Class: M. Sc.

Subject: Mathematical Methods for Physics

I have taken all course materials for Mathematical Methods for Physics from G. B. Arfken and H.J. Weber, Mathematical Methods for Physicists.

Students can download this book form given web address;

Web address: https://b-ok.cc/book/446170/6e805d

All topics have been taken from **Chapter-2 & Chapter-4** from above said book (**https://b-ok.cc/book/446170/6e805d**). I am sending pdf file of Chapter 2 and Chapter 4 .

Name of Topics

Vector Analysis in Curved Coordinates and Tensors, Orthogonal Coordinates in R3, Differential Vector Operators, Special Coordinate Systems, Circular Cylinder Coordinates, Spherical and Polar Coordinates. Tensor Analysis, Contraction, Direct Product, Quotient Rule, Pseudotensors, Dual Tensors, General Tensors, Tensor Derivative Operators. Group Theory, Introduction, Generators of Continuous Group, Orbital Angular Momentum, Angular Momentum Coupling, Homogeneous Lorentz Group, Discrete Groups

CHAPTER 2

VECTOR ANALYSIS IN CURVED COORDINATES AND TENSORS

In Chapter 1 we restricted ourselves almost completely to rectangular or Cartesian coordinate systems. A Cartesian coordinate system offers the unique advantage that all three unit vectors, $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$, are constant in direction as well as in magnitude. We did introduce the radial distance r, but even this was treated as a function of x, y, and z. Unfortunately, not all physical problems are well adapted to a solution in Cartesian coordinates. For instance, if we have a central force problem, $\mathbf{F} = \hat{\mathbf{r}} F(r)$, such as gravitational or electrostatic force, Cartesian coordinates may be unusually inappropriate. Such a problem demands the use of a coordinate system in which the radial distance is taken to be one of the coordinates, that is, spherical polar coordinates.

The point is that the coordinate system should be chosen to fit the problem, to exploit any constraint or symmetry present in it. Then it is likely to be more readily soluble than if we had forced it into a Cartesian framework.

Naturally, there is a price that must be paid for the use of a non-Cartesian coordinate system. We have not yet written expressions for gradient, divergence, or curl in any of the non-Cartesian coordinate systems. Such expressions are developed in general form in Section 2.2. First, we develop a system of curvilinear coordinates, a general system that may be specialized to any of the particular systems of interest. We shall specialize to circular cylindrical coordinates in Section 2.4 and to spherical polar coordinates in Section 2.5.

2.1 ORTHOGONAL COORDINATES IN $\mathbb{R}^3$

In Cartesian coordinates we deal with three mutually perpendicular families of planes: $x =$ constant, $y =$ constant, and $z =$ constant. Imagine that we superimpose on this system

three other families of surfaces $q_i(x, y, z),\ i = 1, 2, 3$. The surfaces of any one family q_i need not be parallel to each other and they need not be planes. If this is difficult to visualize, the figure of a specific coordinate system, such as Fig. 2.3, may be helpful. The three new families of surfaces need not be mutually perpendicular, but for simplicity we impose this condition (Eq. (2.7)) because orthogonal coordinates are common in physical applications. This orthogonality has many advantages: Orthogonal coordinates are almost like Cartesian coordinates where infinitesimal areas and volumes are products of coordinate differentials.

In this section we develop the general formalism of orthogonal coordinates, derive from the geometry the coordinate differentials, and use them for line, area, and volume elements in multiple integrals and vector operators. We may describe any point (x, y, z) as the intersection of three planes in Cartesian coordinates or as the intersection of the three surfaces that form our new, curvilinear coordinates. Describing the curvilinear coordinate surfaces by $q_1 =$ constant, $q_2 =$ constant, $q_3 =$ constant, we may identify our point by (q_1, q_2, q_3) as well as by (x, y, z):

$$\begin{array}{cc} \text{General curvilinear coordinates} & \text{Circular cylindrical coordinates} \\ q_1, q_2, q_3 & \rho, \varphi, z \\ x = x(q_1, q_2, q_3) & -\infty < x = \rho\cos\varphi < \infty \\ y = y(q_1, q_2, q_3) & -\infty < y = \rho\sin\varphi < \infty \\ z = z(q_1, q_2, q_3) & -\infty < z = z < \infty \end{array} \tag{2.1}$$

specifying x, y, z in terms of q_1, q_2, q_3 and the inverse relations

$$\begin{array}{ll} q_1 = q_1(x, y, z) & 0 \leqslant \rho = \left(x^2 + y^2\right)^{1/2} < \infty \\ q_2 = q_2(x, y, z) & 0 \leqslant \varphi = \arctan(y/x) < 2\pi \\ q_3 = q_3(x, y, z) & -\infty < z = z < \infty. \end{array} \tag{2.2}$$

As a specific illustration of the general, abstract q_1, q_2, q_3, the transformation equations for circular cylindrical coordinates (Section 2.4) are included in Eqs. (2.1) and (2.2). With each family of surfaces $q_i =$ constant, we can associate a unit vector $\hat{\mathbf{q}}_i$ normal to the surface $q_i =$ constant and in the direction of increasing q_i. In general, these unit vectors will depend on the position in space. Then a vector $\mathbf{V}$ may be written

$$\boxed{\mathbf{V} = \hat{\mathbf{q}}_1 V_1 + \hat{\mathbf{q}}_2 V_2 + \hat{\mathbf{q}}_3 V_3,} \tag{2.3}$$

but the coordinate or position vector is different in general,

$$\mathbf{r} \neq \hat{\mathbf{q}}_1 q_1 + \hat{\mathbf{q}}_2 q_2 + \hat{\mathbf{q}}_3 q_3,$$

as the special cases $\mathbf{r} = r\hat{\mathbf{r}}$ for spherical polar coordinates and $\mathbf{r} = \rho\hat{\boldsymbol{\rho}} + z\hat{z}$ for cylindrical coordinates demonstrate. The $\hat{\mathbf{q}}_i$ are normalized to $\hat{\mathbf{q}}_i^2 = 1$ and form a right-handed coordinate system with volume $\hat{\mathbf{q}}_1 \cdot (\hat{\mathbf{q}}_2 \times \hat{\mathbf{q}}_3) > 0$.

Differentiation of x in Eqs. (2.1) leads to the total variation or differential

$$dx = \frac{\partial x}{\partial q_1} dq_1 + \frac{\partial x}{\partial q_2} dq_2 + \frac{\partial x}{\partial q_3} dq_3, \tag{2.4}$$

and similarly for differentiation of y and z. In vector notation $d\mathbf{r} = \sum_i \frac{\partial \mathbf{r}}{\partial q_i} dq_i$. From the Pythagorean theorem in Cartesian coordinates the square of the distance between two neighboring points is

$$ds^2 = dx^2 + dy^2 + dz^2.$$

Substituting $d\mathbf{r}$ shows that in our curvilinear coordinate space the square of the distance element can be written as a quadratic form in the differentials dq_i:

$$\begin{aligned} ds^2 = d\mathbf{r}\cdot d\mathbf{r} = d\mathbf{r}^2 &= \sum_{ij} \frac{\partial \mathbf{r}}{\partial q_i}\cdot\frac{\partial \mathbf{r}}{\partial q_j}\,dq_i\,dq_j \\ &= g_{11}\,dq_1^2 + g_{12}\,dq_1\,dq_2 + g_{13}\,dq_1\,dq_3 \\ &\quad + g_{21}\,dq_2\,dq_1 + g_{22}\,dq_2^2 + g_{23}\,dq_2\,dq_3 \\ &\quad + g_{31}\,dq_3\,dq_1 + g_{32}\,dq_3\,dq_2 + g_{33}\,dq_3^2 \\ &= \sum_{ij} g_{ij}\,dq_i\,dq_j, \end{aligned} \tag{2.5}$$

where nonzero mixed terms $dq_i\,dq_j$ with $i \neq j$ signal that these coordinates are not orthogonal, that is, that the tangential directions $\hat{\mathbf{q}}_i$ are not mutually orthogonal. Spaces for which Eq. (2.5) is a legitimate expression are called *metric* or *Riemannian*.

Writing Eq. (2.5) more explicitly, we see that

$$g_{ij}(q_1,q_2,q_3) = \frac{\partial x}{\partial q_i}\frac{\partial x}{\partial q_j} + \frac{\partial y}{\partial q_i}\frac{\partial y}{\partial q_j} + \frac{\partial z}{\partial q_i}\frac{\partial z}{\partial q_j} = \frac{\partial \mathbf{r}}{\partial q_i}\cdot\frac{\partial \mathbf{r}}{\partial q_j} \tag{2.6}$$

are scalar products of the **tangent vectors** $\frac{\partial \mathbf{r}}{\partial q_i}$ to the curves $\mathbf{r}$ for $q_j = \text{const.}$, $j \neq i$. These coefficient functions g_{ij}, which we now proceed to investigate, may be viewed as specifying the nature of the coordinate system (q_1, q_2, q_3). Collectively these coefficients are referred to as the **metric** and in Section 2.10 will be shown to form a second-rank symmetric tensor.[1] In general relativity the metric components are determined by the properties of matter; that is, the g_{ij} are solutions of Einstein's field equations with the energy–momentum tensor as driving term; this may be articulated as "geometry is merged with physics."

At usual we limit ourselves to orthogonal (mutually perpendicular surfaces) coordinate systems, which means (see Exercise 2.1.1)[2]

$$g_{ij} = 0, \qquad i \neq j, \tag{2.7}$$

and $\hat{\mathbf{q}}_i \cdot \hat{\mathbf{q}}_j = \delta_{ij}$. (Nonorthogonal coordinate systems are considered in some detail in Sections 2.10 and 2.11 in the framework of tensor analysis.) Now, to simplify the notation, we write $g_{ii} = h_i^2 > 0$, so

$$\boxed{ds^2 = (h_1\,dq_1)^2 + (h_2\,dq_2)^2 + (h_3\,dq_3)^2 = \sum_i (h_i\,dq_i)^2.} \tag{2.8}$$

[1] The tensor nature of the set of g_{ij}'s follows from the quotient rule (Section 2.8). Then the tensor transformation law yields Eq. (2.5).

[2] In relativistic cosmology the nondiagonal elements of the metric g_{ij} are usually set equal to zero as a consequence of physical assumptions such as no rotation, as for $d\varphi\,dt, d\theta\,dt$.

The specific orthogonal coordinate systems are described in subsequent sections by specifying these (positive) scale factors h_1, h_2, and h_3. Conversely, the scale factors may be conveniently identified by the relation

$$ds_i = h_i\,dq_i, \qquad \frac{\partial \mathbf{r}}{\partial q_i} = h_i \hat{\mathbf{q}}_i \tag{2.9}$$

for any given dq_i, holding all other q constant. Here, ds_i is a differential length along the direction $\hat{\mathbf{q}}_i$. Note that the three curvilinear coordinates q_1, q_2, q_3 need not be lengths. The scale factors h_i may depend on q and they may have dimensions. The **product** $h_i\,dq_i$ must have a dimension of length. The differential distance vector $d\mathbf{r}$ may be written

$$d\mathbf{r} = h_1\,dq_1\,\hat{\mathbf{q}}_1 + h_2\,dq_2\,\hat{\mathbf{q}}_2 + h_3\,dq_3\,\hat{\mathbf{q}}_3 = \sum_i h_i\,dq_i\,\hat{\mathbf{q}}_i.$$

Using this curvilinear component form, we find that a line integral becomes

$$\boxed{\int \mathbf{V}\cdot d\mathbf{r} = \sum_i \int V_i h_i\,dq_i.}$$

From Eqs. (2.9) we may immediately develop the area and volume elements

$$d\sigma_{ij} = ds_i\,ds_j = h_i h_j\,dq_i\,dq_j \tag{2.10}$$

and

$$d\tau = ds_1\,ds_2\,ds_3 = h_1h_2h_3\,dq_1\,dq_2\,dq_3. \tag{2.11}$$

The expressions in Eqs. (2.10) and (2.11) agree, of course, with the results of using the transformation equations, Eq. (2.1), and Jacobians (described shortly; see also Exercise 2.1.5).

From Eq. (2.10) an area element may be expanded:

$$\begin{aligned} d\boldsymbol{\sigma} &= ds_2\,ds_3\,\hat{\mathbf{q}}_1 + ds_3\,ds_1\,\hat{\mathbf{q}}_2 + ds_1\,ds_2\,\hat{\mathbf{q}}_3 \\ &= h_2h_3\,dq_2\,dq_3\,\hat{\mathbf{q}}_1 + h_3h_1\,dq_3\,dq_1\,\hat{\mathbf{q}}_2 \\ &\quad + h_1h_2\,dq_1\,dq_2\,\hat{\mathbf{q}}_3. \end{aligned}$$

A surface integral becomes

$$\boxed{\begin{aligned} \int \mathbf{V}\cdot d\boldsymbol{\sigma} &= \int V_1h_2h_3\,dq_2\,dq_3 + \int V_2h_3h_1\,dq_3\,dq_1 \\ &\quad + \int V_3h_1h_2\,dq_1\,dq_2. \end{aligned}}$$

(Examples of such line and surface integrals appear in Sections 2.4 and 2.5.)

In anticipation of the new forms of equations for vector **calculus** that appear in the next section, let us emphasize that vector **algebra** is the same in orthogonal curvilinear coordinates as in Cartesian coordinates. Specifically, for the dot product,

$$\boxed{\begin{aligned}\mathbf{A}\cdot\mathbf{B} &= \sum_{ik} A_i\hat{\mathbf{q}}_i\cdot\hat{\mathbf{q}}_k B_k = \sum_{ik} A_i B_k\delta_{ik}\\ &= \sum_i A_i B_i = A_1B_1 + A_2B_2 + A_3B_3,\end{aligned}} \tag{2.12}$$

where the subscripts indicate curvilinear components. For the cross product,

$$\boxed{\mathbf{A}\times\mathbf{B} = \begin{vmatrix}\hat{\mathbf{q}}_1 & \hat{\mathbf{q}}_2 & \hat{\mathbf{q}}_3\\ A_1 & A_2 & A_3\\ B_1 & B_2 & B_3\end{vmatrix},} \tag{2.13}$$

as in Eq. (1.40).

