducing the speed of seismic waves through rock from a knowledge of the rock's density and elastic moduli, using elasticity theory (Chap. 11). The principle of relativity, in operational form, dictates that the results of the electric and magnetic experiments must be independent of the Lorentz frame in which one chooses to perform them; therefore, the speed of light is frame-independent—as we argued by a different route in Sec. 1.2. It is this frame independence that enables us to introduce geometrized units with $c = 1$.

Another example of frame independence (Lorentz invariance) is provided by the *interval between two neighboring events*. 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

$$\begin{aligned}(\Delta s)^2 \equiv \Delta\vec{x}\cdot\Delta\vec{x} = \mathbf{g}(\Delta\vec{x},\Delta\vec{x}) &= g_{\alpha\beta}\Delta x^\alpha \Delta x^\beta \\ &= -(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \ . \end{aligned} \tag{23.6}$$

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. 1.2.

Spacetime diagrams will 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. 1.10 and Ex. 1.11.

****

EXERCISES

Exercise 23.1 *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 $\mathbf{g}$. Your friend claims that the metric is some other frame-independent tensor $\tilde{\mathbf{g}}$ that differs from $\mathbf{g}$. Suppose that your correct metric $\mathbf{g}$ and his wrong one $\tilde{\mathbf{g}}$ 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, i.e. $\mathbf{g}(\Delta\vec{x},\Delta\vec{x}) \neq \tilde{\mathbf{g}}(\Delta\vec{x},\Delta\vec{x})$. Prove that $\tilde{\mathbf{g}}$ and $\mathbf{g}$ differ solely by a scalar multiplicative factor. [*Hint*: pick some Lorentz frame and perform computations there, then lift yourself back up to a frame-independent viewpoint.]

Exercise 23.2 *Problem: Causality*
If two events occur at the same point though 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.

****

23.3 Differential Geometry in General Bases and in Curved Manifolds

The tensor-analysis formalism reviewed in the last section is inadequate for general relativity in several ways:

First, in general relativity we shall need to use bases $\vec{e}_\alpha$ that are not orthornormal, 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 shall generalize our formalism to treat such non-orthonormal bases.

Second, in the curved spacetime of general relativity (and in any other curved manifold, e.g. the two-dimensional surface of the earth) the definition of a vector as an arrow connecting two points 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 shall refine the concept of a vector to deal with this problem, and in the process we shall find ourselves introducing 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 manifold.

Third, once we have been forced to think of a tensor as residing in a specific tangent space at a specific point in the manifold, there arises the question of how one can 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 will have to rework the notion of a gradient and the gradient's connection coefficients; and since, in doing an integral, one must add contributions that live at different points in the manifold, we must also rework the notion of integration.

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

23.3.1 Non-Orthonormal Bases

Consider an n-dimensional manifold, e.g. 4-dimensional spacetime or 3-dimensional Euclidean space or the 2-dimensional surface of a sphere. At some point $\mathcal{P}$ in the manifold, 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. 23.2 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, $F_{\alpha\beta} = \mathbf{F}(\vec{e}_\alpha, \vec{e}_\beta)$, and that the tensor itself can be reconstructed from its contravariant components as $\mathbf{F} = F^{\mu\nu}\vec{e}_\mu \otimes \vec{e}_\nu$, and that the two sets of components are computable from

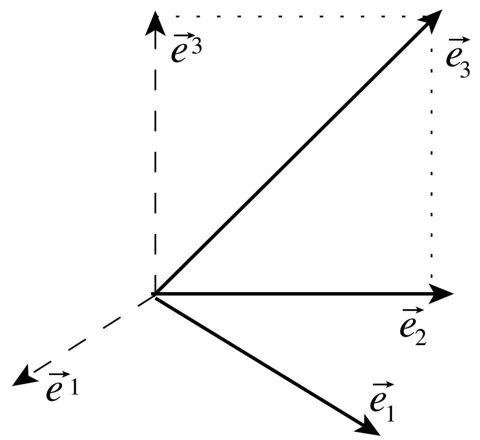

Fig. 23.2: Non-orthonormal basis vectors $\vec{e}_j$ in Euclidean 3-space and two members $\vec{e}^1$ and $\vec{e}^3$ of the dual basis. The vectors $\vec{e}_1$ and $\vec{e}_2$ lie in the horizontal plane, so $\vec{e}^3$ is orthogonal to that plane, i.e. it points vertically upward, and its inner product 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.

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; i.e. 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 determinate).

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

$$\vec{e}^\mu \cdot \vec{e}_\beta \equiv \mathbf{g}(\vec{e}^\mu, \vec{e}_\beta) = \delta^\mu{}_\beta \tag{23.7}$$

where $\delta^\alpha{}_\beta$ 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 (23.7) 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. (23.7) thereby determines the value of $\vec{e}^\mu$ on every other vector.

The duality relation (23.7) says that $\vec{e}^1$ is always perpendicular to all the $\vec{e}_\alpha$ 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. 23.2. In Minkowski spacetime, if $\vec{e}_\alpha$ 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 (23.7) 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

$$F^{\mu\nu} = \mathbf{F}(\vec{e}^\mu, \vec{e}^\nu)\ , \quad F_{\alpha\beta} = \mathbf{F}(\vec{e}_\alpha, \vec{e}_\beta)\ , \quad F^\mu{}_\beta = \mathbf{F}(\vec{e}^\mu, \vec{e}_\beta)\ ; \tag{23.8}$$

Box 23.1
Dual Bases in Other Contexts

Vector spaces appear in a wide variety of contexts in mathematics and physics, and wherever they appear it can be useful to introduce dual bases.

When a vector space does not posses a metric, the basis $\{\vec{e}^\mu\}$ lives in a different space from $\{\vec{e}_\alpha\}$, and the two spaces are said to be dual to each other. An important example occurs in manifolds that do not have metrics. There the vectors in the space spanned by $\{\vec{e}^\mu\}$ are often called a *one forms* and are represented pictorially as families of parallel surfaces; the vectors in the space spanned by $\{\vec{e}_\alpha\}$ are called *tangent vectors* and are represented pictorially as arrows; the one forms are linear functions of tangent vectors, and the result that a one form $\tilde{\beta}$ gives when a tangent vector $\vec{a}$, is inserted into its slot, $\tilde{\beta}(\vec{a})$, is the number of surfaces of $\tilde{\beta}$ pierced by the arrow $\vec{a}$; see, e.g., MTW. A metric produces a one-to-one mapping between the one forms and the tangent vectors. In this book we regard this mapping as equating each one form to a tangent vector and thereby as making the space of one forms and the space of tangent vectors be identical. This permits us to avoid ever speaking about one forms, except here in this box.

Quantum mechanics provides another example of dual spaces. The kets $|\psi\rangle$ are the tangent vectors and the bras $\langle\phi|$ are the one-forms: linear *complex valued* functions of kets with the value that $\langle\phi|$ gives when $|\psi\rangle$ is inserted into its slot being the inner product $\langle\phi|\psi\rangle$.

see Ex. 23.4. Among the consequences of this duality are the following: (i)

$$g^{\mu\beta}g_{\nu\beta} = \delta^\mu_\nu \ , \tag{23.9}$$

i.e., the matrix of contravariant components of the metric is inverse to that of the covariant components, $||g^{\mu\nu}|| = ||g_{\alpha\beta}||^{-1}$; this relation guarantees that when one raises indices on a tensor $F_{\alpha\beta}$ with $g^{\mu\alpha}$ and then lowers them back down with $g_{\nu\beta}$, one recovers one's original covariant components $F_{\alpha\beta}$ unaltered. (ii)

$$\mathsf{F} = F^{\mu\nu}\vec{e}_\mu \otimes \vec{e}_\nu = F_{\alpha\beta}\vec{e}^\alpha \otimes \vec{e}^\beta = F^\mu{}_\beta \vec{e}_\mu \otimes \vec{e}^\beta \ , \tag{23.10}$$

i.e., one can reconstruct a tensor from its components by lining up the indices in a manner that accords with the rules of index manipulation. (iii)

$$\mathsf{F}(\vec{p}, \vec{q}) = F^{\alpha\beta}p_\alpha p_\beta \ , \tag{23.11}$$

i.e., the component versions of tensorial equations are identical in mathematical symbology to the slot-naming-index-notation versions.

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

$$\vec{e}_\alpha \equiv \frac{\partial \mathcal{P}}{\partial x^\alpha} \ . \tag{23.12}$$

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).

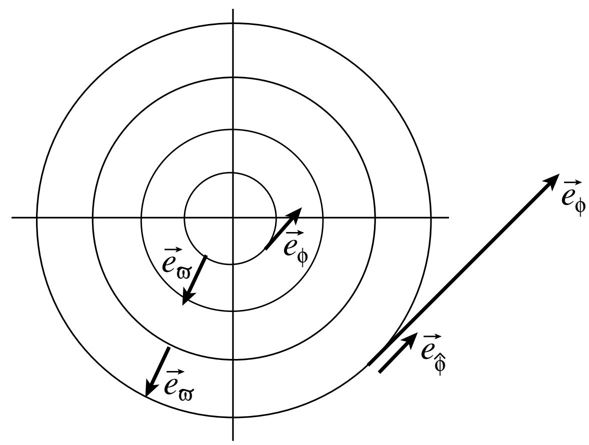

Fig. 23.3: A circular coordinate system $\{\varpi, \phi\}$ and its coordinate basis vectors $\vec{e}_\varpi = \partial\mathcal{P}/\partial\varpi$, $\vec{e}_\phi = \partial\mathcal{P}/\partial\phi$ at several locations in the coordinate system. Also shown is the orthonormal basis vector $\vec{e}_{\hat\phi}$.