Previously, we specialized to locally rectangular coordinates that are adapted to special symmetries. Let us now briefly look at the more general case, where the coordinates are not necessarily orthogonal. Surface and volume elements are part of multiple integrals, which are common in physical applications, such as center of mass determinations and moments of inertia. Typically, we choose coordinates according to the symmetry of the particular problem. In Chapter 1 we used Gauss' theorem to transform a volume integral into a surface integral and Stokes' theorem to transform a surface integral into a line integral. For orthogonal coordinates, the surface and volume elements are simply products of the line elements $h_i\,dq_i$ (see Eqs. (2.10) and (2.11)). For the general case, we use the geometric meaning of $\partial\mathbf{r}/\partial q_i$ in Eq. (2.5) as tangent vectors. We start with the Cartesian surface element $dx\,dy$, which becomes an infinitesimal rectangle in the new coordinates q_1, q_2 formed by the two incremental vectors

$$\begin{aligned}d\mathbf{r_1} &= \mathbf{r}(q_1+dq_1,q_2) - \mathbf{r}(q_1,q_2) = \frac{\partial\mathbf{r}}{\partial q_1}dq_1,\\ d\mathbf{r_2} &= \mathbf{r}(q_1,q_2+dq_2) - \mathbf{r}(q_1,q_2) = \frac{\partial\mathbf{r}}{\partial q_2}dq_2,\end{aligned} \tag{2.14}$$

whose area is the z-component of their cross product, or

$$\begin{aligned}dx\,dy = d\mathbf{r_1}\times d\mathbf{r_2}\big|_z &= \left[\frac{\partial x}{\partial q_1}\frac{\partial y}{\partial q_2} - \frac{\partial x}{\partial q_2}\frac{\partial y}{\partial q_1}\right]dq_1\,dq_2\\ &= \begin{vmatrix}\frac{\partial x}{\partial q_1} & \frac{\partial x}{\partial q_2}\\ \frac{\partial y}{\partial q_1} & \frac{\partial y}{\partial q_2}\end{vmatrix}dq_1\,dq_2.\end{aligned} \tag{2.15}$$

The transformation coefficient in determinant form is called the **Jacobian**.

Similarly, the volume element $dx\,dy\,dz$ becomes the triple scalar product of the three infinitesimal displacement vectors $d\mathbf{r_i} = dq_i\frac{\partial\mathbf{r}}{\partial q_i}$ along the q_i directions $\hat{\mathbf{q}}_\mathbf{i}$, which, according

to Section 1.5, takes on the form

$$dx\,dy\,dz = \begin{vmatrix} \frac{\partial x}{\partial q_1} & \frac{\partial x}{\partial q_2} & \frac{\partial x}{\partial q_3} \\ \frac{\partial y}{\partial q_1} & \frac{\partial y}{\partial q_2} & \frac{\partial y}{\partial q_3} \\ \frac{\partial z}{\partial q_1} & \frac{\partial z}{\partial q_2} & \frac{\partial z}{\partial q_3} \end{vmatrix} dq_1\,dq_2\,dq_3. \tag{2.16}$$

Here the determinant is also called the Jacobian, and so on in higher dimensions.

For orthogonal coordinates the Jacobians simplify to products of the orthogonal vectors in Eq. (2.9). It follows that they are just products of the h_i; for example, the volume Jacobian becomes

$$h_1h_2h_3(\hat{\mathbf{q}}_1 \times \hat{\mathbf{q}}_2) \cdot \hat{\mathbf{q}}_3 = h_1h_2h_3,$$

and so on.

Example 2.1.1 JACOBIANS FOR POLAR COORDINATES

Let us illustrate the transformation of the Cartesian two-dimensional volume element $dx\,dy$ to polar coordinates ρ, φ, with $x = \rho\cos\varphi,\ y = \rho\sin\varphi$. (See also Section 2.4.) Here,

$$dxdy = \begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \varphi} \end{vmatrix} d\rho\,d\varphi = \begin{vmatrix} \cos\varphi & -\rho\sin\varphi \\ \sin\varphi & \rho\cos\varphi \end{vmatrix} d\rho\,d\varphi = \rho\,d\rho\,d\varphi.$$

Similarly, in spherical coordinates (see Section 2.5) we get, from $x = r\sin\theta\cos\varphi$, $y = r\sin\theta\sin\varphi$, $z = r\cos\theta$, the Jacobian

$$\begin{aligned} J &= \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \varphi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \varphi} \end{vmatrix} = \begin{vmatrix} \sin\theta\cos\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\theta & -r\sin\theta & 0 \end{vmatrix} \\ &= \cos\theta \begin{vmatrix} r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \end{vmatrix} + r\sin\theta \begin{vmatrix} \sin\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & r\sin\theta\cos\varphi \end{vmatrix} \\ &= r^2\left(\cos^2\theta\sin\theta + \sin^3\theta\right) = r^2\sin\theta \end{aligned}$$

by expanding the determinant along the third line. Hence the volume element becomes $dx\,dy\,dz = r^2\,dr\sin\theta\,d\theta\,d\varphi$. The volume integral can be written as

$$\int f(x, y, z)\,dx\,dy\,dz = \int f\big(x(r,\theta,\varphi), y(r,\theta,\varphi), z(r,\theta,\varphi)\big)r^2\,dr\sin\theta\,d\theta\,d\varphi. \qquad \blacksquare$$

In summary, we have developed the general formalism for vector analysis in orthogonal curvilinear coordinates in $\mathbb{R}^3$. For most applications, locally orthogonal coordinates can be chosen for which surface and volume elements in multiple integrals are products of line elements. For the general nonorthogonal case, Jacobian determinants apply.

Exercises

2.1.1 Show that limiting our attention to orthogonal coordinate systems implies that $g_{ij} = 0$ for $i \neq j$ (Eq. (2.7)).
Hint. Construct a triangle with sides ds_1, ds_2, and ds_2. Equation (2.9) must hold regardless of whether $g_{ij} = 0$. Then compare ds^2 from Eq. (2.5) with a calculation using the law of cosines. Show that $\cos\theta_{12} = g_{12}/\sqrt{g_{11}g_{22}}$.

2.1.2 In the spherical polar coordinate system, $q_1 = r$, $q_2 = \theta$, $q_3 = \varphi$. The transformation equations corresponding to Eq. (2.1) are

$$x = r\sin\theta\cos\varphi, \qquad y = r\sin\theta\sin\varphi, \qquad z = r\cos\theta.$$

(a) Calculate the spherical polar coordinate scale factors: h_r, h_θ, and h_φ.
(b) Check your calculated scale factors by the relation $ds_i = h_i\, dq_i$.

2.1.3 The u-, v-, z-coordinate system frequently used in electrostatics and in hydrodynamics is defined by

$$xy = u, \qquad x^2 - y^2 = v, \qquad z = z.$$

This u-, v-, z-system is orthogonal.

(a) In words, describe briefly the nature of each of the three families of coordinate surfaces.
(b) Sketch the system in the xy-plane showing the intersections of surfaces of constant u and surfaces of constant v with the xy-plane.
(c) Indicate the directions of the unit vector $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ in all four quadrants.
(d) Finally, is this u-, v-, z-system right-handed ($\hat{\mathbf{u}} \times \hat{\mathbf{v}} = +\hat{\mathbf{z}}$) or left-handed ($\hat{\mathbf{u}} \times \hat{\mathbf{v}} = -\hat{\mathbf{z}}$)?

2.1.4 The elliptic cylindrical coordinate system consists of three families of surfaces:

$$1)\ \frac{x^2}{a^2\cosh^2 u} + \frac{y^2}{a^2\sinh^2 u} = 1; \qquad 2)\ \frac{x^2}{a^2\cos^2 v} - \frac{y^2}{a^2\sin^2 v} = 1; \qquad 3)\ z = z.$$

Sketch the coordinate surfaces $u =$ constant and $v =$ constant as they intersect the first quadrant of the xy-plane. Show the unit vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$. The range of u is $0 \leqslant u < \infty$. The range of v is $0 \leqslant v \leqslant 2\pi$.

2.1.5 A two-dimensional orthogonal system is described by the coordinates q_1 and q_2. Show that the Jacobian

$$J\left(\frac{x, y}{q_1, q_2}\right) \equiv \frac{\partial(x, y)}{\partial(q_1, q_2)} \equiv \frac{\partial x}{\partial q_1}\frac{\partial y}{\partial q_2} - \frac{\partial x}{\partial q_2}\frac{\partial y}{\partial q_1} = h_1 h_2$$

is in agreement with Eq. (2.10).
Hint. It's easier to work with the square of each side of this equation.

2.1.6 In Minkowski space we define $x_1 = x$, $x_2 = y$, $x_3 = z$, and $x_0 = ct$. This is done so that the metric interval becomes $ds^2 = dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2$ (with $c =$ velocity of light). Show that the metric in Minkowski space is

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

We use Minkowski space in Sections 4.5 and 4.6 for describing Lorentz transformations.

2.2 DIFFERENTIAL VECTOR OPERATORS

We return to our restriction to orthogonal coordinate systems.

Gradient

The starting point for developing the gradient, divergence, and curl operators in curvilinear coordinates is the geometric interpretation of the gradient as the vector having the magnitude and direction of the maximum space rate of change (compare Section 1.6). From this interpretation the component of $\nabla\psi(q_1, q_2, q_3)$ in the direction normal to the family of surfaces $q_1 =$ constant is given by[3]

$$\hat{\mathbf{q}}_1 \cdot \nabla\psi = \nabla\psi|_1 = \frac{\partial\psi}{\partial s_1} = \frac{1}{h_1}\frac{\partial\psi}{\partial q_1}, \tag{2.17}$$

since this is the rate of change of ψ for varying q_1, holding q_2 and q_3 fixed. The quantity ds_1 is a differential length in the direction of increasing q_1 (compare Eqs. (2.9)). In Section 2.1 we introduced a unit vector $\hat{\mathbf{q}}_1$ to indicate this direction. By repeating Eq. (2.17) for q_2 and again for q_3 and adding vectorially, we see that the gradient becomes

$$\begin{aligned} \nabla\psi(q_1, q_2, q_3) &= \hat{\mathbf{q}}_1\frac{\partial\psi}{\partial s_1} + \hat{\mathbf{q}}_2\frac{\partial\psi}{\partial s_2} + \hat{\mathbf{q}}_3\frac{\partial\psi}{\partial s_3} \\ &= \hat{\mathbf{q}}_1\frac{1}{h_1}\frac{\partial\psi}{\partial q_1} + \hat{\mathbf{q}}_2\frac{1}{h_2}\frac{\partial\psi}{\partial q_2} + \hat{\mathbf{q}}_3\frac{1}{h_3}\frac{\partial\psi}{\partial q_3} \\ &= \sum_i \hat{\mathbf{q}}_i\frac{1}{h_i}\frac{\partial\psi}{\partial q_i}. \end{aligned} \tag{2.18}$$

Exercise 2.2.4 offers a mathematical alternative independent of this physical interpretation of the gradient. The total variation of a function,

$$d\psi = \nabla\psi \cdot d\mathbf{r} = \sum_i \frac{1}{h_i}\frac{\partial\psi}{\partial q_i} ds_i = \sum_i \frac{\partial\psi}{\partial q_i} dq_i$$

is consistent with Eq. (2.18), of course.

[3] Here the use of φ to label a function is avoided because it is conventional to use this symbol to denote an azimuthal coordinate.

Divergence

The divergence operator may be obtained from the second definition (Eq. (1.98)) of Chapter 1 or equivalently from Gauss' theorem, Section 1.11. Let us use Eq. (1.98),

$$\nabla \cdot \mathbf{V}(q_1, q_2, q_3) = \lim_{\int d\tau \to 0} \frac{\int \mathbf{V} \cdot d\boldsymbol{\sigma}}{\int d\tau}, \tag{2.19}$$

with a differential volume $h_1 h_2 h_3\, dq_1\, dq_2\, dq_3$ (Fig. 2.1). Note that the positive directions have been chosen so that $(\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \hat{\mathbf{q}}_3)$ form a right-handed set, $\hat{\mathbf{q}}_1 \times \hat{\mathbf{q}}_2 = \hat{\mathbf{q}}_3$.

The difference of area integrals for the two faces $q_1 =$ constant is given by

$$\left[V_1 h_2 h_3 + \frac{\partial}{\partial q_1}(V_1 h_2 h_3)\, dq_1\right] dq_2\, dq_3 - V_1 h_2 h_3\, dq_2\, dq_3$$
$$= \frac{\partial}{\partial q_1}(V_1 h_2 h_3)\, dq_1\, dq_2\, dq_3, \tag{2.20}$$

exactly as in Sections 1.7 and 1.10.[4] Here, $V_i = \mathbf{V} \cdot \hat{\mathbf{q}}_i$ is the projection of $\mathbf{V}$ onto the $\hat{\mathbf{q}}_i$-direction. Adding in the similar results for the other two pairs of surfaces, we obtain

$$\int \mathbf{V}(q_1, q_2, q_3) \cdot d\boldsymbol{\sigma}$$
$$= \left[\frac{\partial}{\partial q_1}(V_1 h_2 h_3) + \frac{\partial}{\partial q_2}(V_2 h_3 h_1) + \frac{\partial}{\partial q_3}(V_3 h_1 h_2)\right] dq_1\, dq_2\, dq_3.$$

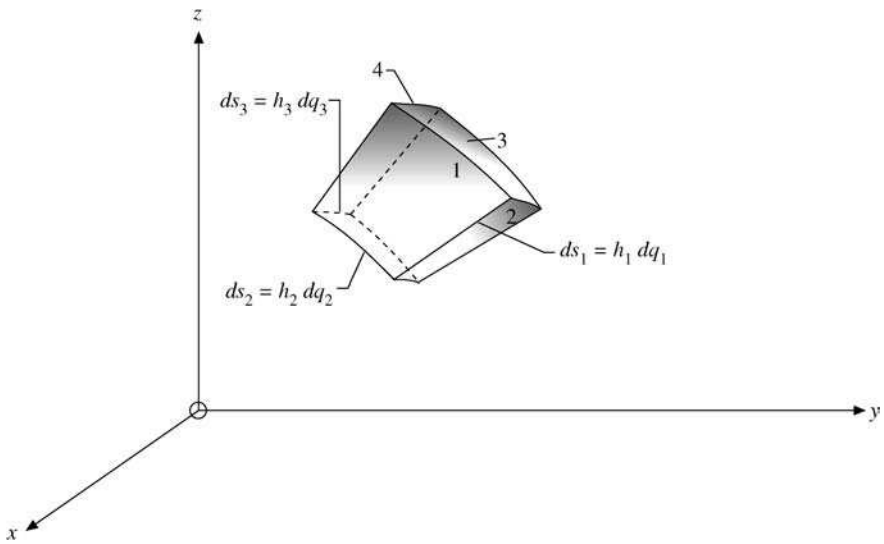

FIGURE 2.1 Curvilinear volume element.

[4] Since we take the limit $dq_1, dq_2, dq_3 \to 0$, the second- and higher-order derivatives will drop out.

Now, using Eq. (2.19), division by our differential volume yields

$$\boxed{\nabla \cdot \mathbf{V}(q_1, q_2, q_3) = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}(V_1 h_2 h_3) + \frac{\partial}{\partial q_2}(V_2 h_3 h_1) + \frac{\partial}{\partial q_3}(V_3 h_1 h_2)\right].} \tag{2.21}$$

We may obtain the Laplacian by combining Eqs. (2.18) and (2.21), using $\mathbf{V} = \nabla\psi(q_1, q_2, q_3)$. This leads to

$$\boxed{\begin{aligned}&\nabla \cdot \nabla\psi(q_1, q_2, q_3)\\ &\quad = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial\psi}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_3 h_1}{h_2}\frac{\partial\psi}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial\psi}{\partial q_3}\right)\right].\end{aligned}} \tag{2.22}$$

Curl

Finally, to develop $\nabla \times \mathbf{V}$, let us apply Stokes' theorem (Section 1.12) and, as with the divergence, take the limit as the surface area becomes vanishingly small. Working on one component at a time, we consider a differential surface element in the curvilinear surface $q_1 =$ constant. From

$$\int_s \nabla \times \mathbf{V} \cdot d\boldsymbol{\sigma} = \hat{\mathbf{q}}_1 \cdot (\nabla \times \mathbf{V}) h_2 h_3\, dq_2\, dq_3 \tag{2.23}$$

(mean value theorem of integral calculus), Stokes' theorem yields

$$\hat{\mathbf{q}}_1 \cdot (\nabla \times \mathbf{V}) h_2 h_3\, dq_2\, dq_3 = \oint \mathbf{V} \cdot d\mathbf{r}, \tag{2.24}$$

with the line integral lying in the surface $q_1 =$ constant. Following the loop (1, 2, 3, 4) of Fig. 2.2,

$$\begin{aligned}\oint \mathbf{V}(q_1, q_2, q_3) \cdot d\mathbf{r} &= V_2 h_2\, dq_2 + \left[V_3 h_3 + \frac{\partial}{\partial q_2}(V_3 h_3)\, dq_2\right] dq_3\\ &\quad - \left[V_2 h_2 + \frac{\partial}{\partial q_3}(V_2 h_2) dq_3\right] dq_2 - V_3 h_3\, dq_3\\ &= \left[\frac{\partial}{\partial q_2}(h_3 V_3) - \frac{\partial}{\partial q_3}(h_2 V_2)\right] dq_2\, dq_3.\end{aligned} \tag{2.25}$$

We pick up a positive sign when going in the positive direction on parts 1 and 2 and a negative sign on parts 3 and 4 because here we are going in the negative direction. (Higher-order terms in Maclaurin or Taylor expansions have been omitted. They will vanish in the limit as the surface becomes vanishingly small ($dq_2 \to 0$, $dq_3 \to 0$).)

From Eq. (2.24),

$$\nabla \times \mathbf{V}|_1 = \frac{1}{h_2 h_3}\left[\frac{\partial}{\partial q_2}(h_3 V_3) - \frac{\partial}{\partial q_3}(h_2 V_2)\right]. \tag{2.26}$$

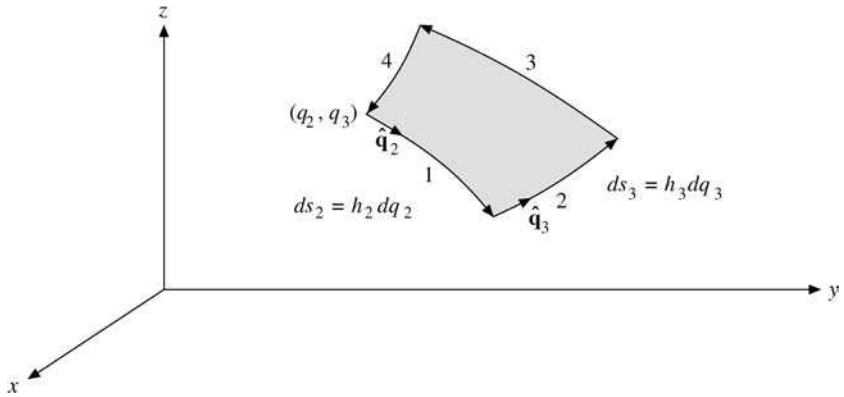

FIGURE 2.2 Curvilinear surface element with $q_1 =$ constant.

The remaining two components of $\nabla \times \mathbf{V}$ may be picked up by cyclic permutation of the indices. As in Chapter 1, it is often convenient to write the curl in determinant form:

$$\nabla \times \mathbf{V} = \frac{1}{h_1 h_2 h_3} \begin{vmatrix} \hat{\mathbf{q}}_1 h_1 & \hat{\mathbf{q}}_2 h_2 & \hat{\mathbf{q}}_3 h_3 \\ \dfrac{\partial}{\partial q_1} & \dfrac{\partial}{\partial q_2} & \dfrac{\partial}{\partial q_3} \\ h_1 V_1 & h_2 V_2 & h_3 V_3 \end{vmatrix}. \tag{2.27}$$

Remember that, because of the presence of the differential operators, this determinant must be expanded from the top down. Note that this equation is **not** identical with the form for the cross product of two vectors, Eq. (2.13). ∇ is not an ordinary vector; it is a vector **operator**.

Our geometric interpretation of the gradient and the use of Gauss' and Stokes' theorems (or integral definitions of divergence and curl) have enabled us to obtain these quantities **without having to differentiate the unit vectors $\hat{\mathbf{q}}_i$**. There exist alternate ways to determine grad, div, and curl based on direct differentiation of the $\hat{\mathbf{q}}_i$. One approach resolves the $\hat{\mathbf{q}}_i$ of a specific coordinate system into its Cartesian components (Exercises 2.4.1 and 2.5.1) and differentiates this Cartesian form (Exercises 2.4.3 and 2.5.2). The point here is that the derivatives of the Cartesian $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ vanish since $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are constant in direction as well as in magnitude. A second approach [L. J. Kijewski, *Am. J. Phys.* **33**: 816 (1965)] assumes the equality of $\partial^2\mathbf{r}/\partial q_i\,\partial q_j$ and $\partial^2\mathbf{r}/\partial q_j\,\partial q_i$ and develops the derivatives of $\hat{\mathbf{q}}_i$ in a general curvilinear form. Exercises 2.2.3 and 2.2.4 are based on this method.

Exercises

2.2.1 Develop arguments to show that dot and cross products (not involving ∇) in orthogonal curvilinear coordinates in $\mathbb{R}^3$ proceed, as in Cartesian coordinates, **with no involvement of scale factors**.

2.2.2 With $\hat{\mathbf{q}}_1$ a unit vector in the direction of increasing q_1, show that

(a) $$\nabla \cdot \hat{\mathbf{q}}_1 = \frac{1}{h_1 h_2 h_3} \frac{\partial (h_2 h_3)}{\partial q_1}$$

(b) $$\nabla \times \hat{\mathbf{q}}_1 = \frac{1}{h_1} \left[\hat{\mathbf{q}}_2 \frac{1}{h_3} \frac{\partial h_1}{\partial q_3} - \hat{\mathbf{q}}_3 \frac{1}{h_2} \frac{\partial h_1}{\partial q_2} \right].$$

Note that even though $\hat{\mathbf{q}}_1$ is a unit vector, its divergence and curl **do not necessarily vanish**.

2.2.3 Show that the orthogonal unit vectors $\hat{\mathbf{q}}_j$ may be defined by

$$\hat{\mathbf{q}}_i = \frac{1}{h_i} \frac{\partial \mathbf{r}}{\partial q_i}. \tag{a}$$

In particular, show that $\hat{\mathbf{q}}_i \cdot \hat{\mathbf{q}}_i = 1$ leads to an expression for h_i in agreement with Eqs. (2.9).
Equation (a) may be taken as a starting point for deriving

$$\frac{\partial \hat{\mathbf{q}}_i}{\partial q_j} = \hat{\mathbf{q}}_j \frac{1}{h_i} \frac{\partial h_j}{\partial q_i}, \qquad i \neq j$$

and

$$\frac{\partial \hat{\mathbf{q}}_i}{\partial q_i} = -\sum_{j \neq i} \hat{\mathbf{q}}_j \frac{1}{h_j} \frac{\partial h_i}{\partial q_j}.$$

2.2.4 Derive

$$\nabla \psi = \hat{\mathbf{q}}_1 \frac{1}{h_1} \frac{\partial \psi}{\partial q_1} + \hat{\mathbf{q}}_2 \frac{1}{h_2} \frac{\partial \psi}{\partial q_2} + \hat{\mathbf{q}}_3 \frac{1}{h_3} \frac{\partial \psi}{\partial q_3}$$

by direct application of Eq. (1.97),

$$\nabla \psi = \lim_{\int d\tau \to 0} \frac{\int \psi \, d\boldsymbol{\sigma}}{\int d\tau}.$$

Hint. Evaluation of the surface integral will lead to terms like $(h_1 h_2 h_3)^{-1} (\partial / \partial q_1) \times (\hat{\mathbf{q}}_1 h_2 h_3)$. The results listed in Exercise 2.2.3 will be helpful. Cancellation of unwanted terms occurs when the contributions of all three pairs of surfaces are added together.

2.3 SPECIAL COORDINATE SYSTEMS: INTRODUCTION

There are at least 11 coordinate systems in which the three-dimensional Helmholtz equation can be separated into three ordinary differential equations. Some of these coordinate systems have achieved prominence in the historical development of quantum mechanics. Other systems, such as bipolar coordinates, satisfy special needs. Partly because the needs are rather infrequent but mostly because the development of computers and efficient programming techniques reduce the need for these coordinate systems, the discussion in this chapter is limited to (1) Cartesian coordinates, (2) spherical polar coordinates, and (3) circular cylindrical coordinates. Specifications and details of the other coordinate systems will be found in the first two editions of this work and in Additional Readings at the end of this chapter (Morse and Feshbach, Margenau and Murphy).

2.4 CIRCULAR CYLINDER COORDINATES

In the circular cylindrical coordinate system the three curvilinear coordinates (q_1, q_2, q_3) are relabeled (ρ, φ, z). We are using ρ for the perpendicular distance from the z-axis and saving r for the distance from the origin. The limits on ρ, φ and z are

$$0 \leqslant \rho < \infty, \qquad 0 \leqslant \varphi \leqslant 2\pi, \qquad \text{and} \qquad -\infty < z < \infty.$$

For $\rho = 0$, φ is not well defined. The coordinate surfaces, shown in Fig. 2.3, are:

1. Right circular cylinders having the z-axis as a common axis,
$$\rho = \left(x^2 + y^2\right)^{1/2} = \text{constant}.$$
2. Half-planes through the z-axis,
$$\varphi = \tan^{-1}\left(\frac{y}{x}\right) = \text{constant}.$$
3. Planes parallel to the xy-plane, as in the Cartesian system,
$$z = \text{constant}.$$

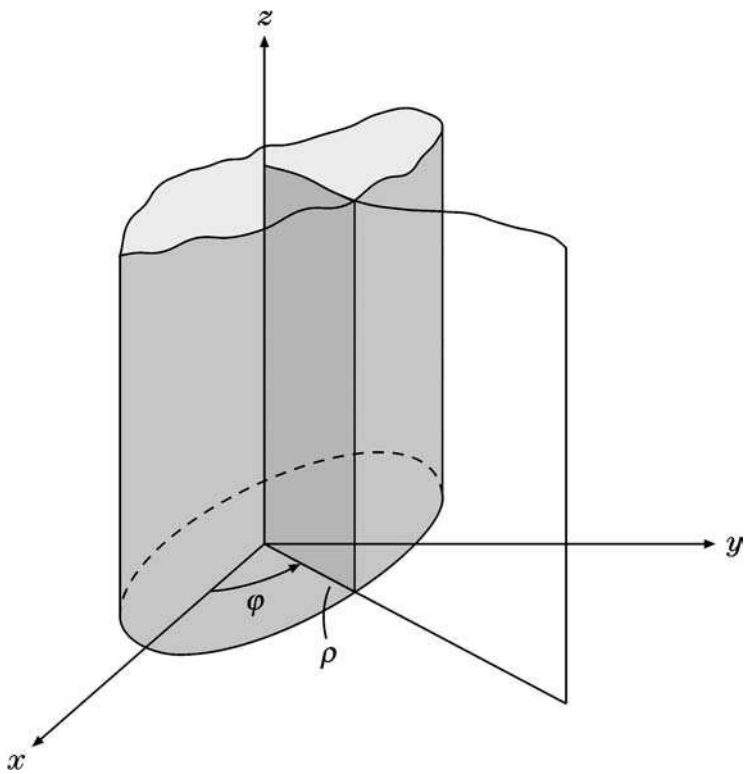

FIGURE 2.3 Circular cylinder coordinates.

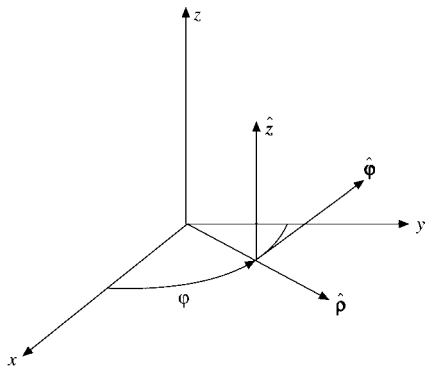

FIGURE 2.4 Circular cylindrical coordinate unit vectors.

Inverting the preceding equations for ρ and φ (or going directly to Fig. 2.3), we obtain the transformation relations

$$x = \rho\cos\varphi, \qquad y = \rho\sin\varphi, \qquad z = z. \tag{2.28}$$

The z-axis remains unchanged. This is essentially a two-dimensional curvilinear system with a Cartesian z-axis added on to form a three-dimensional system.

According to Eq. (2.5) or from the length elements ds_i, the scale factors are

$$h_1 = h_\rho = 1, \qquad h_2 = h_\varphi = \rho, \qquad h_3 = h_z = 1. \tag{2.29}$$

The unit vectors $\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \hat{\mathbf{q}}_3$ are relabeled $(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\varphi}}, \hat{\mathbf{z}})$, as in Fig. 2.4. The unit vector $\hat{\boldsymbol{\rho}}$ is normal to the cylindrical surface, pointing in the direction of increasing radius ρ. The unit vector $\hat{\boldsymbol{\varphi}}$ is tangential to the cylindrical surface, perpendicular to the half plane $\varphi =$ constant and pointing in the direction of increasing azimuth angle φ. The third unit vector, $\hat{\mathbf{z}}$, is the usual Cartesian unit vector. They are mutually orthogonal,

$$\hat{\boldsymbol{\rho}} \cdot \hat{\boldsymbol{\varphi}} = \hat{\boldsymbol{\varphi}} \cdot \hat{\mathbf{z}} = \hat{\mathbf{z}} \cdot \hat{\boldsymbol{\rho}} = 0,$$

and the coordinate vector and a general vector $\mathbf{V}$ are expressed as

$$\mathbf{r} = \hat{\boldsymbol{\rho}}\rho + \hat{\mathbf{z}}z, \qquad \mathbf{V} = \hat{\boldsymbol{\rho}}V_\rho + \hat{\boldsymbol{\varphi}}V_\varphi + \hat{\mathbf{z}}V_z.$$

A differential displacement $d\mathbf{r}$ may be written

$$\begin{aligned} d\mathbf{r} &= \hat{\boldsymbol{\rho}}\,ds_\rho + \hat{\boldsymbol{\varphi}}\,ds_\varphi + \hat{\mathbf{z}}\,dz \\ &= \hat{\boldsymbol{\rho}}\,d\rho + \hat{\boldsymbol{\varphi}}\rho\,d\varphi + \hat{\mathbf{z}}\,dz. \end{aligned} \tag{2.30}$$

Example 2.4.1 AREA LAW FOR PLANETARY MOTION

First we derive Kepler's law in cylindrical coordinates, saying that the radius vector sweeps out equal areas in equal time, from angular momentum conservation.