In an orthogonal curvilinear coordinate system, e.g. circular polar coordinates (ϖ, ϕ) in Euclidean 2-space, 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 vector at small radii [cf. Fig. 23.3]; the corresponding unit-length vector is $\vec{e}_{\hat\phi} = (1/\varpi)\vec{e}_\phi$. 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 $g_{\phi\phi} = \varpi^2$, $g_{\varpi\varpi} = 0$, $g_{\varpi\phi} = g_{\phi\varpi} = 0$ which is in accord with the equation for the squared distance (interval) between adjacent points $ds^2 = g_{ij}dx^i dx^j = d\varpi^2 + \varpi^2 d\phi^2$. The metric components in the orthonormal basis, of course, are $g_{\hat i\hat j} = \delta_{ij}$.

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

In general, 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 $x^\alpha(\mathcal{P})$:

$$\vec{e}^\mu = \vec{\nabla} x^\mu \; . \tag{23.13}$$

It is straightforward to verify the duality relation (23.7) for these two bases:

$$\vec{e}^\mu \cdot \vec{e}_\alpha = \vec{e}_\alpha \cdot \vec{\nabla} x^\mu = \nabla_{\vec{e}_\alpha} x^\mu = \nabla_{\partial\mathcal{P}/\partial x^\alpha} x^\mu = \frac{\partial x^\mu}{\partial x^\alpha} = \delta^\mu_\alpha \; . \tag{23.14}$$

In any coordinate system, the expansion of the metric in terms of the dual basis, $\mathbf{g} = g_{\alpha\beta}\vec{e}^\alpha \otimes \vec{e}^\beta = g_{\alpha\beta}\vec{\nabla} x^\alpha \otimes \vec{\nabla} x^\beta$ 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

$$ds^2 = g_{\alpha\beta}\boldsymbol{\nabla} x^\alpha \otimes \boldsymbol{\nabla} x^\beta(d\vec{x}, d\vec{x}) = g_{\alpha\beta}(d\vec{x}\cdot\boldsymbol{\nabla} x^\alpha)(d\vec{x}\cdot\boldsymbol{\nabla} x^\beta) = g_{\alpha\beta}dx^\alpha dx^\beta \; . \tag{23.15}$$

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, $d\vec{x}\cdot\boldsymbol{\nabla}\psi$ 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:

$$\vec{e}_\alpha = \vec{e}_{\bar{\mu}} L^{\bar{\mu}}{}_\alpha \; , \quad \vec{e}_{\bar{\mu}} = \vec{e}_\alpha L^\alpha{}_{\bar{\mu}} \; . \tag{23.16}$$

(Note: 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

$$L^{\bar{\mu}}{}_\alpha L^\alpha{}_{\bar{\nu}} = \delta^{\bar{\mu}}{}_{\bar{\nu}} \; , \quad L^\alpha{}_{\bar{\mu}} L^{\bar{\mu}}{}_\beta = \delta^\alpha{}_\beta \; . \tag{23.17}$$

These $||L^{\bar{\mu}}{}_\alpha||$ are the generalizations of Lorentz transformations to arbitrary bases; cf. Eqs. (1.77), (1.78). As in the Lorentz-transformation case, the transformation laws (23.16) for the basis vectors imply corresponding transformation laws for components of vectors and tensors—laws that entail lining up indices in the obvious manner; e.g.

$$A_{\bar{\mu}} = L^\alpha{}_{\bar{\mu}} A_\alpha \; , \quad T^{\bar{\mu}\bar{\nu}}{}_{\bar{\rho}} = L^{\bar{\mu}}{}_\alpha L^{\bar{\nu}}{}_\beta L^\gamma{}_{\bar{\rho}} T^{\alpha\beta}{}_\gamma \; , \quad \text{and similarly in the opposite direction.} \tag{23.18}$$

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

$$L^{\bar{\mu}}{}_\alpha = \frac{\partial x^{\bar{\mu}}}{\partial x^\alpha} \; , \quad L^\alpha{}_{\bar{\mu}} = \frac{\partial x^\alpha}{\partial x^{\bar{\mu}}} \; , \tag{23.19}$$

as one can easily deduce via

$$\vec{e}_\alpha = \frac{\partial \mathcal{P}}{\partial x^\alpha} = \frac{\partial x^\mu}{\partial x^\alpha} \frac{\partial \mathcal{P}}{\partial x^\mu} = \vec{e}_\mu \frac{\partial x^\mu}{\partial x^\alpha} \; . \tag{23.20}$$

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

$$F_{\alpha\beta} = F_{\mu\nu} \frac{\partial x^\mu}{\partial x^\alpha} \frac{\partial x^\nu}{\partial x^\beta} \; . \tag{23.21}$$

This definition is in accord with Eqs. (23.18) and (23.19), though it hides the true and very simple nature of a tensor as a linear function of frame-independent vectors.

23.3.2 Vectors as Differential Operators; Tangent Space; Commutators

As was discussed above, 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

$$\vec{A} \equiv \frac{d\mathcal{P}}{d\zeta} \equiv \lim_{\Delta\zeta \to 0} \frac{\mathcal{P}(\Delta\zeta) - \mathcal{P}(0)}{\Delta\zeta} \; ; \tag{23.22}$$

cf. Eq. (23.2). 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; and in such an arbitrarily small region of the manifold, the

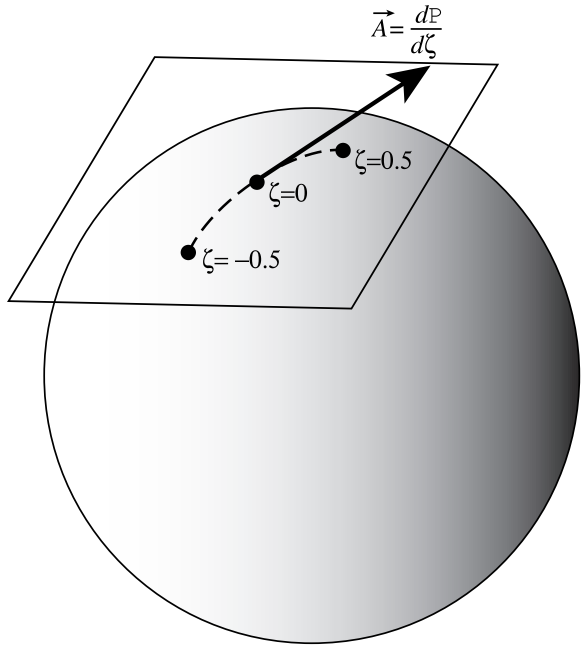

Fig. 23.4: A curve $\mathcal{P}(\zeta)$ on the surface of a sphere and the curve's tangent vector $\vec{A} = d\mathcal{P}/d\zeta$ at $\mathcal{P}(\zeta = 0) \equiv \mathcal{P}_o$. The tangent vector lives in the tangent space at $\mathcal{P}_o$, i.e. in the flat plane that is tangent to the sphere there as seen in the flat Euclidean 3-space in which the sphere's surface is embedded.

effects of the manifold's curvature become arbitrarily small and 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. 23.4). 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 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.

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 Fig. 23.4). 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 $\vec{A} \cdot \vec{B} = \mathbf{g}(\vec{A}, \vec{B})$ using the metric $\mathbf{g}$ that also resides in 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 $d/d\zeta$ which differentiates scalar fields along the curve.* This derivative operator is very 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 they define this vector space to be the tangent space at $\mathcal{P}_o$.

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 $\partial_{\vec{A}}$. There therefore is a one-to-one correspondence between the directional derivatives $\partial_{\vec{A}}$ at $\mathcal{P}_o$ and the vectors $\vec{A}$ there, and a complete isomorphism between the tangent-space manipulations that a mathematician will perform treating the directional derivatives as vectors, and those that a physicist will perform 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. With these motivations in mind, and because one always gains conceptual and computational power by having multiple viewpoints at one's finger tips (see, e.g., Feynman, 1966), we shall regard vectors henceforth *both* as arrows living in a tangent space and as directional derivatives. Correspondingly, we shall assert the equalities

$$\frac{\partial\mathcal{P}}{\partial x^\alpha} = \frac{\partial}{\partial x^\alpha} \quad , \quad \vec{A} = \partial_{\vec{A}} \; , \tag{23.23}$$

and shall often expand vectors in a coordinate basis using the notation

$$\vec{A} = A^\alpha \frac{\partial}{\partial x^\alpha} \; . \tag{23.24}$$

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 which, when viewed 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

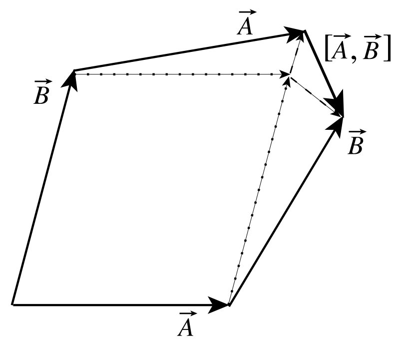

Fig. 23.5: The commutator $[\vec{A}, \vec{B}]$ of two vector fields. In this diagram the vectors are assumed to be so small that the curvature of the manifold is negligible in the region of the diagram, so all the vectors can be drawn lying in the surface itself rather than in their respective tangent spaces. In evaluating the two terms in the commutator (23.25), 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 short dashed curve 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 other short dashed curve). According to Eq. (23.25), the difference of these two short-dashed curves 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.

components of the commutator in a coordinate basis:

$$[\vec{A}, \vec{B}] \equiv \left[A^\alpha \frac{\partial}{\partial x^\alpha}, B^\beta \frac{\partial}{\partial x^\beta} \right] = \left(A^\alpha \frac{\partial B^\beta}{\partial x^\alpha} - B^\alpha \frac{\partial A^\beta}{\partial x^\alpha} \right) \frac{\partial}{\partial x^\beta} \tag{23.25}$$

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 23.5 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 $\vec{B}$.