We consider the sun at the origin as a source of the **central** gravitational force $\mathbf{F} = f(r)\hat{\mathbf{r}}$. Then the orbital angular momentum $\mathbf{L} = m\mathbf{r} \times \mathbf{v}$ of a planet of mass m and velocity $\mathbf{v}$ is conserved, because the torque

$$\frac{d\mathbf{L}}{dt} = m\frac{d\mathbf{r}}{dt} \times \frac{d\mathbf{r}}{dt} + \mathbf{r} \times m\frac{d\mathbf{v}}{dt} = \mathbf{r} \times \mathbf{F} = \frac{f(r)}{r}\mathbf{r} \times \mathbf{r} = 0.$$

Hence $\mathbf{L} = \text{const}$. Now we can choose the z-axis to lie along the direction of the orbital angular momentum vector, $\mathbf{L} = L\hat{\mathbf{z}}$, and work in cylindrical coordinates $\mathbf{r} = (\rho, \varphi, z) = \rho\hat{\boldsymbol{\rho}}$ with $z = 0$. The planet moves in the xy-plane because $\mathbf{r}$ and $\mathbf{v}$ are perpendicular to $\mathbf{L}$. Thus, we expand its velocity as follows:

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \dot{\rho}\hat{\boldsymbol{\rho}} + \rho\frac{d\hat{\boldsymbol{\rho}}}{dt}.$$

From

$$\hat{\boldsymbol{\rho}} = (\cos\varphi, \sin\varphi), \qquad \frac{\partial\hat{\boldsymbol{\rho}}}{d\varphi} = (-\sin\varphi, \cos\varphi) = \hat{\boldsymbol{\varphi}},$$

we find that $\frac{d\hat{\boldsymbol{\rho}}}{dt} = \frac{d\hat{\boldsymbol{\rho}}}{d\varphi}\frac{d\varphi}{dt} = \dot{\varphi}\hat{\boldsymbol{\varphi}}$ using the chain rule, so $\mathbf{v} = \dot{\rho}\hat{\rho} + \rho\frac{d\hat{\rho}}{dt} = \dot{\rho}\hat{\boldsymbol{\rho}} + \rho\dot{\varphi}\hat{\boldsymbol{\varphi}}$. When we substitute the expansions of $\hat{\boldsymbol{\rho}}$ and $\mathbf{v}$ in polar coordinates, we obtain

$$\mathbf{L} = m\boldsymbol{\rho} \times \mathbf{v} = m\rho(\rho\dot{\varphi})(\hat{\boldsymbol{\rho}} \times \hat{\boldsymbol{\varphi}}) = m\rho^2\dot{\varphi}\hat{\mathbf{z}} = \text{constant}.$$

The triangular area swept by the radius vector $\boldsymbol{\rho}$ in the time dt (**area law**), when integrated over one revolution, is given by

$$A = \frac{1}{2}\int \rho(\rho\, d\varphi) = \frac{1}{2}\int \rho^2\dot{\varphi}\, dt = \frac{L}{2m}\int dt = \frac{L\tau}{2m}, \tag{2.31}$$

if we substitute $m\rho^2\dot{\varphi} = L = \text{const}$. Here τ is the period, that is, the time for one revolution of the planet in its orbit.

Kepler's first law says that the orbit is an ellipse. Now we derive the orbit equation $\rho(\varphi)$ of the ellipse in polar coordinates, where in Fig. 2.5 the sun is at one focus, which is the origin of our cylindrical coordinates. From the geometrical construction of the ellipse we know that $\rho' + \rho = 2a$, where a is the major half-axis; we shall show that this is equivalent to the conventional form of the ellipse equation. The distance between both foci is $0 < 2a\epsilon < 2a$, where $0 < \epsilon < 1$ is called the eccentricity of the ellipse. For a circle $\epsilon = 0$ because both foci coincide with the center. There is an angle, as shown in Fig. 2.5, where the distances $\rho' = \rho = a$ are equal, and Pythagoras' theorem applied to this right triangle

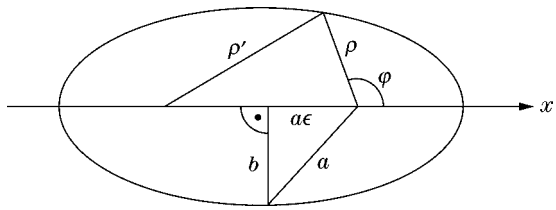

FIGURE 2.5 Ellipse in polar coordinates.

gives $b^2 + a^2\epsilon^2 = a^2$. As a result, $\sqrt{1-\epsilon^2} = b/a$ is the ratio of the minor half-axis (b) to the major half-axis, a.

Now consider the triangle with the sides labeled by ρ', ρ, $2a\epsilon$ in Fig. 2.5 and angle opposite ρ' equal to $\pi - \varphi$. Then, applying the law of cosines, gives

$$\rho'^2 = \rho^2 + 4a^2\epsilon^2 + 4\rho a\epsilon\cos\varphi.$$

Now substituting $\rho' = 2a - \rho$, canceling ρ^2 on both sides and dividing by $4a$ yields

$$\rho(1 + \epsilon\cos\varphi) = a\left(1 - \epsilon^2\right) \equiv p, \tag{2.32}$$

the **Kepler orbit equation in polar coordinates**.

Alternatively, we revert to Cartesian coordinates to find, from Eq. (2.32) with $x = \rho\cos\varphi$, that

$$\rho^2 = x^2 + y^2 = (p - x\epsilon)^2 = p^2 + x^2\epsilon^2 - 2px\epsilon,$$

so the familiar ellipse equation in Cartesian coordinates,

$$\left(1-\epsilon^2\right)\left(x + \frac{p\epsilon}{1-\epsilon^2}\right)^2 + y^2 = p^2 + \frac{p^2\epsilon^2}{1-\epsilon^2} = \frac{p^2}{1-\epsilon^2},$$

obtains. If we compare this result with the standard form of the ellipse,

$$\frac{(x-x_0)^2}{a^2} + \frac{y^2}{b^2} = 1,$$

we confirm that

$$b = \frac{p}{\sqrt{1-\epsilon^2}} = a\sqrt{1-\epsilon^2}, \qquad a = \frac{p}{1-\epsilon^2},$$

and that the distance x_0 between the center and focus is $a\epsilon$, as shown in Fig. 2.5. ■

The differential operations involving $\boldsymbol{\nabla}$ follow from Eqs. (2.18), (2.21), (2.22), and (2.27):

$$\boldsymbol{\nabla}\psi(\rho,\varphi,z) = \hat{\boldsymbol{\rho}}\frac{\partial\psi}{\partial\rho} + \hat{\boldsymbol{\varphi}}\frac{1}{\rho}\frac{\partial\psi}{\partial\varphi} + \hat{\mathbf{z}}\frac{\partial\psi}{\partial z}, \tag{2.33}$$

$$\boldsymbol{\nabla}\cdot\mathbf{V} = \frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho V_\rho) + \frac{1}{\rho}\frac{\partial V_\varphi}{\partial\varphi} + \frac{\partial V_z}{\partial z}, \tag{2.34}$$

$$\boldsymbol{\nabla}^2\psi = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\psi}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2\psi}{\partial\varphi^2} + \frac{\partial^2\psi}{\partial z^2}, \tag{2.35}$$

$$\boldsymbol{\nabla}\times\mathbf{V} = \frac{1}{\rho}\begin{vmatrix} \hat{\boldsymbol{\rho}} & \rho\hat{\boldsymbol{\varphi}} & \hat{\mathbf{z}} \\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\varphi} & \dfrac{\partial}{\partial z} \\ V_\rho & \rho V_\varphi & V_z \end{vmatrix}. \tag{2.36}$$

Finally, for problems such as circular wave guides and cylindrical cavity resonators the vector Laplacian $\nabla^2\mathbf{V}$ resolved in circular cylindrical coordinates is

$$\begin{aligned}\nabla^2\mathbf{V}|_\rho &= \nabla^2 V_\rho - \frac{1}{\rho^2}V_\rho - \frac{2}{\rho^2}\frac{\partial V_\varphi}{\partial\varphi},\\ \nabla^2\mathbf{V}|_\varphi &= \nabla^2 V_\varphi - \frac{1}{\rho^2}V_\varphi + \frac{2}{\rho^2}\frac{\partial V_\rho}{\partial\varphi},\\ \nabla^2\mathbf{V}|_z &= \nabla^2 V_z,\end{aligned} \tag{2.37}$$

which follow from Eq. (1.85). The basic reason for this particular form of the z-component is that the z-axis is a Cartesian axis; that is,

$$\begin{aligned}\nabla^2(\hat{\boldsymbol{\rho}}V_\rho + \hat{\boldsymbol{\varphi}}V_\varphi + \hat{\mathbf{z}}V_z) &= \nabla^2(\hat{\boldsymbol{\rho}}V_\rho + \hat{\boldsymbol{\varphi}}V_\varphi) + \hat{\mathbf{z}}\nabla^2 V_z\\ &= \hat{\boldsymbol{\rho}}f(V_\rho, V_\varphi) + \hat{\boldsymbol{\varphi}}g(V_\rho, V_\varphi) + \hat{\mathbf{z}}\nabla^2 V_z.\end{aligned}$$

Finally, the operator ∇^2 operating on the $\hat{\boldsymbol{\rho}}$, $\hat{\boldsymbol{\varphi}}$ unit vectors stays in the $\hat{\boldsymbol{\rho}}\hat{\boldsymbol{\varphi}}$-plane.

Example 2.4.2 A Navier–Stokes Term

The Navier–Stokes equations of hydrodynamics contain a nonlinear term

$$\nabla\times\big[\mathbf{v}\times(\nabla\times\mathbf{v})\big],$$

where $\mathbf{v}$ is the fluid velocity. For fluid flowing through a cylindrical pipe in the z-direction,

$$\mathbf{v} = \hat{\mathbf{z}}v(\rho).$$

From Eq. (2.36),

$$\nabla\times\mathbf{v} = \frac{1}{\rho}\begin{vmatrix}\hat{\boldsymbol{\rho}} & \rho\hat{\boldsymbol{\varphi}} & \hat{\mathbf{z}}\\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\varphi} & \dfrac{\partial}{\partial z}\\ 0 & 0 & v(\rho)\end{vmatrix} = -\hat{\boldsymbol{\varphi}}\frac{\partial v}{\partial\rho}$$

$$\mathbf{v}\times(\nabla\times\mathbf{v}) = \begin{vmatrix}\hat{\boldsymbol{\rho}} & \hat{\boldsymbol{\varphi}} & \hat{\mathbf{z}}\\ 0 & 0 & v\\ 0 & -\dfrac{\partial v}{\partial\rho} & 0\end{vmatrix} = \hat{\boldsymbol{\rho}}v(\rho)\frac{\partial v}{\partial\rho}.$$

Finally,

$$\nabla\times\big(\mathbf{v}\times(\nabla\times\mathbf{v})\big) = \frac{1}{\rho}\begin{vmatrix}\hat{\boldsymbol{\rho}} & \rho\hat{\boldsymbol{\varphi}} & \hat{\mathbf{z}}\\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\varphi} & \dfrac{\partial}{\partial z}\\ v\dfrac{\partial v}{\partial\rho} & 0 & 0\end{vmatrix} = 0,$$

so, for this particular case, the nonlinear term vanishes. ■

Exercises

2.4.1 Resolve the circular cylindrical unit vectors into their Cartesian components (Fig. 2.6).

ANS. $\hat{\boldsymbol{\rho}} = \hat{\mathbf{x}}\cos\varphi + \hat{\mathbf{y}}\sin\varphi,$
$\hat{\boldsymbol{\varphi}} = -\hat{\mathbf{x}}\sin\varphi + \hat{\mathbf{y}}\cos\varphi,$
$\hat{\mathbf{z}} = \hat{\mathbf{z}}.$

2.4.2 Resolve the Cartesian unit vectors into their circular cylindrical components (Fig. 2.6).

ANS. $\hat{\mathbf{x}} = \hat{\boldsymbol{\rho}}\cos\varphi - \hat{\boldsymbol{\varphi}}\sin\varphi,$
$\hat{\mathbf{y}} = \hat{\boldsymbol{\rho}}\sin\varphi + \hat{\boldsymbol{\varphi}}\cos\varphi,$
$\hat{\mathbf{z}} = \hat{\mathbf{z}}.$

2.4.3 From the results of Exercise 2.4.1 show that

$$\frac{\partial\hat{\boldsymbol{\rho}}}{\partial\varphi} = \hat{\boldsymbol{\varphi}}, \qquad \frac{\partial\hat{\boldsymbol{\varphi}}}{\partial\varphi} = -\hat{\boldsymbol{\rho}}$$

and that all other first derivatives of the circular cylindrical unit vectors with respect to the circular cylindrical coordinates vanish.

2.4.4 Compare $\nabla\cdot\mathbf{V}$ (Eq. (2.34)) with the gradient operator

$$\nabla = \hat{\boldsymbol{\rho}}\frac{\partial}{\partial\rho} + \hat{\boldsymbol{\varphi}}\frac{1}{\rho}\frac{\partial}{\partial\varphi} + \hat{\mathbf{z}}\frac{\partial}{\partial z}$$

(Eq. (2.33)) dotted into $\mathbf{V}$. Note that the differential operators of ∇ differentiate **both** the unit vectors and the components of $\mathbf{V}$.
Hint. $\hat{\boldsymbol{\varphi}}(1/\rho)(\partial/\partial\varphi)\cdot\hat{\boldsymbol{\rho}}V_\rho$ becomes $\hat{\boldsymbol{\varphi}}\cdot\frac{1}{\rho}\frac{\partial}{\partial\varphi}(\hat{\boldsymbol{\rho}}V_\rho)$ and does **not** vanish.

2.4.5 (a) Show that $\mathbf{r} = \hat{\boldsymbol{\rho}}\rho + \hat{\mathbf{z}}z$.

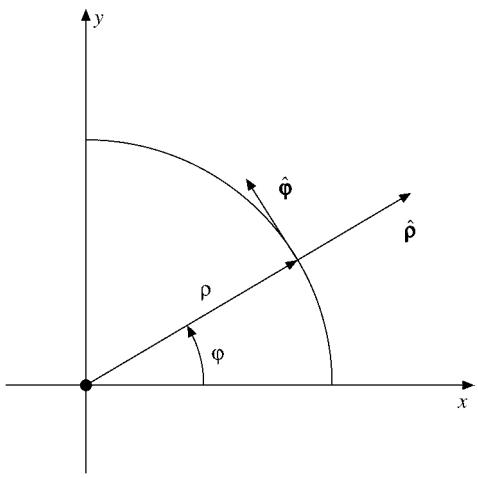

FIGURE 2.6 Plane polar coordinates.

(b) Working entirely in circular cylindrical coordinates, show that

$$\nabla \cdot \mathbf{r} = 3 \qquad \text{and} \qquad \nabla \times \mathbf{r} = 0.$$

2.4.6 (a) Show that the parity operation (reflection through the origin) on a point (ρ, φ, z) relative to **fixed** x-, y-, z-axes consists of the transformation

$$\rho \to \rho, \qquad \varphi \to \varphi \pm \pi, \qquad z \to -z.$$

(b) Show that $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\varphi}}$ have odd parity (reversal of direction) and that $\hat{\mathbf{z}}$ has even parity.

Note. The Cartesian unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ remain constant.

2.4.7 A rigid body is rotating about a fixed axis with a constant angular velocity $\boldsymbol{\omega}$. Take $\boldsymbol{\omega}$ to lie along the z-axis. Express the position vector $\mathbf{r}$ in circular cylindrical coordinates and using circular cylindrical coordinates,
(a) calculate $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, (b) calculate $\nabla \times \mathbf{v}$.

ANS. (a) $\mathbf{v} = \hat{\boldsymbol{\varphi}}\omega\rho$,
(b) $\nabla \times \mathbf{v} = 2\boldsymbol{\omega}$.

2.4.8 Find the circular cylindrical components of the velocity and acceleration of a moving particle,

$$\begin{aligned} v_\rho &= \dot{\rho}, & a_\rho &= \ddot{\rho} - \rho\dot{\varphi}^2, \\ v_\varphi &= \rho\dot{\varphi}, & a_\varphi &= \rho\ddot{\varphi} + 2\dot{\rho}\dot{\varphi}, \\ v_z &= \dot{z}, & a_z &= \ddot{z}. \end{aligned}$$

Hint.

$$\begin{aligned} \mathbf{r}(t) &= \hat{\boldsymbol{\rho}}(t)\rho(t) + \hat{\mathbf{z}}z(t) \\ &= \big[\hat{\mathbf{x}}\cos\varphi(t) + \hat{\mathbf{y}}\sin\varphi(t)\big]\rho(t) + \hat{\mathbf{z}}z(t). \end{aligned}$$

Note. $\dot{\rho} = d\rho/dt$, $\ddot{\rho} = d^2\rho/dt^2$, and so on.

2.4.9 Solve Laplace's equation, $\nabla^2\psi = 0$, in cylindrical coordinates for $\psi = \psi(\rho)$.

ANS. $\psi = k \ln \dfrac{\rho}{\rho_0}$.

2.4.10 In right circular cylindrical coordinates a particular vector function is given by

$$\mathbf{V}(\rho, \varphi) = \hat{\boldsymbol{\rho}}V_\rho(\rho, \varphi) + \hat{\boldsymbol{\varphi}}V_\varphi(\rho, \varphi).$$

Show that $\nabla \times \mathbf{V}$ has only a z-component. Note that this result will hold for any vector confined to a surface $q_3 =$ constant as long as the products h_1V_1 and h_2V_2 are each independent of q_3.

2.4.11 For the flow of an incompressible viscous fluid the Navier–Stokes equations lead to

$$-\nabla \times \big(\mathbf{v} \times (\nabla \times \mathbf{v})\big) = \frac{\eta}{\rho_0}\nabla^2(\nabla \times \mathbf{v}).$$

Here η is the viscosity and ρ_0 is the density of the fluid. For axial flow in a cylindrical pipe we take the velocity $\mathbf{v}$ to be

$$\mathbf{v} = \hat{\mathbf{z}}v(\rho).$$

From Example 2.4.2,

$$\nabla \times \big(\mathbf{v} \times (\nabla \times \mathbf{v})\big) = 0$$

for this choice of **v**.
Show that

$$\nabla^2(\nabla \times \mathbf{v}) = 0$$

leads to the differential equation

$$\frac{1}{\rho}\frac{d}{d\rho}\left(\rho\frac{d^2v}{d\rho^2}\right) - \frac{1}{\rho^2}\frac{dv}{d\rho} = 0$$

and that this is satisfied by

$$v = v_0 + a_2\rho^2.$$

2.4.12 A conducting wire along the z-axis carries a current I. The resulting magnetic vector potential is given by

$$\mathbf{A} = \hat{\mathbf{z}}\frac{\mu I}{2\pi}\ln\left(\frac{1}{\rho}\right).$$

Show that the magnetic induction **B** is given by

$$\mathbf{B} = \hat{\boldsymbol{\varphi}}\frac{\mu I}{2\pi\rho}.$$

2.4.13 A force is described by

$$\mathbf{F} = -\hat{\mathbf{x}}\frac{y}{x^2+y^2} + \hat{\mathbf{y}}\frac{x}{x^2+y^2}.$$

(a) Express **F** in circular cylindrical coordinates.

Operating entirely in circular cylindrical coordinates for (b) and (c),

(b) calculate the curl of **F** and
(c) calculate the work done by **F** in travers the unit circle once counterclockwise.
(d) How do you reconcile the results of (b) and (c)?

2.4.14 A transverse electromagnetic wave (TEM) in a coaxial waveguide has an electric field $\mathbf{E} = \mathbf{E}(\rho, \varphi)e^{i(kz-\omega t)}$ and a magnetic induction field of $\mathbf{B} = \mathbf{B}(\rho, \varphi)e^{i(kz-\omega t)}$. Since the wave is transverse, neither **E** nor **B** has a z component. The two fields satisfy the **vector** Laplacian equation

$$\nabla^2\mathbf{E}(\rho, \varphi) = 0$$

$$\nabla^2\mathbf{B}(\rho, \varphi) = 0.$$

(a) Show that $\mathbf{E} = \hat{\boldsymbol{\rho}}E_0(a/\rho)e^{i(kz-\omega t)}$ and $\mathbf{B} = \hat{\boldsymbol{\varphi}}B_0(a/\rho)e^{i(kz-\omega t)}$ are solutions. Here a is the radius of the inner conductor and E_0 and B_0 are constant amplitudes.

(b) Assuming a vacuum inside the waveguide, verify that Maxwell's equations are satisfied with

$$B_0/E_0 = k/\omega = \mu_0\varepsilon_0(\omega/k) = 1/c.$$

2.4.15 A calculation of the magnetohydrodynamic pinch effect involves the evaluation of $(\mathbf{B}\cdot\nabla)\mathbf{B}$. If the magnetic induction $\mathbf{B}$ is taken to be $\mathbf{B} = \hat{\boldsymbol{\varphi}}B_\varphi(\rho)$, show that

$$(\mathbf{B}\cdot\nabla)\mathbf{B} = -\hat{\boldsymbol{\rho}}B_\varphi^2/\rho.$$

2.4.16 The linear velocity of particles in a rigid body rotating with angular velocity ω is given by

$$\mathbf{v} = \hat{\boldsymbol{\varphi}}\rho\omega.$$

Integrate $\oint \mathbf{v}\cdot d\boldsymbol{\lambda}$ around a circle in the xy-plane and verify that

$$\frac{\oint \mathbf{v}\cdot d\boldsymbol{\lambda}}{\text{area}} = \nabla\times\mathbf{v}|_z.$$

2.4.17 A proton of mass m, charge $+e$, and (asymptotic) momentum $p = mv$ is incident on a nucleus of charge $+Ze$ at an impact parameter b. Determine the proton's distance of closest approach.

2.5 SPHERICAL POLAR COORDINATES

Relabeling (q_1, q_2, q_3) as (r, θ, φ), we see that the spherical polar coordinate system consists of the following:

1. Concentric spheres centered at the origin,

$$r = \left(x^2 + y^2 + z^2\right)^{1/2} = \text{constant}.$$

2. Right circular cones centered on the z-(polar) axis, vertices at the origin,

$$\theta = \arccos\frac{z}{(x^2+y^2+z^2)^{1/2}} = \text{constant}.$$

3. Half-planes through the z-(polar) axis,

$$\varphi = \arctan\frac{y}{x} = \text{constant}.$$

By our arbitrary choice of definitions of θ, the polar angle, and φ, the azimuth angle, the z-axis is singled out for special treatment. The transformation equations corresponding to Eq. (2.1) are

$$x = r\sin\theta\cos\varphi, \qquad y = r\sin\theta\sin\varphi, \qquad z = r\cos\theta, \tag{2.38}$$

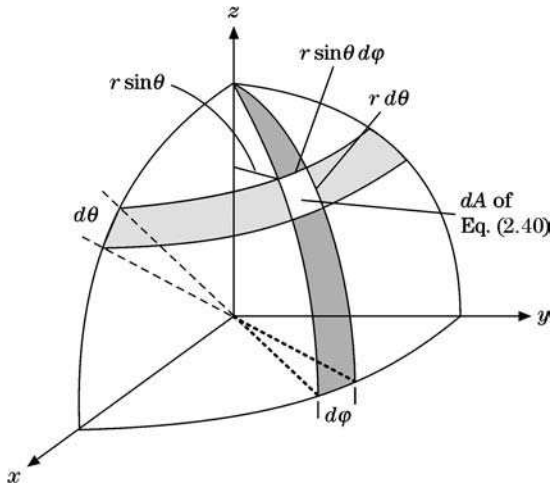

FIGURE 2.7 Spherical polar coordinate area elements.

measuring θ from the positive z-axis and φ in the xy-plane from the positive x-axis. The ranges of values are $0 \leqslant r < \infty$, $0 \leqslant \theta \leqslant \pi$, and $0 \leqslant \varphi \leqslant 2\pi$. At $r = 0$, θ and φ are undefined. From differentiation of Eq. (2.38),

$$\begin{aligned} h_1 &= h_r = 1, \\ h_2 &= h_\theta = r, \\ h_3 &= h_\varphi = r\sin\theta. \end{aligned} \tag{2.39}$$

This gives a line element

$$d\mathbf{r} = \hat{\mathbf{r}}\,dr + \hat{\boldsymbol{\theta}} r\,d\theta + \hat{\boldsymbol{\varphi}} r\sin\theta\,d\varphi,$$

so

$$ds^2 = d\mathbf{r}\cdot d\mathbf{r} = dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\,d\varphi^2,$$

the coordinates being obviously orthogonal. In this spherical coordinate system the area element (for $r =$ constant) is

$$dA = d\sigma_{\theta\varphi} = r^2\sin\theta\,d\theta\,d\varphi, \tag{2.40}$$

the light, unshaded area in Fig. 2.7. Integrating over the azimuth φ, we find that the area element becomes a ring of width $d\theta$,

$$dA_\theta = 2\pi r^2\sin\theta\,d\theta. \tag{2.41}$$

This form will appear repeatedly in problems in spherical polar coordinates with azimuthal symmetry, such as the scattering of an unpolarized beam of particles. By definition of solid radians, or steradians, an element of solid angle $d\Omega$ is given by

$$d\Omega = \frac{dA}{r^2} = \sin\theta\,d\theta\,d\varphi. \tag{2.42}$$

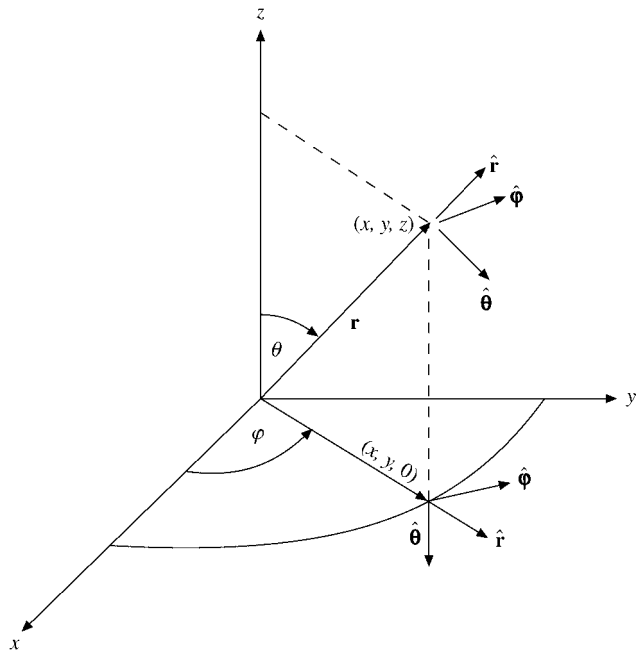

FIGURE 2.8 Spherical polar coordinates.

Integrating over the entire spherical surface, we obtain

$$\int d\Omega = 4\pi.$$

From Eq. (2.11) the volume element is

$$d\tau = r^2\,dr\sin\theta\,d\theta\,d\varphi = r^2\,dr\,d\Omega. \tag{2.43}$$

The spherical polar coordinate unit vectors are shown in Fig. 2.8.

It must be emphasized that **the unit vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\varphi}}$ vary in direction as the angles θ and φ vary.** Specifically, the θ and φ derivatives of these spherical polar coordinate unit vectors do not vanish (Exercise 2.5.2). When differentiating vectors in spherical polar (or in any non-Cartesian system), this variation of the unit vectors with position must not be neglected. In terms of the fixed-direction Cartesian unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ (cp. Eq. (2.38)),

$$\begin{aligned}
\hat{\mathbf{r}} &= \hat{\mathbf{x}}\sin\theta\cos\varphi + \hat{\mathbf{y}}\sin\theta\sin\varphi + \hat{\mathbf{z}}\cos\theta,\\
\hat{\boldsymbol{\theta}} &= \hat{\mathbf{x}}\cos\theta\cos\varphi + \hat{\mathbf{y}}\cos\theta\sin\varphi - \hat{\mathbf{z}}\sin\theta = \frac{\partial\hat{\mathbf{r}}}{\partial\theta},\\
\hat{\boldsymbol{\varphi}} &= -\hat{\mathbf{x}}\sin\varphi + \hat{\mathbf{y}}\cos\varphi = \frac{1}{\sin\theta}\frac{\partial\hat{\mathbf{r}}}{\partial\varphi},
\end{aligned} \tag{2.44}$$

which follow from

$$0 = \frac{\partial\hat{\mathbf{r}}^2}{\partial\theta} = 2\hat{\mathbf{r}}\cdot\frac{\partial\hat{\mathbf{r}}}{\partial\theta}, \qquad 0 = \frac{\partial\hat{\mathbf{r}}^2}{\partial\varphi} = 2\hat{\mathbf{r}}\cdot\frac{\partial\hat{\mathbf{r}}}{\partial\varphi}.$$

Note that Exercise 2.5.5 gives the inverse transformation and that a given vector can now be expressed in a number of different (but equivalent) ways. For instance, the position vector $\mathbf{r}$ may be written

$$\begin{aligned}\mathbf{r} &= \hat{\mathbf{r}}r = \hat{\mathbf{r}}\left(x^2+y^2+z^2\right)^{1/2}\\ &= \hat{\mathbf{x}}x+\hat{\mathbf{y}}y+\hat{\mathbf{z}}z\\ &= \hat{\mathbf{x}}r\sin\theta\cos\varphi+\hat{\mathbf{y}}r\sin\theta\sin\varphi+\hat{\mathbf{z}}r\cos\theta. \end{aligned}\tag{2.45}$$

Select the form that is most useful for your particular problem.