The commutator is useful as a tool for distinguishing between coordinate bases and non-coordinate bases (also called non-holonomic 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. (23.5 explains), 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 non-coordinate basis, at least one of the commutators $[\vec{e}_\alpha, \vec{e}_\beta]$ will be nonzero.

23.3.3 Differentiation of Vectors and Tensors; Connection Coefficients

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. 1.9). Converting to the above notation, we began by defining the directional derivative of a tensor field $\mathbf{F}(\mathcal{P})$ along the tangent vector $\vec{A} = d/d\zeta$ to a curve $\mathcal{P}(\zeta)$:

$$\nabla_{\vec{A}}\mathbf{F} \equiv \lim_{\Delta\zeta\to 0} \frac{\mathbf{F}[\mathcal{P}(\Delta\zeta)] - \mathbf{F}[\mathcal{P}(0)]}{\Delta\zeta} \ . \tag{23.26}$$

This definition is problematic because $\mathbf{F}[\mathcal{P}(\Delta\zeta))]$ lives in a different tangent space than $\mathbf{F}[\mathcal{P}(0)]$. 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 $\mathbf{F}$ from one tangent space to the other.

In flat space or flat spacetime, and when $\mathbf{F}$ 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 $\mathbf{F}$ the parallel transport law is the same: keep its components fixed in an orthonormal coordinate basis.

In curved spacetime there is no such thing as an orthonormal coordinate basis. Just as the curvature of the earth's surface prevents one from placing a Cartesian coordinate system on it, so the spacetime curvature prevents one from introducing Lorentz coordinates; see Chap. 24. However, in an arbitrarily small region on the 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 $\mathcal{O}(L^2/R^2)$, where L is the size of the region and R is the earth's radius (see Sec.24.3). Correspondingly, 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. Such coordinates are sufficiently like their flat space counterparts that they can be used to define parallel transport in the curved manifolds: In Eq. (23.26) one must transport $\mathbf{F}$ 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 $\nabla_{\vec{A}}\mathbf{F}$ of $\mathbf{F}$.

Having made the directional derivative meaningful, one can proceed as in Sec. 1.9, and define the gradient of $\mathbf{F}$ by $\nabla_{\vec{A}}\mathbf{F} = \vec{\nabla}\mathbf{F}(\_,\_,\vec{A})$ [i.e., put $\vec{A}$ in the last, differentiation, slot of $\vec{\nabla}\mathbf{F}$; Eq. (1.89)].

As in Chap. 1, in any basis we denote the components of $\vec{\nabla}\mathbf{F}$ by $F_{\alpha\beta;\gamma}$; and as in Sec. 10.3 (elasticity theory), we can compute these components in any basis with the aid of that basis's *connection coefficients* (also called *Christoffel symbols*).

In Sec. 10.3 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, with non-othonormal bases and in spacetime, we must distinguish covariant and contravariant indices. Accordingly, by analogy with Eq. (10.18), we define the connection coefficients $\Gamma^\mu{}_{\alpha\beta}$ as

$$\nabla_\beta \vec{e}_\alpha = \Gamma^\mu{}_{\alpha\beta}\vec{e}_\mu \ , \tag{23.27}$$

where $\nabla_\beta \equiv \nabla_{\vec{e}_\beta}$. The duality between bases $\vec{e}^\nu \cdot \vec{e}_\alpha = \delta^\nu_\alpha$ then implies

$$\nabla_\beta \vec{e}^\mu = -\Gamma^\mu{}_{\alpha\beta}\vec{e}^\alpha \ . \tag{23.28}$$

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. Duality also implies that Eqs. (23.27) and (23.28) can be rewritten as

$$\Gamma^\mu{}_{\alpha\beta} = \vec{e}^\mu \cdot \nabla_\beta \vec{e}_\alpha = -\vec{e}_\alpha \nabla_\beta \vec{e}^\mu \ . \tag{23.29}$$

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

$$\begin{aligned} A^\mu{}_{;\beta}\vec{e}_\mu &= \nabla_\beta \vec{A} = \nabla_\beta(\vec{A}^\mu \vec{e}_\mu) = (\nabla_\beta A^\mu)\vec{e}_\mu + \vec{A}^\mu \nabla_\beta \vec{e}_\mu \\ &= A^\mu{}_{,\beta}\vec{e}_\mu + A^\mu \Gamma^\alpha{}_{\mu\beta}\vec{e}_\alpha \\ &= (A^\mu{}_{,\beta} + A^\alpha \Gamma^\mu{}_{\alpha\beta})\vec{e}_\mu \ . \end{aligned} \tag{23.30}$$

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

$$A^\mu{}_{,\beta} \equiv \partial_{\vec{e}_\beta} A^\mu \ ; \tag{23.31}$$

i.e. 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 (23.30) to the third, we have renamed the summed-over index α μ and renamed μ α. By comparing the first and last expressions in Eq. (23.30), we conclude that

$$A^\mu{}_{;\beta} = A^\mu{}_{,\beta} + A^\alpha \Gamma^\mu{}_{\alpha\beta} \ . \tag{23.32}$$

The first term in this equation dscribes the changes in $\vec{A}$ associated with changes of its components; the second term corrects for artificial changes of components that are induced by turning and length changes of the basis vectors.

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

$$A_{\alpha;\beta} = A_{\alpha,\beta} - \Gamma^\mu{}_{\alpha\beta} \ , \tag{23.33}$$

where again $A_{\alpha,\beta} \equiv \partial_\beta A_\alpha$. Notice that when the index being "corrected" is down [Eq. (23.33)], the connection coefficient has a minus sign; when it is up [Eq. (23.32)], the connection coefficient has a plus sign. This is in accord with the signs in Eqs. (23.28)–(23.29).

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

$$F^{\alpha\beta}{}_{;\gamma} = F^{\alpha\beta}{}_{,\gamma} + \Gamma^\alpha{}_{\mu\gamma}F^{\mu\beta} + \Gamma^\beta{}_{\mu\gamma}F^{\alpha\mu} \ , \quad F_{\alpha\beta;\gamma} = F_{\alpha\beta,\gamma} - \Gamma^\mu{}_{\alpha\gamma}F_{\mu\beta} - \Gamma^\mu{}_{\beta\gamma}F_{\alpha\mu} \ . \tag{23.34}$$

Notice that each index of $\mathbf{F}$ 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. 10.3). 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 below for a proof:

(i) Evaluate the *commutation coefficients* $c_{\alpha\beta}{}^{\rho}$ of the basis, which are defined by the two equivalent relations

$$[\vec{e}_\alpha, \vec{e}_\beta] \equiv c_{\alpha\beta}{}^{\rho}\vec{e}_\rho \quad , \qquad c_{\alpha\beta}{}^{\rho} \equiv \vec{e}^{\rho} \cdot [\vec{e}_\alpha, \vec{e}_\beta] \; . \tag{23.35}$$

[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, $c_{\alpha\beta}{}^{\rho}{}_{\rm here} = c^{\rho}_{\alpha\beta\,{\rm Lie\ groups}}$.] ($ii$) Lower the last index on the commutation coefficients using the metric components in the basis:

$$c_{\alpha\beta\gamma} \equiv c_{\alpha\beta}{}^{\rho} g_{\rho\gamma} \; . \tag{23.36}$$

(iii) Compute the *covariant Christoffel symbols*

$$\Gamma_{\alpha\beta\gamma} \equiv \frac{1}{2}(g_{\alpha\beta,\gamma} + g_{\alpha\gamma,\beta} - g_{\beta\gamma,\alpha} + c_{\alpha\beta\gamma} + c_{\alpha\gamma\beta} - c_{\beta\gamma\alpha}) \; . \tag{23.37}$$

Here the commas denote differentiation with respect to the basis vectors as though the connection coefficients were scalar fields [Eq. (23.31)]. Notice that the pattern of indices is the same on the g's and on the c's. 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's will vanish and $\Gamma_{\alpha\beta\gamma}$ will be symmetric in its last two indices; in an orthonormal basis $g_{\mu\nu}$ are constant so the g's 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's and the g's will vanish, so $\Gamma_{\alpha\beta\gamma}$ will vanish. (iv) Raise the first index on the covariant Christoffel symbols to obtain the connection coefficients, which are also sometimes called the *mixed Christoffel symbols*

$$\Gamma^{\mu}{}_{\beta\gamma} = g^{\mu\alpha}\Gamma_{\alpha\beta\gamma} \; . \tag{23.38}$$

The gradient operator $\vec{\nabla}$ is an example of a geometric object that is not a tensor. The connection coefficients can be regarded as the components of $\vec{\nabla}$; and because $\vec{\nabla}$ is not a tensor, these components $\Gamma^{\alpha}{}_{\beta\gamma}$ do not obey the tensorial transformation law (23.18) when switching from one basis to another. Their transformation law is far more complicated and is very rarely used. Normally one computes them from scratch in the new basis, using the above prescription or some other, equivalent prescription (cf. Chap. 14 of MTW). 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 manipulations software such as Macsyma, or GRTensor (running under Maple or Mathematica), or Mathtensor (under Mathematica). Such software is easily found on the Internet using a search engine.