From Section 2.2, relabeling the curvilinear coordinate unit vectors $\hat{\mathbf{q}}_1$, $\hat{\mathbf{q}}_2$, and $\hat{\mathbf{q}}_3$ as $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\varphi}}$ gives

$$\nabla\psi = \hat{\mathbf{r}}\frac{\partial\psi}{\partial r}+\hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial\psi}{\partial\theta}+\hat{\boldsymbol{\varphi}}\frac{1}{r\sin\theta}\frac{\partial\psi}{\partial\varphi}, \tag{2.46}$$

$$\nabla\cdot\mathbf{V} = \frac{1}{r^2\sin\theta}\left[\sin\theta\frac{\partial}{\partial r}(r^2V_r)+r\frac{\partial}{\partial\theta}(\sin\theta V_\theta)+r\frac{\partial V_\varphi}{\partial\varphi}\right], \tag{2.47}$$

$$\nabla\cdot\nabla\psi = \frac{1}{r^2\sin\theta}\left[\sin\theta\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)+\frac{1}{\sin\theta}\frac{\partial^2\psi}{\partial\varphi^2}\right], \tag{2.48}$$

$$\nabla\times\mathbf{V} = \frac{1}{r^2\sin\theta}\begin{vmatrix}\hat{\mathbf{r}} & r\hat{\boldsymbol{\theta}} & r\sin\theta\hat{\boldsymbol{\varphi}}\\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial\theta} & \dfrac{\partial}{\partial\varphi}\\ V_r & rV_\theta & r\sin\theta V_\varphi\end{vmatrix}. \tag{2.49}$$

Occasionally, the vector Laplacian $\nabla^2\mathbf{V}$ is needed in spherical polar coordinates. It is best obtained by using the vector identity (Eq. (1.85)) of Chapter 1. For reference

$$\begin{aligned}\nabla^2\mathbf{V}|_r &= \left(-\frac{2}{r^2}+\frac{2}{r}\frac{\partial}{\partial r}+\frac{\partial^2}{\partial r^2}+\frac{\cos\theta}{r^2\sin\theta}\frac{\partial}{\partial\theta}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right)V_r\\ &\quad+\left(-\frac{2}{r^2}\frac{\partial}{\partial\theta}-\frac{2\cos\theta}{r^2\sin\theta}\right)V_\theta+\left(-\frac{2}{r^2\sin\theta}\frac{\partial}{\partial\varphi}\right)V_\varphi\\ &= \nabla^2V_r-\frac{2}{r^2}V_r-\frac{2}{r^2}\frac{\partial V_\theta}{\partial\theta}-\frac{2\cos\theta}{r^2\sin\theta}V_\theta-\frac{2}{r^2\sin\theta}\frac{\partial V_\varphi}{\partial\varphi}, \end{aligned}\tag{2.50}$$

$$\nabla^2\mathbf{V}|_\theta = \nabla^2V_\theta-\frac{1}{r^2\sin^2\theta}V_\theta+\frac{2}{r^2}\frac{\partial V_r}{\partial\theta}-\frac{2\cos\theta}{r^2\sin^2\theta}\frac{\partial V_\varphi}{\partial\varphi}, \tag{2.51}$$

$$\nabla^2\mathbf{V}|_\varphi = \nabla^2V_\varphi-\frac{1}{r^2\sin^2\theta}V_\varphi+\frac{2}{r^2\sin\theta}\frac{\partial V_r}{\partial\varphi}+\frac{2\cos\theta}{r^2\sin^2\theta}\frac{\partial V_\theta}{\partial\varphi}. \tag{2.52}$$

These expressions for the components of $\nabla^2\mathbf{V}$ are undeniably messy, but sometimes they are needed.

Example 2.5.1 ∇, $\nabla\cdot$, $\nabla\times$ FOR A CENTRAL FORCE

Using Eqs. (2.46) to (2.49), we can reproduce by inspection some of the results derived in Chapter 1 by laborious application of Cartesian coordinates.

From Eq. (2.46),

$$\begin{aligned}\nabla f(r) &= \hat{\mathbf{r}}\frac{df}{dr},\\ \nabla r^n &= \hat{\mathbf{r}} n r^{n-1}.\end{aligned} \tag{2.53}$$

For the Coulomb potential $V = Ze/(4\pi\varepsilon_0 r)$, the electric field is $\mathbf{E} = -\nabla V = \frac{Ze}{4\pi\varepsilon_0 r^2}\hat{\mathbf{r}}$.

From Eq. (2.47),

$$\begin{aligned}\nabla\cdot\hat{\mathbf{r}} f(r) &= \frac{2}{r}f(r) + \frac{df}{dr},\\ \nabla\cdot\hat{\mathbf{r}} r^n &= (n+2)r^{n-1}.\end{aligned} \tag{2.54}$$

For $r > 0$ the charge density of the electric field of the Coulomb potential is $\rho = \nabla\cdot\mathbf{E} = \frac{Ze}{4\pi\varepsilon_0}\nabla\cdot\frac{\hat{\mathbf{r}}}{r^2} = 0$ because $n = -2$.

From Eq. (2.48),

$$\nabla^2 f(r) = \frac{2}{r}\frac{df}{dr} + \frac{d^2 f}{dr^2}, \tag{2.55}$$

$$\nabla^2 r^n = n(n+1)r^{n-2}, \tag{2.56}$$

in contrast to the ordinary radial second derivative of r^n involving $n-1$ instead of $n+1$.

Finally, from Eq. (2.49),

$$\nabla\times\hat{\mathbf{r}} f(r) = 0. \tag{2.57}$$

■

Example 2.5.2 MAGNETIC VECTOR POTENTIAL

The computation of the magnetic vector potential of a single current loop in the xy-plane uses Oersted's law, $\nabla\times\mathbf{H} = \mathbf{J}$, in conjunction with $\mu_0\mathbf{H} = \mathbf{B} = \nabla\times\mathbf{A}$ (see Examples 1.9.2 and 1.12.1), and involves the evaluation of

$$\mu_0\mathbf{J} = \nabla\times\big[\nabla\times\hat{\boldsymbol{\varphi}}A_\varphi(r,\theta)\big].$$

In spherical polar coordinates this reduces to

$$\begin{aligned}\mu_0\mathbf{J} &= \nabla\times\frac{1}{r^2\sin\theta}\begin{vmatrix}\hat{\mathbf{r}} & r\hat{\boldsymbol{\theta}} & r\sin\theta\hat{\boldsymbol{\varphi}}\\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial\theta} & \dfrac{\partial}{\partial\varphi}\\ 0 & 0 & r\sin\theta A_\varphi(r,\theta)\end{vmatrix}\\ &= \nabla\times\frac{1}{r^2\sin\theta}\left[\hat{\mathbf{r}}\frac{\partial}{\partial\theta}(r\sin\theta A_\varphi) - r\hat{\boldsymbol{\theta}}\frac{\partial}{\partial r}(r\sin\theta A_\varphi)\right].\end{aligned}$$

Taking the curl a second time, we obtain

$$\mu_0\mathbf{J} = \frac{1}{r^2\sin\theta}\begin{vmatrix} \hat{\mathbf{r}} & r\hat{\boldsymbol{\theta}} & r\sin\theta\hat{\boldsymbol{\varphi}} \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial\theta} & \dfrac{\partial}{\partial\varphi} \\ \dfrac{1}{r^2\sin\theta}\dfrac{\partial}{\partial\theta}(r\sin\theta A_\varphi) & -\dfrac{1}{r\sin\theta}\dfrac{\partial}{\partial r}(r\sin\theta A_\varphi) & 0 \end{vmatrix}.$$

By expanding the determinant along the top row, we have

$$\begin{aligned}\mu_0\mathbf{J} &= -\hat{\boldsymbol{\varphi}}\left\{\frac{1}{r}\frac{\partial^2}{\partial r^2}(rA_\varphi) + \frac{1}{r^2}\frac{\partial}{\partial\theta}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta A_\varphi)\right]\right\} \\ &= -\hat{\boldsymbol{\varphi}}\left[\nabla^2 A_\varphi(r,\theta) - \frac{1}{r^2\sin^2\theta}A_\varphi(r,\theta)\right]. \end{aligned} \tag{2.58}$$

■

Exercises

2.5.1 Express the spherical polar unit vectors in Cartesian unit vectors.

ANS. $\hat{\mathbf{r}} = \hat{\mathbf{x}}\sin\theta\cos\varphi + \hat{\mathbf{y}}\sin\theta\sin\varphi + \hat{\mathbf{z}}\cos\theta,$
$\hat{\boldsymbol{\theta}} = \hat{\mathbf{x}}\cos\theta\cos\varphi + \hat{\mathbf{y}}\cos\theta\sin\varphi - \hat{\mathbf{z}}\sin\theta,$
$\hat{\boldsymbol{\varphi}} = -\hat{\mathbf{x}}\sin\varphi + \hat{\mathbf{y}}\cos\varphi.$

2.5.2 (a) From the results of Exercise 2.5.1, calculate the partial derivatives of $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\varphi}}$ with respect to r, θ, and φ.

(b) With ∇ given by

$$\hat{\mathbf{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial\theta} + \hat{\boldsymbol{\varphi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial\varphi}$$

(greatest space rate of change), use the results of part (a) to calculate $\nabla\cdot\nabla\psi$. This is an alternate derivation of the Laplacian.

Note. The derivatives of the left-hand ∇ operate on the unit vectors of the right-hand ∇ **before** the unit vectors are dotted together.

2.5.3 A rigid body is rotating about a fixed axis with a constant angular velocity $\boldsymbol{\omega}$. Take $\boldsymbol{\omega}$ to be along the z-axis. Using spherical polar coordinates,

(a) Calculate

$$\mathbf{v} = \boldsymbol{\omega}\times\mathbf{r}.$$

(b) Calculate

$$\nabla\times\mathbf{v}.$$

ANS. (a) $\mathbf{v} = \hat{\boldsymbol{\varphi}}\omega r\sin\theta,$
(b) $\nabla\times\mathbf{v} = 2\boldsymbol{\omega}.$

2.5.4 The coordinate system (x, y, z) is rotated through an angle Φ counterclockwise about an axis defined by the unit vector $\mathbf{n}$ into system (x', y', z'). In terms of the new coordinates the radius vector becomes

$$\mathbf{r}' = \mathbf{r}\cos\Phi + \mathbf{r}\times\mathbf{n}\sin\Phi + \mathbf{n}(\mathbf{n}\cdot\mathbf{r})(1-\cos\Phi).$$

(a) Derive this expression from geometric considerations.
(b) Show that it reduces as expected for $\mathbf{n} = \hat{\mathbf{z}}$. The answer, in matrix form, appears in Eq. (3.90).
(c) Verify that $r'^2 = r^2$.

2.5.5 Resolve the Cartesian unit vectors into their spherical polar components:

$$\hat{\mathbf{x}} = \hat{\mathbf{r}}\sin\theta\cos\varphi + \hat{\boldsymbol{\theta}}\cos\theta\cos\varphi - \hat{\boldsymbol{\varphi}}\sin\varphi,$$
$$\hat{\mathbf{y}} = \hat{\mathbf{r}}\sin\theta\sin\varphi + \hat{\boldsymbol{\theta}}\cos\theta\sin\varphi + \hat{\boldsymbol{\varphi}}\cos\varphi,$$
$$\hat{\mathbf{z}} = \hat{\mathbf{r}}\cos\theta - \hat{\boldsymbol{\theta}}\sin\theta.$$

2.5.6 The direction of one vector is given by the angles θ_1 and φ_1. For a second vector the corresponding angles are θ_2 and φ_2. Show that the cosine of the included angle γ is given by

$$\cos\gamma = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\varphi_1 - \varphi_2).$$

See Fig. 12.15.

2.5.7 A certain vector $\mathbf{V}$ has no radial component. Its curl has no tangential components. What does this imply about the radial dependence of the tangential components of $\mathbf{V}$?

2.5.8 Modern physics lays great stress on the property of parity — whether a quantity remains invariant or changes sign under an inversion of the coordinate system. In Cartesian coordinates this means $x \to -x$, $y \to -y$, and $z \to -z$.

(a) Show that the inversion (reflection through the origin) of a point (r, θ, φ) relative to **fixed** x-, y-, z-axes consists of the transformation

$$r \to r, \qquad \theta \to \pi - \theta, \qquad \varphi \to \varphi \pm \pi.$$

(b) Show that $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\varphi}}$ have odd parity (reversal of direction) and that $\hat{\boldsymbol{\theta}}$ has even parity.

2.5.9 With $\mathbf{A}$ any vector,

$$\mathbf{A}\cdot\nabla\mathbf{r} = \mathbf{A}.$$

(a) Verify this result in Cartesian coordinates.
(b) Verify this result using spherical polar coordinates. (Equation (2.46) provides ∇.)

2.5.10 Find the spherical coordinate components of the velocity and acceleration of a moving particle:

$$
\begin{aligned}
v_r &= \dot{r},\\
v_\theta &= r\dot{\theta},\\
v_\varphi &= r\sin\theta\dot{\varphi},\\
a_r &= \ddot{r} - r\dot{\theta}^2 - r\sin^2\theta\dot{\varphi}^2,\\
a_\theta &= r\ddot{\theta} + 2\dot{r}\dot{\theta} - r\sin\theta\cos\theta\dot{\varphi}^2,\\
a_\varphi &= r\sin\theta\ddot{\varphi} + 2\dot{r}\sin\theta\dot{\varphi} + 2r\cos\theta\dot{\theta}\dot{\varphi}.
\end{aligned}
$$

Hint.

$$
\begin{aligned}
\mathbf{r}(t) &= \hat{\mathbf{r}}(t)r(t)\\
&= \left[\hat{\mathbf{x}}\sin\theta(t)\cos\varphi(t) + \hat{\mathbf{y}}\sin\theta(t)\sin\varphi(t) + \hat{\mathbf{z}}\cos\theta(t)\right]r(t).
\end{aligned}
$$

Note. Using the Lagrangian techniques of Section 17.3, we may obtain these results somewhat more elegantly. The dot in $\dot{r}, \dot{\theta}, \dot{\varphi}$ means time derivative, $\dot{r} = dr/dt, \dot{\theta} = d\theta/dt, \dot{\varphi} = d\varphi/dt$. The notation was originated by Newton.

2.5.11 A particle m moves in response to a central force according to Newton's second law,

$$m\ddot{\mathbf{r}} = \hat{\mathbf{r}}f(r).$$

Show that $\mathbf{r}\times\dot{\mathbf{r}} = \mathbf{c}$, a constant, and that the geometric interpretation of this leads to Kepler's second law.

2.5.12 Express $\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$ in spherical polar coordinates.

$$
\begin{aligned}
\text{ANS. } \frac{\partial}{\partial x} &= \sin\theta\cos\varphi\frac{\partial}{\partial r} + \cos\theta\cos\varphi\frac{1}{r}\frac{\partial}{\partial\theta} - \frac{\sin\varphi}{r\sin\theta}\frac{\partial}{\partial\varphi},\\
\frac{\partial}{\partial y} &= \sin\theta\sin\varphi\frac{\partial}{\partial r} + \cos\theta\sin\varphi\frac{1}{r}\frac{\partial}{\partial\theta} + \frac{\cos\varphi}{r\sin\theta}\frac{\partial}{\partial\varphi},\\
\frac{\partial}{\partial z} &= \cos\theta\frac{\partial}{\partial r} - \sin\theta\frac{1}{r}\frac{\partial}{\partial\theta}.
\end{aligned}
$$

Hint. Equate ∇_{xyz} and $\nabla_{r\theta\varphi}$.

2.5.13 From Exercise 2.5.12 show that

$$-i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) = -i\frac{\partial}{\partial\varphi}.$$

This is the quantum mechanical operator corresponding to the z-component of orbital angular momentum.

2.5.14 With the quantum mechanical orbital angular momentum operator defined as $\mathbf{L} = -i(\mathbf{r}\times\nabla)$, show that

(a) $L_x + iL_y = e^{i\varphi}\left(\dfrac{\partial}{\partial\theta} + i\cot\theta\dfrac{\partial}{\partial\varphi}\right)$,

(b) $L_x - iL_y = -e^{-i\varphi}\left(\dfrac{\partial}{\partial\theta} - i\cot\theta\dfrac{\partial}{\partial\varphi}\right)$.

(These are the raising and lowering operators of Section 4.3.)

2.5.15 Verify that $\mathbf{L}\times\mathbf{L} = i\mathbf{L}$ in spherical polar coordinates. $\mathbf{L} = -i(\mathbf{r}\times\boldsymbol{\nabla})$, the quantum mechanical orbital angular momentum operator.
Hint. Use spherical polar coordinates for $\mathbf{L}$ but Cartesian components for the cross product.

2.5.16 (a) From Eq. (2.46) show that

$$\mathbf{L} = -i(\mathbf{r}\times\boldsymbol{\nabla}) = i\left(\hat{\boldsymbol{\theta}}\frac{1}{\sin\theta}\frac{\partial}{\partial\varphi} - \hat{\boldsymbol{\varphi}}\frac{\partial}{\partial\theta}\right).$$

(b) Resolving $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ into Cartesian components, determine L_x, L_y, and L_z in terms of θ, φ, and their derivatives.

(c) From $L^2 = L_x^2 + L_y^2 + L_z^2$ show that

$$\begin{aligned}\mathbf{L}^2 &= -\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) - \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\\ &= -r^2\boldsymbol{\nabla}^2 + \frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right).\end{aligned}$$

This latter identity is useful in relating orbital angular momentum and Legendre's differential equation, Exercise 9.3.8.

2.5.17 With $\mathbf{L} = -i\mathbf{r}\times\boldsymbol{\nabla}$, verify the operator identities

(a) $\boldsymbol{\nabla} = \hat{\mathbf{r}}\dfrac{\partial}{\partial r} - i\dfrac{\mathbf{r}\times\mathbf{L}}{r^2}$,

(b) $\mathbf{r}\boldsymbol{\nabla}^2 - \boldsymbol{\nabla}\left(1 + r\dfrac{\partial}{\partial r}\right) = i\boldsymbol{\nabla}\times\mathbf{L}$.

2.5.18 Show that the following three forms (spherical coordinates) of $\boldsymbol{\nabla}^2\psi(r)$ are equivalent:

(a) $\dfrac{1}{r^2}\dfrac{d}{dr}\left[r^2\dfrac{d\psi(r)}{dr}\right]$; (b) $\dfrac{1}{r}\dfrac{d^2}{dr^2}\big[r\psi(r)\big]$; (c) $\dfrac{d^2\psi(r)}{dr^2} + \dfrac{2}{r}\dfrac{d\psi(r)}{dr}$.

The second form is particularly convenient in establishing a correspondence between spherical polar and Cartesian descriptions of a problem.

2.5.19 One model of the solar corona assumes that the steady-state equation of heat flow,

$$\boldsymbol{\nabla}\cdot(k\boldsymbol{\nabla}T) = 0,$$

is satisfied. Here, k, the thermal conductivity, is proportional to $T^{5/2}$. Assuming that the temperature T is proportional to r^n, show that the heat flow equation is satisfied by $T = T_0(r_0/r)^{2/7}$.

2.5.20 A certain force field is given by

$$\mathbf{F} = \hat{\mathbf{r}}\frac{2P\cos\theta}{r^3} + \hat{\boldsymbol{\theta}}\frac{P}{r^3}\sin\theta, \qquad r \geqslant P/2$$

(in spherical polar coordinates).

(a) Examine $\nabla \times \mathbf{F}$ to see if a potential exists.
(b) Calculate $\oint \mathbf{F} \cdot d\boldsymbol{\lambda}$ for a unit circle in the plane $\theta = \pi/2$. What does this indicate about the force being conservative or nonconservative?
(c) If you believe that $\mathbf{F}$ may be described by $\mathbf{F} = -\nabla\psi$, find ψ. Otherwise simply state that no acceptable potential exists.

2.5.21 (a) Show that $\mathbf{A} = -\hat{\boldsymbol{\varphi}}\cot\theta/r$ is a solution of $\nabla \times \mathbf{A} = \hat{\mathbf{r}}/r^2$.
(b) Show that this spherical polar coordinate solution agrees with the solution given for Exercise 1.13.6:

$$\mathbf{A} = \hat{\mathbf{x}}\frac{yz}{r(x^2+y^2)} - \hat{\mathbf{y}}\frac{xz}{r(x^2+y^2)}.$$

Note that the solution diverges for $\theta = 0, \pi$ corresponding to $x, y = 0$.
(c) Finally, show that $\mathbf{A} = -\hat{\boldsymbol{\theta}}\varphi\sin\theta/r$ is a solution. Note that although this solution does not diverge ($r \neq 0$), it is no longer single-valued for all possible azimuth angles.

2.5.22 A magnetic vector potential is given by

$$\mathbf{A} = \frac{\mu_0}{4\pi}\frac{\mathbf{m}\times\mathbf{r}}{r^3}.$$

Show that this leads to the magnetic induction $\mathbf{B}$ of a point magnetic dipole with dipole moment $\mathbf{m}$.

ANS. for $\mathbf{m} = \hat{\mathbf{z}}m$,

$$\nabla \times \mathbf{A} = \hat{\mathbf{r}}\frac{\mu_0}{4\pi}\frac{2m\cos\theta}{r^3} + \hat{\boldsymbol{\theta}}\frac{\mu_0}{4\pi}\frac{m\sin\theta}{r^3}.$$

Compare Eqs. (12.133) and (12.134)

2.5.23 At large distances from its source, electric dipole radiation has fields

$$\mathbf{E} = a_E\sin\theta\frac{e^{i(kr-\omega t)}}{r}\hat{\boldsymbol{\theta}}, \qquad \mathbf{B} = a_B\sin\theta\frac{e^{i(kr-\omega t)}}{r}\hat{\boldsymbol{\varphi}}.$$

Show that Maxwell's equations

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{and} \qquad \nabla \times \mathbf{B} = \varepsilon_0\mu_0\frac{\partial \mathbf{E}}{\partial t}$$

are satisfied, if we take

$$\frac{a_E}{a_B} = \frac{\omega}{k} = c = (\varepsilon_0\mu_0)^{-1/2}.$$

Hint. Since r is large, terms of order r^{-2} may be dropped.

2.5.24 The magnetic vector potential for a uniformly charged rotating spherical shell is

$$\mathbf{A} = \begin{cases} \hat{\boldsymbol{\varphi}} \dfrac{\mu_0 a^4 \sigma \omega}{3} \cdot \dfrac{\sin\theta}{r^2}, & r > a \\ \hat{\boldsymbol{\varphi}} \dfrac{\mu_0 a \sigma \omega}{3} \cdot r\cos\theta, & r < a. \end{cases}$$

(a = radius of spherical shell, σ = surface charge density, and ω = angular velocity.) Find the magnetic induction $\mathbf{B} = \nabla \times \mathbf{A}$.

$$\text{ANS. } B_r(r,\theta) = \frac{2\mu_0 a^4 \sigma \omega}{3} \cdot \frac{\cos\theta}{r^3}, \qquad r > a,$$

$$B_\theta(r,\theta) = \frac{\mu_0 a^4 \sigma \omega}{3} \cdot \frac{\sin\theta}{r^3}, \qquad r > a,$$

$$\mathbf{B} = \hat{\mathbf{z}} \frac{2\mu_0 a \sigma \omega}{3}, \qquad r < a.$$

2.5.25 (a) Explain why ∇^2 in plane polar coordinates follows from ∇^2 in circular cylindrical coordinates with z = constant.

(b) Explain why taking ∇^2 in spherical polar coordinates and restricting θ to $\pi/2$ does **not** lead to the plane polar form of ∇.

Note.

$$\nabla^2(\rho,\varphi) = \frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2}.$$

2.6 TENSOR ANALYSIS

Introduction, Definitions

Tensors are important in many areas of physics, including general relativity and electrodynamics. Scalars and vectors are special cases of tensors. In Chapter 1, a quantity that did not change under rotations of the coordinate system in three-dimensional space, an invariant, was labeled a scalar. A **scalar** is specified by one real number and is a **tensor of rank 0**. A quantity whose components transformed under rotations like those of the distance of a point from a chosen origin (Eq. (1.9), Section 1.2) was called a vector. The transformation of the components of the vector under a rotation of the coordinates preserves the vector as a geometric entity (such as an arrow in space), independent of the orientation of the reference frame. In three-dimensional space, a **vector** is specified by $3 = 3^1$ real numbers, for example, its Cartesian components, and is a **tensor of rank 1**. A **tensor of rank** n has 3^n components that transform in a definite way.[5] This transformation philosophy is of central importance for tensor analysis and conforms with the mathematician's concept of vector and vector (or linear) space and the physicist's notion that physical observables must not depend on the choice of coordinate frames. There is a physical basis for such a philosophy: We describe the physical world by mathematics, but any physical predictions we make

[5] In N-dimensional space a tensor of rank n has N^n components.

must be independent of our mathematical conventions, such as a coordinate system with its arbitrary origin and orientation of its axes.

There is a possible ambiguity in the transformation law of a vector

$$A_i' = \sum_j a_{ij} A_j, \tag{2.59}$$

in which a_{ij} is the cosine of the angle between the x_i'-axis and the x_j-axis.

If we start with a differential distance vector $d\mathbf{r}$, then, taking dx_i' to be a function of the unprimed variables,

$$dx_i' = \sum_j \frac{\partial x_i'}{\partial x_j}\, dx_j \tag{2.60}$$

by partial differentiation. If we set

$$a_{ij} = \frac{\partial x_i'}{\partial x_j}, \tag{2.61}$$

Eqs. (2.59) and (2.60) are consistent. Any set of quantities A^j transforming according to

$$A'^i = \sum_j \frac{\partial x_i'}{\partial x_j} A^j \tag{2.62a}$$

is defined as a **contravariant** vector, whose indices we write as **superscript**; this includes the Cartesian coordinate vector $x^i = x_i$ from now on.

However, we have already encountered a slightly different type of vector transformation. The gradient of a scalar $\nabla\varphi$, defined by

$$\nabla\varphi = \hat{\mathbf{x}}\frac{\partial\varphi}{\partial x^1} + \hat{\mathbf{y}}\frac{\partial\varphi}{\partial x^2} + \hat{\mathbf{z}}\frac{\partial\varphi}{\partial x^3} \tag{2.63}$$

(using x^1, x^2, x^3 for x, y, z), transforms as

$$\frac{\partial\varphi'}{\partial x'^i} = \sum_j \frac{\partial\varphi}{\partial x^j}\frac{\partial x^j}{\partial x'^i}, \tag{2.64}$$

using $\varphi = \varphi(x, y, z) = \varphi(x', y', z') = \varphi'$, φ defined as a scalar quantity. Notice that this differs from Eq. (2.62) in that we have $\partial x^j/\partial x'^i$ instead of $\partial x'^i/\partial x^j$. Equation (2.64) is taken as the definition of a **covariant** vector, with the gradient as the prototype. The covariant analog of Eq. (2.62a) is

$$A_i' = \sum_j \frac{\partial x^j}{\partial x'^i}\, A_j. \tag{2.62b}$$

Only in Cartesian coordinates is

$$\frac{\partial x^j}{\partial x'^i} = \frac{\partial x'^i}{\partial x^j} = a_{ij} \tag{2.65}$$

so that there no difference between contravariant and covariant transformations. In other systems, Eq. (2.65) in general does not apply, and the distinction between contravariant and covariant is real and must be observed. This is of prime importance in the curved Riemannian space of general relativity.

In the remainder of this section the components of any **contravariant** vector are denoted by a **superscript**, A^i, whereas a **subscript** is used for the components of a **covariant** vector A_i.[6]

Definition of Tensors of Rank 2

Now we proceed to define **contravariant, mixed, and covariant tensors of rank 2** by the following equations for their components under coordinate transformations:

$$A'^{ij} = \sum_{kl} \frac{\partial x'^i}{\partial x^k} \frac{\partial x'^j}{\partial x^l} A^{kl},$$

$$B'^i{}_j = \sum_{kl} \frac{\partial x'^i}{\partial x^k} \frac{\partial x^l}{\partial x'^j} B^k{}_l, \tag{2.66}$$

$$C'_{ij} = \sum_{kl} \frac{\partial x^k}{\partial x'^i} \frac{\partial x^l}{\partial x'^j} C_{kl}.$$

Clearly, the rank goes as the number of partial derivatives (or direction cosines) in the definition: 0 for a scalar, 1 for a vector, 2 for a second-rank tensor, and so on. Each index (subscript or superscript) ranges over the number of dimensions of the space. The number of indices (equal to the rank of tensor) is independent of the dimensions of the space. We see that A^{kl} is contravariant with respect to both indices, C_{kl} is covariant with respect to both indices, and $B^k{}_l$ transforms contravariantly with respect to the first index k but covariantly with respect to the second index l. Once again, if we are using Cartesian coordinates, all three forms of the tensors of second rank contravariant, mixed, and covariant are — the same.