The above prescription for computing the connection coefficients follows from two key properties of the gradient $\vec{\nabla}$: *First*, The gradient of the metric tensor vanishes,

$$\vec{\nabla}\mathbf{g} = 0 \; . \tag{23.39}$$

This can be seen by introducing a locally orthornormal coordinate basis at the arbitrary point $\mathcal{P}$ where the gradient is to be evaluated. In such a basis, the effects of curvature show up only at quadratic order in distance away from $\mathcal{P}$, which means that the coordinate bases $\vec{e}_\alpha \equiv \partial/\partial x^\alpha$ behave, at first order in distance, just like those of an orthonormal coordinate system in flat space. Since $\nabla_\beta \vec{e}_\alpha$ involves only first derivatives and it vanishes in an orthonormal coordinate system in flat space, it must also vanish here—which means that the connection coefficients vanish at $\mathcal{P}$ in this basis. Therefore, the components of $\vec{\nabla}\mathbf{g}$ at $\mathcal{P}$ are $g_{\alpha\beta;\gamma} = g_{\alpha\beta,\gamma} = \partial g_{\alpha\beta}/\partial x^\gamma$, which vanishes since the components of $\mathbf{g}$ in this basis are all 0 or ± 1 plus corrections second order in distance from $\mathcal{P}$. This vanishing of the components of $\vec{\nabla}\mathbf{g}$ in our special basis guarantees that $\vec{\nabla}\mathbf{g}$ itself vanishes at $\mathcal{P}$; and since $\mathcal{P}$ was an arbitrary point, $\vec{\nabla}\mathbf{g}$ must vanish everywhere and always.

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

$$\nabla_{\vec{A}}\vec{B} - \nabla_{\vec{B}}\vec{A} = [\vec{A}, \vec{B}] \,. \tag{23.40}$$

This relation, like $\vec{\nabla}\mathbf{g} = 0$, is most easily derived by introducing a locally orthonormal coordinate basis at the point $\mathcal{P}$ where one wishes to check its validity. Since $\Gamma^\mu{}_{\alpha\beta} = 0$ at $\mathcal{P}$ in that basis, the components of $\nabla_{\vec{A}}\vec{B} - \nabla_{\vec{B}}\vec{A}$ are $B^\alpha{}_{;\beta}A^\beta - A^\alpha{}_{;\beta}B^\beta = B^\alpha{}_{,\beta}A^\beta - A^\alpha{}_{,\beta}B^\beta$ [cf. Eq. (23.32)]. But these components are identical to those of the commutator $[\vec{A}, \vec{B}]$ [Eq. (23.25)]. Since the components of these two vectors [the left and right sides of (23.40)] are identical at $\mathcal{P}$ in this special basis, the vectors must be identical, and since the point $\mathcal{P}$ was arbitrary, they must always be identical.

Turn, now, to the derivation of our prescription for computing the connection coefficients in an arbitrary basis. By virtue of the relation $\Gamma^\mu{}_{\beta\gamma} = g^{\mu\alpha}\Gamma_{\alpha\beta\gamma}$ [Eq. (23.38)] and its inverse

$$\Gamma_{\alpha\beta\gamma} = g_{\alpha\mu}\Gamma^\mu{}_{\beta\gamma} \,, \tag{23.41}$$

a knowledge of $\Gamma_{\alpha\beta\gamma}$ is equivalent to a knowledge of $\Gamma^\mu{}_{\beta\gamma}$. Thus, our task reduces to deriving expression (23.37) for $\Gamma_{\alpha\beta\gamma}$, in which the $c_{\alpha\beta\gamma}$ are defined by equations (23.35) and (23.36). As a first step in the derivation, notice that the constancy of the metric tensor, $\vec{\nabla}\mathbf{g} = 0$, when expressed in component notation using Eq. (23.34), and when combined with Eq. (23.41), becomes $0 = g_{\alpha\beta;\gamma} = g_{\alpha\beta,\gamma} - \Gamma_{\beta\alpha\gamma} - \Gamma_{\alpha\beta\gamma}$; i.e.,

$$\Gamma_{\alpha\beta\gamma} + \Gamma_{\beta\alpha\gamma} = g_{\alpha\beta,\gamma} \,. \tag{23.42}$$

This determines the part of $\Gamma_{\alpha\beta\gamma}$ that is symmetric in the first two indices. The commutator of the basis vectors determines the part antisymmetric in the last two indices: From

$$c_{\alpha\beta}{}^\mu \vec{e}_\mu = [\vec{e}_\alpha,\ \vec{e}_\beta] = \nabla_\alpha \vec{e}_\beta - \nabla_\beta \vec{e}_\alpha = (\Gamma^\mu{}_{\beta\alpha} - \Gamma^\mu{}_{\alpha\beta})\vec{e}_\mu \tag{23.43}$$

(where the first equality is the definition (23.35) of the commutation coefficient, the second is expression (23.40) for the commutator in terms of the gradient, and the third follows from the definition (23.27) of the connection coefficient), we infer, by equating the components and lowering the μ index, that

$$\Gamma_{\gamma\beta\alpha} - \Gamma_{\gamma\alpha\beta} = c_{\alpha\beta\gamma} \,. \tag{23.44}$$

By combining equations (23.42) and (23.44) and performing some rather tricky algebra (cf. Ex. 8.15 of MTW), we obtain the computational rule (23.37).

23.3.4 Integration

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

First, the volume elements used in integration involve the Levi-Civita tensor [Eqs. (1.109)–(1.111) and (1.116)], so we need to know the components of the Levi-Civita tensor in a general basis. It turns out [see, e.g., Ex. 8.3 of MTW] that the covariant components differ from those in an orthonormal basis by a factor $\sqrt{|g|}$ and the contravariant by $1/\sqrt{|g|}$, where

$$g \equiv \det ||g_{\alpha\beta}|| \tag{23.45}$$

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. (1.97)]:

$$\begin{aligned}
[12\ldots N] &= +1 \;, \\
[\alpha\beta\ldots\nu] &= +1 \;\; \text{if } \alpha,\, \beta,\, \ldots,\, \nu \text{ is an even permutation of } 1,\, 2,\, \ldots,\, N \\
&= -1 \;\; \text{if } \alpha,\, \beta,\, \ldots,\, \nu \text{ is an odd permutation of } 1,\, 2,\, \ldots,\, N \\
&= 0 \quad\;\, \text{if } \alpha,\, \beta,\, \ldots,\, \nu \text{ are not all different.}
\end{aligned} \tag{23.46}$$

(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

$$\epsilon_{\alpha\beta\ldots\nu} = \sqrt{|g|}\; [\alpha\beta\ldots\nu] \;, \quad \epsilon^{\alpha\beta\ldots\nu} = \pm\frac{1}{\sqrt{|g|}}\; [\alpha\beta\ldots\nu] \;, \tag{23.47}$$

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 three-dimensional Euclidean space, and use the three infinitesimal vectors $dx^j(\partial/\partial x^j)$ to construct the volume element $d\Sigma$ [cf. Eq. (1.110)]:

$$d\Sigma = \boldsymbol{\epsilon}\left(dr\frac{\partial}{\partial r}, d\theta\frac{\partial}{\partial\theta}, d\phi\frac{\partial}{\partial\phi}\right) = \epsilon_{r\theta\phi}drd\theta d\phi = \sqrt{g}drd\theta d\phi = r^2\sin\theta drd\theta d\phi \;. \tag{23.48}$$

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 (23.47) for the components, and the fourth equality entails the determinant of the metric coefficients, which in spherical coordinates are $g_{rr} = 1$, $g_{\theta\theta} = r^2$, $g_{\phi\phi} = r^2\sin^2\theta$, all other g_{jk} vanish, so $g = r^4\sin^2\theta$. The resulting volume element $r^2\sin\theta d\theta d\phi$ should be familiar and obvious.

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

$$\int_{\partial\mathcal{V}} T^{\alpha\beta}d\Sigma_\beta \tag{23.49}$$

[Eq. (1.134)] involve constructing a vector $T^{\alpha\beta}d\Sigma_\beta$ in each infinitesimal region $d\Sigma_\beta$ of the surface of integration, and then adding up the contributions from all the infinitesimal regions.

A major difficulty arises from the fact that 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. As we shall see in Chap. 24, 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. As a result, integrals such as (23.49) 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 energy, momentum, and angular momentum conservation.

****

EXERCISES

Exercise 23.3 *Problem: Practice with Frame-Independent Tensors*
Let $\mathbf{A}, \mathbf{B}$ be second rank tensors.

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

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

(c) Show that the contraction of $\mathbf{A} \otimes \mathbf{B}$ on its first and fourth slots is a second rank tensor. (If necessary, consult Chap. 1 for a discussion of contraction).

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

Exercise 23.4 *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 (23.7), the definition (23.18) of components of a tensor and the relation $\vec{A} \cdot \vec{B} = \mathbf{g}(\vec{A}, \vec{B})$ between the metric and the inner product to deduce the following results:

(a) The relations

$$\vec{e}^{\mu} = g^{\mu\alpha} \vec{e}_\alpha \ , \quad \vec{e}_\alpha = g_{\alpha\mu} \vec{e}^{\mu} \ . \tag{23.50}$$

(b) The fact that indices on the components of tensors can be raised and lowered using the components of the metric, e.g.

$$F^{\mu\nu} = g^{\mu\alpha} F_\alpha{}^\nu \ , \quad p_\alpha = g_{\alpha\beta} p^\beta \ . \tag{23.51}$$

(c) The fact that a tensor can be reconstructed from its components in the manner of Eq. (23.10).

Exercise 23.5 *Practice: Transformation Matrices for Circular Polar Bases*
Consider the circular coordinate system $\{\varpi, \phi\}$ and its coordinate bases and orthonormal bases as discussed in Fig. 23.3 and 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) Evaluate, similarly, 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?

Exercise 23.6 *Practice: Commutation and Connection Coefficients for Circular Polar Bases*
As in the previous exercise, consider the circular coordinates $\{\varpi, \phi\}$ of Fig. 23.3 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. (23.35)–(23.38). [Note: the answer for the orthonormal basis was worked out by a different method in our study of elasticity theory; Eq. (10.20).]

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

Exercise 23.7 *Practice: Connection Coefficients for Spherical Polar Coordinates*

(a) Consider spherical polar coordinates in 3-dimensional space and verify that the non-zero connection coefficients assuming an orthonormal basis are given by Eq. (10.21).

(b) Repeat the exercise assuming a coordinate basis with

$$\mathbf{e}_r \equiv \frac{\partial}{\partial r}, \qquad \mathbf{e}_\theta \equiv \frac{\partial}{\partial \theta}, \qquad \mathbf{e}_\phi \equiv \frac{\partial}{\partial \phi} . \tag{23.52}$$

(c) Repeat both computations using symbolic manipulation software on a computer.

Exercise 23.8 *Practice: Index Gymnastics — Geometric Optics*
In the geometric optics approximation (Chap. 6), 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 Chap. 6, one can deduce that the wave vector, defined by $\vec{k} \equiv \boldsymbol{\nabla}\varphi$, is null: $\vec{k} \cdot \vec{k} = 0$.

(a) Rewrite all of 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 vector field) satisfies the geodesic equation, $\nabla_{\vec{k}}\vec{k} = 0$. The geodesics, to which $\vec{k}$ is the tangent vector, are the rays discussed in Chap. 6, along which the waves propagate.

Exercise 23.9 *Practice: Index Gymnastics — Irreducible Tensorial Parts of the Gradient of a 4-Velocity Field*
In our study of elasticity theory, we introduced the concept of the irreducible tensorial parts of a second-rank tensor in Euclidean space (Box. 10.1). Consider a fluid flowing through spacetime, with a 4-velocity $\vec{u}(\mathcal{P})$. The fluid's gradient $\boldsymbol{\nabla}\vec{u}$ ($u_{\alpha;\beta}$ 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:

$$u_{\alpha;\beta} = -a_\alpha u_\beta + \frac{1}{3}\theta P_{\alpha\beta} + \sigma_{\alpha\beta} + \omega_{\alpha\beta} \; . \tag{23.53}$$

Here: (i) $P_{\alpha\beta}$ is defined by

$$P_{\alpha\beta} \equiv g_{\alpha\beta} + u_\alpha u_\beta \; , \tag{23.54}$$

(ii) $\sigma_{\alpha\beta}$ is symmetric and is orthogonal to the 4-velocity, and (iii) $\omega_{\alpha\beta}$ is antisymmetric and is orthogonal to the 4-velocity.

(a) In quantum mechanics one deals with "projection operators" $\hat{P}$, which satisfy the equation $\hat{P}^2 = \hat{P}$. Show that $P_{\alpha\beta}$ is a projection tensor, in the sense that $P_{\alpha\beta}P^{\beta}{}_{\gamma} = P_{\alpha\gamma}$.

(b) This suggests that $P_{\alpha\beta}$ may project vectors into some subspace of 4-dimensional spacetime. Indeed it does: Show that for any vector A^α, $P_{\alpha\beta}A^\beta$ is orthogonal to $\vec{u}$; and if A^α is already perpendicular to $\vec{u}$, then $P_{\alpha\beta}A^\beta = A_\alpha$, i.e. the projection leaves the vector unchanged. Thus, $P_{\alpha\beta}$ projects vectors into the 3-space orthogonal to $\vec{u}$.

(c) What are the components of $P_{\alpha\beta}$ in the fluid's local rest frame, i.e. in an orthonormal basis where $\vec{u} = \vec{e}_{\hat{0}}$?

(d) 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 (23.53). Show, further, that $\vec{a}\cdot\vec{u} = 0$.

(e) 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 (23.53).

(f) The quantities $\sigma_{\alpha\beta}$ and $\omega_{\alpha\beta}$ are the relativistic versions of the fluid's shear and rotation tensors. Derive equations for these tensors in terms of $u_{\alpha;\beta}$ and $P_{\mu\nu}$.

(g) 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 are true: (i) $u^0 = 1$, $u^j = v^j$; (ii) θ is the nonrelativistic expansion of the fluid, $\theta = \boldsymbol{\nabla}\cdot\mathbf{v} \equiv v^j{}_{,j}$ [Eq. (12.57)]; (iii) σ_{jk} is the fluid's nonrelativistic shear [Eq. (12.57)]; (iv) ω_{jk} is the fluid's nonrelativist rotation tensor [denoted r_{jk} in Eq. (12.57)].

Exercise 23.10 *Practice: Integration — Gauss's Law*
In 3-dimensional Euclidean space the Maxwell equation $\boldsymbol{\nabla}\cdot\mathbf{E} = \rho_e/\epsilon_0$ can be combined with Gauss's law 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:

$$\int_{\partial\mathcal{V}} \mathbf{E}\cdot d\boldsymbol{\Sigma} = \int_{\mathcal{V}} (\rho_e/\epsilon_0) d\Sigma \; . \tag{23.55}$$

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 $d\boldsymbol{\Sigma} = \epsilon(..., d\theta\partial/\partial\theta, d\phi\partial/\partial\phi)$. Similarly, evaluate $d\Sigma$ in terms of vectorial legs in the sphere's interior. Then use these results for $d\Sigma_j$ and $d\Sigma$ to convert Eq. (23.55) into an explicit form in terms of integrals over r, θ, ϕ. The final answer should be obvious, but the above steps in deriving it are informative.

****

23.4 The Stress-Energy Tensor Revisited

In Sec. 1.9 we defined the *stress-energy tensor* $\mathbf{T}$ 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 $\mathbf{P}$ that flows through some small 3-volume $\boldsymbol{\Sigma}$, going from the negative side of $\boldsymbol{\Sigma}$ to its positive side, is

$$\mathbf{T}(\ldots, \vec{\Sigma}) = (\text{total 4-momentum } \vec{P} \text{ that flows through } \vec{\Sigma}); \quad \text{i.e., } T^{\alpha\beta}\Sigma_\beta = P^\alpha \tag{23.56}$$

[Eq. (1.121)]. 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 $\mathbf{T}(\mathcal{P})$.

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. (1.122)–(1.127)]: 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.

We gained some insight into the stress-energy tensor in the context of kinetic theory in Sec. 2.5, and we briefly introduced the stress-energy tensor for a perfect fluid in the last paragraph of Sec. 1.12. Because perfect fluids will play a very important role in this book's applications of general relativity to relativistic stars (Chap. 25) and cosmology (Chap. 27), we shall now explore the perfect-fluid stress-energy tensor in some depth, and shall see how it is related to the Newtonian description of perfect fluids, which we studied in Part IV.

Recall [Eq. (1.138)] 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:

$$T^{00} = \rho \; , \quad T^{jk} = P\delta^{jk} \; . \tag{23.57}$$

From this special form of $T^{\alpha\beta}$ in the local rest frame, one can derive the following expression for the stress-energy tensor in terms of the 4-velocity $\vec{u}$ of the local rest frame (i.e., of the fluid itself), the metric tensor of spacetime $\mathbf{g}$, and the rest-frame energy density ρ and pressure P:

$$T^{\alpha\beta} = (\rho + P)u^\alpha u^\beta + P g^{\alpha\beta} \; ; \quad \text{i.e., } \mathbf{T} = (\rho + P)\vec{u} \otimes \vec{u} + P\mathbf{g} \; ; \tag{23.58}$$

see Ex. 23.11, below. This expression for the stress-energy tensor of a perfect fluid is an example of a geometric, frame-independent description of physics.

It is instructive to evaluate the nonrelativistic limit of this perfect-fluid stress-energy tensor and verify that it has the form we used in our study of nonrelativistic, inviscid fluid mechanics (Table 12.1 on page 19 of Chap. 12, with vanishing gravitational potential $\Phi = 0$). In the nonrelativistic limit the fluid is nearly at rest in the chosen Lorentz reference frame. It moves with ordinary velocity $\mathbf{v} = d\mathbf{x}/dt$ that is small compared to the speed of light, so the temporal part of its 4-velocity $u^0 = 1/\sqrt{1 - v^2}$ and spatial part $\mathbf{u} = u^0\mathbf{v}$ can be approximated as

$$u^0 \simeq 1 + \frac{1}{2}v^2 \; , \quad \mathbf{u} \simeq \left(1 + \frac{1}{2}v^2\right)\mathbf{v} \; . \tag{23.59}$$

In the fluid's rest frame, it has a rest mass density ρ_o, an internal energy per unit rest mass u (not to be confused with the 4-velocity), and a total density of mass-energy

$$\rho = \rho_o(1 + u) \; . \tag{23.60}$$

Now, in our chosen Lorentz frame the volume of each fluid element is Lorentz contracted by the factor $\sqrt{1 - v^2}$ and therefore the rest mass density is increased from ρ_o to $\rho_o/\sqrt{1 - v^2} = \rho_o u^0$; and correspondingly the rest-mass flux is $\rho_o u^0 \mathbf{v} = \rho_o \mathbf{u}$, and the law of rest-mass conservation is $\partial(\rho_o u^0)/\partial t + \partial(\rho_o u^j)/\partial x^j = 0$, i.e. $\vec{\nabla} \cdot (\rho_o \mathbf{v}) = 0$. When taking the Newtonian limit, we should identify the Newtonian mass ρ_N with the low-velocity limit of this rest mass density:

$$\rho_N = \rho_o u^0 \simeq \rho_o \left(1 + \frac{1}{2}v^2\right) \; . \tag{23.61}$$

The nonrelativistic limit regards the specific internal energy u, the kinetic energy per unit mass $\frac{1}{2}v^2$, and the ratio of pressure to rest mass density P/ρ_o as of the same order of smallness

$$u \sim \frac{1}{2}v^2 \sim \frac{P}{\rho_o} \ll 1 \; , \tag{23.62}$$

and it expresses the momentum density T^{j0} accurate to first order in $v \equiv |\mathbf{v}|$, the momentum flux (stress) T^{jk} accurate to second order in v, the energy density T^{00} accurate to second order in v, and the energy flux T^{0j} accurate to third order in v. To these accuracies, the perfect-fluid stress-energy tensor (23.58) takes the following form:

$$\begin{aligned} T^{j0} &= \rho_N v^j \; , & T^{jk} &= P g^{jk} + \rho_N v^j v^k \; , \\ T^{00} &= \rho_N + \frac{1}{2}\rho_N v^2 + \rho_N u \; , & T^{0j} &= \rho_N v^j + \left(\frac{1}{2}v^2 + \frac{P}{\rho_N}\right)\rho_N v^j \; ; \end{aligned} \tag{23.63}$$

see Ex. 23.11(c). These are precisely the same as the momentum density, momentum flux, energy density, and energy flux that we used in our study of nonrelativistic, inviscid fluid mechanics (Chap. 12), aside from the notational change from there to here $\rho \rightarrow \rho_N$, and aside from including the rest mass-energy $\rho_N = \rho_N c^2$ in T_{00} here but not there, and including the rest-mass-energy flux $\rho_N v^j$ in T^{0j} here but not there.

Just as the nonrelativistic equations of fluid mechanics (Euler equation and energy conservation) are derivable by combining the nonrelativistic $T^{\alpha\beta}$ of Eq. (23.63) with the nonrelativistic laws of momentum and energy conservation, so also the relativistic equations of fluid mechanics are derivable by combining the relativistic version (23.58) of $T^{\alpha\beta}$ with the equation of 4-momentum conservation $\vec{\nabla} \cdot \mathbf{T} = 0$. (We shall give such a derivation and shall examine the resulting fluid mechanics equations in the context of general relativity in Chap. 24.) This, together with the fact that the relativistic $\mathbf{T}$ reduces to the nonrelativistic $T^{\alpha\beta}$ in the nonrelativistic limit, guarantees that the special relativistic equations of inviscid fluid mechanics will reduce to the nonrelativistic equations in the nonrelativistic limit.

A second important example of a stress-energy tensor is that for the electromagnetic field. We shall explore it in Ex. 23.13 below.

For a point particle which moves through spacetime along a world line $\mathcal{P}(\zeta)$ (where ζ is the affine parameter such that the particle's 4-momentum is $\vec{p} = d/d\zeta$), the stress-energy tensor will vanish everywhere except on the world line itself. Correspondingly, $\mathbf{T}$ 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 $d\Sigma = dt dx dy dz$), one obtains

$$\int_{\mathcal{V}} f(\mathcal{P}) \delta(\mathcal{Q}, \mathcal{P}) d\Sigma = f(\mathcal{Q}) . \tag{23.64}$$

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 verify that in terms of Lorentz coordinates this delta function can be expressed as

$$\delta(\mathcal{Q}, \mathcal{P}) = \delta(t_{\mathcal{Q}} - t_{\mathcal{P}}) \delta(x_{\mathcal{Q}} - x_{\mathcal{P}}) \delta(y_{\mathcal{Q}} - y_{\mathcal{P}}) \delta(z_{\mathcal{Q}} - z_{\mathcal{P}}) , \tag{23.65}$$

where the deltas on the right-hand side are ordinary one-dimensional Dirac delta functions.

In terms of the spacetime delta function $\delta(\mathcal{Q}, \mathcal{P})$ the stress-energy tensor of a point particle takes the form

$$\mathbf{T}(\mathcal{Q}) = \int_{-\infty}^{+\infty} \vec{p}(\zeta) \otimes \vec{p}(\zeta) \delta\left(\mathcal{Q}, \mathcal{P}(\zeta)\right) d\zeta , \tag{23.66}$$

where the integral is along the world line $\mathcal{P}(\zeta)$ of the particle. It is a straightforward but sophisticated exercise [Ex. 23.14] to verify that the integral of this stress-energy tensor over any 3-surface $\mathcal{S}$ that slices through the particle's world line just once, at an event $\mathcal{P}(\zeta_o)$, is equal to the particle's 4-momentum at the intersection point:

$$\int_{\mathcal{S}} T^{\alpha\beta}(\mathcal{Q}) d\Sigma_\beta = p^\alpha(\zeta_o) . \tag{23.67}$$

****

EXERCISES

Exercise 23.11 *Derivation: Stress-Energy Tensor for a Perfect Fluid*

(a) Derive the frame-independent expression (23.58) for the perfect fluid stress-energy tensor from its rest-frame components (23.57).

(b) Show that for a perfect fluid the inertial mass per unit volume (Ex. 2.4) is isotropic and is equal to $(\rho + P)\delta^{ij}$ when thought of as a tensor, or simply $\rho + P$ when thought of as a scalar.

(c) Show that in the nonrelativistic limit the components of the perfect fluid stress-energy tensor (23.58) take on the forms (23.63), and verify that these agree with the densities and fluxes of energy and momentum that are used in nonrelativistic fluid mechanics (e.g., Table 12.1 on page 19 of Chap. 12).

(d) Show that it is the contribution of the pressure P to the relativistic density of inertial mass that causes the term $(P/\rho_N)\rho_N\mathbf{v} = P\mathbf{v}$ to appear in the nonrelativistic energy flux.

Exercise 23.12 *Problem: Electromagnetic Field Tensor*
As we saw in Sec. 1.6, in 4-dimensional spacetime the electromagnetic field is described by a second-rank tensor $\mathsf{F}(\ldots,\ldots)$ which is antisymmetric on its two slots, $F^{\alpha\beta} = -F^{\beta\alpha}$; and the 4-force (rate of change of 4-momentum) that it exerts on a particle with rest mass m, charge q, proper time τ, 4-velocity $\vec{u} = d/d\tau$, and 4-momentum $\vec{p}$ is

$$\frac{d\vec{p}}{d\tau} = \nabla_{\vec{u}}\vec{p} = q\mathsf{F}(\ldots,\vec{u}) \; ; \quad \text{i.e.,} \; \frac{dp^\alpha}{d\tau} = qF^{\alpha\beta}u_\beta \; . \tag{23.68}$$

Here the second form of the equation, valid in a Lorentz frame, follows from the component form of $\nabla_{\vec{u}}\vec{p}$: $p^\alpha{}_{;\mu}u^\mu = p^\alpha{}_{,\mu}u^\mu = dp^\alpha/d\tau$.

(a) By comparing this with the Lorentz force law for a low-velocity particle, $d\mathbf{p}/dt = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, show that the components of the electromagnetic field tensor in a Lorentz reference frame are

$$|| \, F^{\alpha\beta} \, || = \left\| \begin{array}{cccc} 0 & E^x & E^y & E^z \\ -E^x & 0 & B^z & -B^y \\ -E^y & -B^z & 0 & B^x \\ -E^z & B^y & -B^x & 0 \end{array} \right\| \; ; \tag{23.69}$$

$$\text{i.e.,} \; F^{0i} = -F^{i0} = E^i \; , \quad F^{ij} = -F^{ji} = \epsilon^{ij}{}_k B^k \; , \tag{23.70}$$

where E^j and B^j are the components of the 3-vector electric and magnetic fields that reside in the 3-space of the Lorentz frame.

(b) Define $*\mathsf{F} \equiv$(“dual” of F) by

$$*F_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta} \; , \tag{23.71}$$

where $\boldsymbol{\epsilon}$ is the Levi-Civita tensor of Eq. (1.93). What are the components of $*\mathsf{F}$ in a Lorentz frame in terms of that frame's electric and magnetic fields [analog of Eq. (23.69)]?

(c) Let $\vec{u}$ be the 4-velocity of some observer. Show that the 4-vectors $\mathsf{F}(\ldots,\vec{u}) \equiv \vec{E}_{\vec{u}}$ and $-*\mathsf{F}(\ldots,\vec{u}) \equiv \vec{B}_{\vec{u}}$ lie in the 3-space of that observer's local rest frame (i.e., they are orthogonal to the observer's 4-velocity), and are equal to the electric and magnetic fields of that 3-space, i.e., the electric and magnetic fields measured by that observer.

(d) There are only two independent scalars constructable from the electromagnetic field tensor: $F^{\mu\nu}F_{\mu\nu}$ and $*F^{\mu\nu}F_{\mu\nu}$. Show that, when expressed in terms of the electric and magnetic fields measured by any observer (i.e., of any Lorentz reference frame), these take the form

$$F^{\mu\nu}F_{\mu\nu} = 2(\mathbf{B}^2 - \mathbf{E}^2) \; , \quad *F^{\mu\nu}F_{\mu\nu} = 4\mathbf{B}\cdot\mathbf{E} \; . \tag{23.72}$$

Exercise 23.13 *Problem: Electromagnetic Stress-energy Tensor*

Expressed in geometric, frame-independent language, the Maxwell equations take the form

$$\begin{aligned} F^{\alpha\beta}{}_{;\beta} &= 4\pi J^\alpha \; , \\ *F^{\alpha\beta}{}_{;\beta} &= 0 \; , \quad \text{or equivalently } F_{\alpha\beta;\gamma} + F_{\beta\gamma;\alpha} + F_{\gamma\alpha;\beta} = 0 \; . \end{aligned} \tag{23.73}$$

Here J^α is the density-current 4-vector, whose components in a specific Lorentz frame have the physical meanings

$$J^0 = (\text{charge density}) \; , \quad J^i = (i\text{-component of current density}) \; . \tag{23.74}$$

The stress-energy tensor for the electromagnetic field has the form

$$T^{\mu\nu} = \frac{1}{4\pi}(F^{\mu\alpha}F^\nu{}_\alpha - \frac{1}{4}g^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}) \; . \tag{23.75}$$

(a) Show that in any Lorentz reference frame the electromagnetic energy density T^{00}, energy flux T^{0j}, momentum density T^{j0}, and stress T^{jk} have the following forms when expressed in terms of the electric and magnetic fields measured in that frame:

$$\begin{aligned} T^{00} &= \frac{\mathbf{E}^2 + \mathbf{B}^2}{8\pi} \; , \quad T^{i0} = T^{0i} = \frac{\epsilon^i{}_{jk}E^jB^k}{4\pi} \; , \\ T^{jk} &= \frac{1}{8\pi}\left[(\mathbf{E}^2 + \mathbf{B}^2)g^{jk} - 2(E^jE^k + B^jB^k)\right] \; . \end{aligned} \tag{23.76}$$

Show that, expressed in index-free notation, the energy flux has the standard Poynting-vector form $\mathbf{E}\times\mathbf{B}/4\pi$, and the stress tensor consists of a pressure $P_\perp = \mathbf{E}^2/8\pi$ orthogonal to $\mathbf{E}$, a pressure $P_\perp = \mathbf{B}^2/8\pi$ orthogonal to $\mathbf{B}$, a tension $-P_{||} = \mathbf{E}^2/8\pi$ along $\mathbf{E}$, and a tension $-P_{||} = \mathbf{B}^2/8\pi$ along $\mathbf{B}$.

(b) Show that the divergence of the stress-energy tensor (23.80) is given by

$$T^{\mu\nu}{}_{;\nu} = \frac{1}{4\pi}(F^{\mu\alpha}{}_{;\nu}F^\nu{}_\alpha + F^{\mu\alpha}F^\nu{}_{\alpha;\nu} - \frac{1}{2}F_{\alpha\beta}{}^{;\mu}F^{\alpha\beta}) \; . \tag{23.77}$$

(c) Combine this with the Maxwell equations to show that

$$\boldsymbol{\nabla}\cdot\mathsf{T} = -\mathsf{F}(\ldots,\mathbf{J}) \; ; \quad \text{i.e., } T^{\alpha\beta}{}_{;\beta} = -F^{\alpha\beta}J_\beta \; . \tag{23.78}$$

(d) Show that in a Lorentz reference frame the time and space components of this equation reduce to

$$\frac{\partial}{\partial t}T^{00} + \frac{\partial}{\partial x^j}T^{0j} = -E^j J_j \equiv -(\text{rate of Joule heating}) \,, \tag{23.79}$$

$$\left(\frac{\partial}{\partial t}T^{k0} + \frac{\partial}{\partial x^j}T^{kj}\right) \mathbf{e}_k = -(J^0\mathbf{E} + \mathbf{J}\times\mathbf{B}) = -\left(\begin{array}{c}\text{Lorentz force} \\ \text{per unit volume}\end{array}\right) . \tag{23.80}$$

Explain why these relations guarantee that, although the electromagnetic stress-energy tensor is not divergence-free, the total stress-energy tensor (electromagnetic plus that of the medium or fields that produce the charge-current 4-vector $\vec{J}$) is divergence-free; i.e., the total 4-momentum is conserved.

Exercise 23.14 *Derivation: Stress-Energy Tensor for a Point Particle*
Derive Eq. (23.67).

****

23.5 The Proper Reference Frame of an Accelerated Observer [MTW pp. 163–176, 327–332]

Physics experiments and astronomical measurements almost always use apparatus that accelerates and rotates. For example, if the apparatus is in an earth-bound 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 the earth. It is useful in studying such 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 here shall study them in flat spacetime.

Consider an observer who moves along an accelerated world line through flat spacetime (Fig. 23.6) so she has a nonzero 4-acceleration

$$\vec{a} = \frac{d\vec{u}}{d\tau} \,. \tag{23.81}$$

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. These properties dictate that in the immediate vicinity of her world line the metric has the form

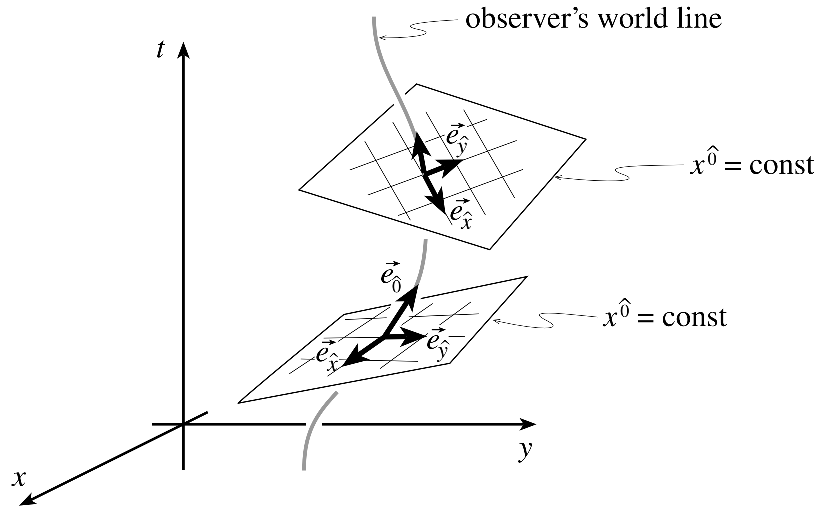

Fig. 23.6: The proper reference frame of an accelerated observer. The spatial basis vectors $\vec{e}_{\hat{x}}$, $\vec{e}_{\hat{y}}$, and $\vec{e}_{\hat{z}}$ are orthogonal to the observer's world line and rotate, relative to local gyroscopes, as they move along the world line. The flat 3-planes spanned 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 "3-space". In each of these flat 3-planes the spatial coordinates $\hat{x}$, $\hat{y}$, $\hat{z}$ are Cartesian, with $\partial/\partial\hat{x} = \vec{e}_{\hat{x}}$, $\partial/\partial\hat{y} = \vec{e}_{\hat{y}}$, $\partial/\partial\hat{z} = \vec{e}_{\hat{z}}$.

$ds^2 = \eta_{\hat{\alpha}\hat{\beta}} dx^{\hat{\alpha}} dx^{\hat{\beta}}$; in other words, all along her world line the coordinate basis vectors are orthonormal:

$$g_{\hat{\alpha}\hat{\beta}} = \frac{\partial}{\partial x^{\hat{\alpha}}} \cdot \frac{\partial}{\partial x^{\hat{\beta}}} = \eta_{\hat{\alpha}\hat{\beta}} \quad \text{at } x^{\hat{j}} = 0 \; . \tag{23.82}$$

Properties (i) and (ii) dictate, moreover, 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 $\vec{U}$,

$$\vec{e}_{\hat{0}} = \frac{\partial}{\partial x^{\hat{0}}} = \vec{U} \; . \tag{23.83}$$

There remains freedom as to how the observer's latticework is oriented, spatially: The observer can lock it to a gyroscopic inertial-guidance system, in which case we shall say that it is "nonrotating", or she can rotate it relative to such gyroscopes. We shall assume that the latticework rotates. Its angular velocity as measured by the observer (by comparing the latticework's orientation with inertial-guidance gyroscopes) is a 3-dimensional, spatial vector $\mathbf{\Omega}$; and as viewed geometrically, it is a 4-vector $\vec{\Omega}$ whose components in the observer's reference frame are $\Omega^{\hat{j}} \neq 0$ and $\Omega^{\hat{0}} = 0$, i.e., it is a 4-vector that is orthogonal to the observer's 4-velocity, $\vec{\Omega} \cdot \vec{U} = 0$; i.e., it is a 4-vector that lies in the observer's 3-space. Similarly, the latticework's acceleration as measured by an accelerometer attached to it is a 3-dimensional spatial vector $\mathbf{a}$ which can be thought of as a 4-vector with components in the observer's frame $a^{\hat{0}} = 0$, $a^{\hat{j}} = (\hat{j}$-component of the measured $\mathbf{a})$. This 4-vector, in fact, 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.

Geometrically the coordinates of the proper reference frame are constructed as follows: (i) Begin with the basis vectors $\vec{e}_{\hat{\alpha}}$ along the observer's world line (Fig. 23.6)—basis vectors that satisfy equations (23.82) and (23.83), 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 $\vec{e}_{\hat{j}} = \partial/\partial x^{\hat{j}}$.

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 shall 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 her 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 which, arbitrarily near this event, coincides with the observer's proper reference frame. Then, if we were doing Newtonian physics, the coordinate transformation from the proper reference frame to the inertial frame would have the form (accurate through terms quadratic in $x^{\hat{\alpha}}$)

$$x^i = x^{\hat{i}} + \frac{1}{2} a^{\hat{i}} (x^{\hat{0}})^2 + \epsilon^{\hat{i}}{}_{\hat{j}\hat{k}} \Omega^{\hat{j}} x^{\hat{k}} x^{\hat{0}} \,, \quad x^0 = x^{\hat{0}} \,. \tag{23.84}$$

Here the term $\frac{1}{2} a^{\hat{j}} (x^{\hat{0}})^2$ is the standard expression for the vectorial displacement produced, after time $x^{\hat{0}}$ by the acceleration $a^{\hat{j}}$; 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 $\Omega^{\hat{j}}$ 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; and 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. (1.84)], accurate only to quadratic order]. Thus, the full transformation to quadratic order is

$$\begin{aligned} x^i &= x^{\hat{i}} + \frac{1}{2} a^{\hat{i}} (x^{\hat{0}})^2 + \epsilon^{\hat{i}}{}_{\hat{j}\hat{k}} \Omega^{\hat{j}} x^{\hat{k}} x^{\hat{0}} \,, \\ x^0 &= x^{\hat{0}} (1 + a_{\hat{j}} x^{\hat{j}}) \,. \end{aligned} \tag{23.85}$$

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:

$$ds^2 = -(1 + 2\mathbf{a} \cdot \mathbf{x})(dx^{\hat{0}})^2 + 2(\boldsymbol{\Omega} \times \mathbf{x}) \cdot d\mathbf{x}\, dx^{\hat{0}} + \delta_{jk} dx^j dx^k \tag{23.86}$$

[Ex. 23.15(a)]. 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 (23.85) was constructed near an arbitrary event on the observer's world line, the metric (23.86) is valid near any and every event on its world line; i.e., it is valid all along the world line. It, in fact, 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, e.g., 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. (23.86)] are $g_{\hat{\alpha}\hat{\beta}} = \eta_{\hat{\alpha}\hat{\beta}}$, in accord with equation (23.82). However, as one moves farther and 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 (23.86) 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,

$$\nabla_{\vec{U}}\vec{e}_{\hat{\alpha}} = \nabla_{\hat{0}}\vec{e}_{\hat{\alpha}} = \Gamma^{\hat{\mu}}{}_{\hat{\alpha}\hat{0}}\vec{e}_{\hat{\mu}} \; . \tag{23.87}$$

The result is (cf. Ex. 23.15)

$$\nabla_{\vec{U}}\vec{e}_{\hat{0}} \equiv \nabla_{\vec{U}}\vec{U} = \vec{a} \; , \tag{23.88}$$

$$\nabla_{\vec{U}}\vec{e}_{\hat{j}} = (\vec{a}\cdot\vec{e}_{\hat{j}})\vec{U} + \boldsymbol{\epsilon}(\vec{U},\vec{\Omega},\vec{e}_{\hat{j}},\; \ldots) \; . \tag{23.89}$$

Equation (23.89) is a special case of a general "law of transport" for 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 (one which the observer carries with herself, applying the forces that make it accelerate precisely at its center of mass so they do not also make it precess), the transport law is (23.89) with $\vec{e}_{\hat{j}}$ replaced by $\vec{S}$ and with $\vec{\Omega} = 0$:

$$\nabla_{\vec{U}}\vec{S} = \vec{U}(\vec{a}\cdot\vec{S}) \; . \tag{23.90}$$

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 $\vec{A}$, which rotates relative to inertial-guidance gyroscopes, the transport law has in addition to this "keep-it-orthogonal-to $\vec{U}$" term, also a second term which is the 4-vector form of $d\mathbf{A}/dt = \boldsymbol{\Omega}\times\mathbf{A}$:

$$\nabla_{\vec{U}}\vec{A} = \vec{U}(\vec{a}\cdot\vec{A}) + \boldsymbol{\epsilon}(\vec{U},\; \vec{\Omega},\; \vec{A},\; \ldots\,) \; . \tag{23.91}$$

Equation (23.89) is this general transport law with $\vec{A}$ replaced by $\vec{e}_{\hat{j}}$.

Consider a particle which moves freely through the neighborhood of an accelerated observer. As seen in an inertial reference frame, the particle moves 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

$$\nabla_{\vec{u}}\vec{u} = 0 \; ; \tag{23.92}$$

i.e., it parallel transports its 4-velocity $\vec{u}$ along itself. 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 (23.92) are

$$u^{\hat{\alpha}}{}_{,\hat{\mu}}u^{\hat{\mu}} + \Gamma^{\hat{\alpha}}{}_{\hat{\mu}\hat{\nu}}u^{\hat{\mu}}u^{\hat{\nu}} = 0 \; ; \tag{23.93}$$

or equivalently

$$\frac{d^2 x^{\hat{\alpha}}}{d\tau^2} + \Gamma^{\hat{\alpha}}{}_{\hat{\mu}\hat{\nu}} \frac{dx^{\hat{\mu}}}{d\tau} \frac{dx^{\hat{\nu}}}{d\tau} = 0 \ . \tag{23.94}$$

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 very nearly equal to $u^{\hat{j}} = dx^{\hat{j}}/d\tau$ and is very small compared to unity (the speed of light), and $u^{\hat{0}} = dx^{\hat{0}}/d\tau$ is very nearly unity. Then to first order in the ordinary velocity $v^{\hat{j}}$, the spatial part of the geodesic equation (23.94) becomes

$$\frac{d^2 x^{\hat{i}}}{(dx^{\hat{0}})^2} = -\Gamma^{\hat{i}}{}_{\hat{0}\hat{0}} - (\Gamma^{\hat{i}}{}_{\hat{j}\hat{0}} + \Gamma^{\hat{i}}{}_{\hat{0}\hat{j}}) v^{\hat{j}} \ . \tag{23.95}$$

By computing the connection coefficients from the metric coefficients of (23.86) [Ex. 23.15], we bring this low-velocity geodesic law of motion into the form

$$\frac{d^2 x^{\hat{i}}}{(dx^{\hat{0}})^2} = -a^{\hat{i}} - 2\epsilon^{\hat{i}}{}_{\hat{j}\hat{k}} \Omega^{\hat{j}} v^{\hat{k}} \ , \quad \text{i.e.,} \ \frac{d^2 \mathbf{x}}{(dx^{\hat{0}})^2} = -\mathbf{a} - 2\boldsymbol{\Omega} \times \mathbf{v} \ . \tag{23.96}$$

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 higher order in distance away from the observer's world line, but it has been lost due to our linearizing the metric (23.86) 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 23.15 *Example: Proper Reference Frame*

(a) Show that the coordinate transformation (23.85) brings the metric $ds^2 = \eta_{\alpha\beta} dx^\alpha dx^\beta$ into the form (23.86), accurate to linear order in separation $x^{\hat{j}}$ from the origin of coordinates.

(b) Compute the connection coefficients for the coordinate basis of (23.86) 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) From those connection coefficients show that the rate of change of the basis vectors $\mathbf{e}_{\hat{\alpha}}$ along the observer's world line is given by (23.88), (23.89).

(d) From the connection coefficients show that the low-velocity limit (23.95) of the geodesic equation is given by (23.96).

Exercise 23.16 *Problem: Uniformly Accelerated Observer*
As a special example of an accelerated observer, consider one whose world line, expressed in terms of a Lorentz coordinate system $(t,\, x,\, y,\, z)$, is

$$t = \frac{1}{a}\sinh(a\tau)\ , \quad x = \frac{1}{a}\cosh(a\tau)\ , \quad y = 0\ , \quad z = 0\ . \tag{23.97}$$

Here a is a constant with dimensions of 1/(length) or equivalently (length)/(time)^2, and τ is a parameter that varies along the accelerated world line.

(a) Show that τ is the observer's proper time, and evaluate the observer's 4-acceleration $\vec{a}$, and show that $|\vec{a}| = a$ where a is the constant in (23.97), so the observer feels constant, time-independent accleration in his proper reference frame.

(b) The basis vectors $\mathbf{e}_{\hat{0}}$ and $\mathbf{e}_{\hat{1}}$ of the observer's proper reference frame lie in the t, x plane in spacetime, $\mathbf{e}_{\hat{2}}$ points along the y-axis, and $\mathbf{e}_{\hat{3}}$ points along the z axis. Draw a spacetime diagram, on it draw the observer's world line, and at several points along it draw the basis vectors $\mathbf{e}_{\hat{\mu}}$. What are $\mathbf{e}_{\hat{\mu}}$ in terms of the Lorentz coordinate basis vectors $\partial/\partial x^\alpha$?

(c) What is the angular velocity $\vec{\Omega}$ of the proper reference frame?

(d) Express the coordinates $x^{\hat{\mu}}$ of the observer's proper reference frame in terms of the Lorentz coordinates $(t,\, x,\, y,\, z)$ accurate to first order in distance away from the observer's world line and accurate for all proper times τ. Show that under this coordinate transformation the Lorentz-frame components of the metric, $g_{\alpha\beta} = \eta_{\alpha\beta}$, are transformed into the components given by Eq. (23.86).

Exercise 23.17 *Challenge: Thomas Precession*
As is well known in quantum mechanics, the spin-orbit contribution to the Hamiltonian for an electron in an atom is

$$H_{SO} = \frac{-e}{2m_e^2c^2r}\frac{d\phi}{dr}\mathbf{L}\cdot\mathbf{S} \tag{23.98}$$

where ϕ is the electrostatic potential and $\mathbf{L}, \mathbf{S}$ are the electron's angular momentum and spin respectively. This is one half the naive value and the difference, known as the *Thomas precession*, is a purely special relativistic kinematic effect. Using the language of this chapter, explain from first principles how the Thomas precession arises.

****

Bibliography

Feynman, R. P., 1966. *The Character of Physical Law*, Cambridge MA: MIT Press.

Lightman, A. P., Press, W. H., Price, R. H. & Teukolsky, S. A. 1975. *Problem Book In Relativity and Gravitation*, Princeton: Princeton University Press.

MTW: Misner, Charles W., Thorne, Kip S., and Wheeler, John A. 1973. *Gravitation*, San Francisco:W. H. Freeman.

Ni, W.-T., and Zimmermann, M., 1978. "Inertial and gravitational effects in the proper reference frame of an accelerated, rotating observer," *Physical Review D*, **17**, 1473.

Schutz, B. 1980. *Geometrical Methods of Mathematical Physics*, Cambridge: Cambridge University Press.

Wald, R. M. 1984. *General Relativity*, Chicago: University of Chicago Press.