As with the components of a vector, the transformation laws for the components of a tensor, Eq. (2.66), yield entities (and properties) that are independent of the choice of reference frame. This is what makes tensor analysis important in physics. The independence of reference frame (invariance) is ideal for expressing and investigating universal physical laws.

The second-rank tensor $\mathbf{A}$ (components A^{kl}) may be conveniently represented by writing out its components in a square array (3×3 if we are in three-dimensional space):

$$\mathbf{A} = \begin{pmatrix} A^{11} & A^{12} & A^{13} \\ A^{21} & A^{22} & A^{23} \\ A^{31} & A^{32} & A^{33} \end{pmatrix}. \tag{2.67}$$

This does not mean that any square array of numbers or functions forms a tensor. The essential condition is that the components transform according to Eq. (2.66).

[6] This means that the coordinates (x, y, z) are written (x^1, x^2, x^3) since $\mathbf{r}$ transforms as a contravariant vector. The ambiguity of x^2 representing both x squared and y is the price we pay.

In the context of matrix analysis the preceding transformation equations become (for Cartesian coordinates) an orthogonal similarity transformation; see Section 3.3. A geometrical interpretation of a second-rank tensor (the inertia tensor) is developed in Section 3.5.

In summary, tensors are systems of components organized by one or more indices that transform according to specific rules under a set of transformations. The number of indices is called the rank of the tensor. If the transformations are coordinate rotations in three-dimensional space, then tensor analysis amounts to what we did in the sections on curvilinear coordinates and in Cartesian coordinates in Chapter 1. *In four dimensions of Minkowski space–time, the transformations are Lorentz transformations, and tensors of rank* 1 *are called four-vectors.*

Addition and Subtraction of Tensors

The addition and subtraction of tensors is defined in terms of the individual elements, just as for vectors. If

$$\mathbf{A} + \mathbf{B} = \mathbf{C}, \tag{2.68}$$

then

$$A^{ij} + B^{ij} = C^{ij}.$$

Of course, **A** and **B** must be tensors of the same rank and both expressed in a space of the same number of dimensions.

Summation Convention

In tensor analysis it is customary to adopt a summation convention to put Eq. (2.66) and subsequent tensor equations in a more compact form. As long as we are distinguishing between contravariance and covariance, let us agree that when an index appears on one side of an equation, once as a superscript and once as a subscript (except for the coordinates where both are subscripts), we automatically sum over that index. Then we may write the second expression in Eq. (2.66) as

$$B'^{i}{}_{j} = \frac{\partial x'^{i}}{\partial x^{k}} \frac{\partial x^{l}}{\partial x'^{j}} B^{k}{}_{l}, \tag{2.69}$$

with the summation of the right-hand side over k and l implied. This is Einstein's summation convention.[7] The index i is superscript because it is associated with the contravariant x'^{i}; likewise j is subscript because it is related to the covariant gradient.

To illustrate the use of the summation convention and some of the techniques of tensor analysis, let us show that the now-familiar Kronecker delta, δ_{kl}, is really a mixed tensor

[7] In this context $\partial x'^{i}/\partial x^{k}$ might better be written as a^{i}_{k} and $\partial x^{l}/\partial x'^{j}$ as b^{l}_{j}.

of rank 2, $\delta^k{}_l$.[8] The question is: Does $\delta^k{}_l$ transform according to Eq. (2.66)? This is our criterion for calling it a tensor. We have, using the summation convention,

$$\delta^k{}_l \frac{\partial x'^i}{\partial x^k} \frac{\partial x^l}{\partial x'^j} = \frac{\partial x'^i}{\partial x^k} \frac{\partial x^k}{\partial x'^j} \tag{2.70}$$

by definition of the Kronecker delta. Now,

$$\frac{\partial x'^i}{\partial x^k} \frac{\partial x^k}{\partial x'^j} = \frac{\partial x'^i}{\partial x'^j} \tag{2.71}$$

by direct partial differentiation of the right-hand side (chain rule). However, x'^i and x'^j are independent coordinates, and therefore the variation of one with respect to the other must be zero if they are different, unity if they coincide; that is,

$$\frac{\partial x'^i}{\partial x'^j} = \delta'^i{}_j. \tag{2.72}$$

Hence

$$\delta'^i{}_j = \frac{\partial x'^i}{\partial x^k} \frac{\partial x^l}{\partial x'^j} \delta^k{}_l,$$

showing that the $\delta^k{}_l$ are indeed the components of a mixed second-rank tensor. Notice that this result is independent of the number of dimensions of our space. The reason for the upper index i and lower index j is the same as in Eq. (2.69).

The Kronecker delta has one further interesting property. It has the same components in all of our rotated coordinate systems and is therefore called **isotropic**. In Section 2.9 we shall meet a third-rank isotropic tensor and three fourth-rank isotropic tensors. No isotropic first-rank tensor (vector) exists.

Symmetry–Antisymmetry

The order in which the indices appear in our description of a tensor is important. In general, A^{mn} is independent of A^{nm}, but there are some cases of special interest. If, for all m and n,

$$A^{mn} = A^{nm}, \tag{2.73}$$

we call the tensor **symmetric**. If, on the other hand,

$$A^{mn} = -A^{nm}, \tag{2.74}$$

the tensor is **antisymmetric**. Clearly, every (second-rank) tensor can be resolved into symmetric and antisymmetric parts by the identity

$$A^{mn} = \tfrac{1}{2}\big(A^{mn} + A^{nm}\big) + \tfrac{1}{2}\big(A^{mn} - A^{nm}\big), \tag{2.75}$$

the first term on the right being a symmetric tensor, the second, an antisymmetric tensor. A similar resolution of functions into symmetric and antisymmetric parts is of extreme importance to quantum mechanics.

[8] It is common practice to refer to a tensor $\mathbf{A}$ by specifying a typical component, A_{ij}. As long as the reader refrains from writing nonsense such as $\mathbf{A} = A_{ij}$, no harm is done.

Spinors

It was once thought that the system of scalars, vectors, tensors (second-rank), and so on formed a complete mathematical system, one that is adequate for describing a physics independent of the choice of reference frame. But the universe and mathematical physics are not that simple. In the realm of elementary particles, for example, spin zero particles[9] (π mesons, α particles) may be described with scalars, spin 1 particles (deuterons) by vectors, and spin 2 particles (gravitons) by tensors. This listing omits the most common particles: electrons, protons, and neutrons, all with spin $\frac{1}{2}$. These particles are properly described by **spinors**. A spinor is not a scalar, vector, or tensor. A brief introduction to spinors in the context of group theory ($J = 1/2$) appears in Section 4.3.

Exercises

2.6.1 Show that if all the components of any tensor of any rank vanish in one particular coordinate system, they vanish in all coordinate systems.
Note. This point takes on special importance in the four-dimensional curved space of general relativity. If a quantity, expressed as a tensor, exists in one coordinate system, it exists in all coordinate systems and is not just a consequence of a **choice** of a coordinate system (as are centrifugal and Coriolis forces in Newtonian mechanics).

2.6.2 The components of tensor **A** are equal to the corresponding components of tensor **B** in one particular coordinate system, denoted by the superscript 0; that is,

$$A^0_{ij} = B^0_{ij}.$$

Show that tensor **A** is equal to tensor **B**, $A_{ij} = B_{ij}$, in all coordinate systems.

2.6.3 The last three components of a four-dimensional vector vanish in each of two reference frames. If the second reference frame is not merely a rotation of the first about the x_0 axis, that is, if at least one of the coefficients a_{i0} $(i = 1, 2, 3) \neq 0$, show that the zeroth component vanishes in all reference frames. Translated into relativistic mechanics this means that if momentum is conserved in two Lorentz frames, then energy is conserved in all Lorentz frames.

2.6.4 From an analysis of the behavior of a general second-rank tensor under 90° and 180° rotations about the coordinate axes, show that an isotropic second-rank tensor in three-dimensional space must be a multiple of δ_{ij}.

2.6.5 The four-dimensional fourth-rank Riemann–Christoffel curvature tensor of general relativity, R_{iklm}, satisfies the symmetry relations

$$R_{iklm} = -R_{ikml} = -R_{kilm}.$$

With the indices running from 0 to 3, show that the number of independent components is reduced from 256 to 36 and that the condition

$$R_{iklm} = R_{lmik}$$

[9]The particle spin is intrinsic angular momentum (in units of $\hbar$). It is distinct from classical, orbital angular momentum due to motion.

further reduces the number of independent components to 21. Finally, if the components satisfy an identity $R_{iklm} + R_{ilmk} + R_{imkl} = 0$, show that the number of independent components is reduced to 20.
Note. The final three-term identity furnishes new information only if all four indices are different. Then it reduces the number of independent components by one-third.

2.6.6 T_{iklm} is antisymmetric with respect to all pairs of indices. How many independent components has it (in three-dimensional space)?

2.7 CONTRACTION, DIRECT PRODUCT

Contraction

When dealing with vectors, we formed a scalar product (Section 1.3) by summing products of corresponding components:

$$\mathbf{A} \cdot \mathbf{B} = A_i B_i \qquad (\textbf{summation convention}). \tag{2.76}$$

The generalization of this expression in tensor analysis is a process known as contraction. Two indices, one covariant and the other contravariant, are set equal to each other, and then (as implied by the summation convention) we sum over this repeated index. For example, let us contract the second-rank mixed tensor $B'^{i}{}_{j}$,

$$B'^{i}{}_{i} = \frac{\partial x'^{i}}{\partial x^{k}} \frac{\partial x^{l}}{\partial x'^{i}} B^{k}{}_{l} = \frac{\partial x^{l}}{\partial x^{k}} B^{k}{}_{l} \tag{2.77}$$

using Eq. (2.71), and then by Eq. (2.72)

$$B'^{i}{}_{i} = \delta^{l}{}_{k} B^{k}{}_{l} = B^{k}{}_{k}. \tag{2.78}$$

Our contracted second-rank mixed tensor is invariant and therefore a scalar.[10] This is exactly what we obtained in Section 1.3 for the dot product of two vectors and in Section 1.7 for the divergence of a vector. In general, the operation of contraction reduces the rank of a tensor by 2. An example of the use of contraction appears in Chapter 4.

Direct Product

The components of a covariant vector (first-rank tensor) a_i and those of a contravariant vector (first-rank tensor) b^j may be multiplied component by component to give the general term $a_i b^j$. This, by Eq. (2.66) is actually a second-rank tensor, for

$$a'_i b'^{j} = \frac{\partial x^{k}}{\partial x'^{i}} a_k \frac{\partial x'^{j}}{\partial x^{l}} b^{l} = \frac{\partial x^{k}}{\partial x'^{i}} \frac{\partial x'^{j}}{\partial x^{l}} \left(a_k b^{l}\right). \tag{2.79}$$

Contracting, we obtain

$$a'_i b'^{i} = a_k b^{k}, \tag{2.80}$$

[10] In matrix analysis this scalar is the **trace** of the matrix, Section 3.2.

as in Eqs. (2.77) and (2.78), to give the regular scalar product.

The operation of adjoining two vectors a_i and b^j as in the last paragraph is known as forming the **direct product**. For the case of two vectors, the direct product is a tensor of second rank. In this sense we may attach meaning to $\nabla\mathbf{E}$, which was not defined within the framework of vector analysis. In general, the direct product of two tensors is a tensor of rank equal to the sum of the two initial ranks; that is,

$$A^i{}_j B^{kl} = C^i{}_j{}^{kl}, \tag{2.81a}$$

where $C^i{}_j{}^{kl}$ is a tensor of fourth rank. From Eqs. (2.66),

$$C'^i{}_j{}^{kl} = \frac{\partial x'^i}{\partial x^m}\frac{\partial x^n}{\partial x'^j}\frac{\partial x'^k}{\partial x^p}\frac{\partial x'^l}{\partial x^q} C^m{}_n{}^{pq}. \tag{2.81b}$$

The direct product is a technique for creating new, higher-rank tensors. Exercise 2.7.1 is a form of the direct product in which the first factor is ∇. Applications appear in Section 4.6.

When $\mathbf{T}$ is an nth-rank Cartesian tensor, $(\partial/\partial x^i)T_{jkl}\ldots$, a component of $\nabla\mathbf{T}$, is a **Cartesian** tensor of rank $n+1$ (Exercise 2.7.1). However, $(\partial/\partial x^i)T_{jkl}\ldots$ is not a tensor in more general spaces. In non-Cartesian systems $\partial/\partial x'^i$ will act on the partial derivatives $\partial x^p/\partial x'^q$ and destroy the simple tensor transformation relation (see Eq. (2.129)).

So far the distinction between a covariant transformation and a contravariant transformation has been maintained because it does exist in non-Euclidean space and because it is of great importance in general relativity. In Sections 2.10 and 2.11 we shall develop differential relations for general tensors. Often, however, because of the simplification achieved, we restrict ourselves to Cartesian tensors. As noted in Section 2.6, the distinction between contravariance and covariance disappears.

Exercises

2.7.1 If $T_{\ldots i}$ is a tensor of rank n, show that $\partial T_{\ldots i}/\partial x^j$ is a tensor of rank $n+1$ (Cartesian coordinates).
Note. In non-Cartesian coordinate systems the coefficients a_{ij} are, in general, functions of the coordinates, and the simple derivative of a tensor of rank n is not a tensor except in the special case of $n=0$. In this case the derivative does yield a covariant vector (tensor of rank 1) by Eq. (2.64).

2.7.2 If $T_{ijk\ldots}$ is a tensor of rank n, show that $\sum_j \partial T_{ijk\ldots}/\partial x^j$ is a tensor of rank $n-1$ (Cartesian coordinates).

2.7.3 The operator

$$\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$$

may be written as

$$\sum_{i=1}^{4} \frac{\partial^2}{\partial x_i^2},$$

using $x_4 = ict$. This is the four-dimensional Laplacian, sometimes called the d'Alembertian and denoted by $\Box^2$. Show that it is a **scalar** operator, that is, is invariant under Lorentz transformations.

2.8 QUOTIENT RULE

If A_i and B_j are vectors, as seen in Section 2.7, we can easily show that $A_i B_j$ is a second-rank tensor. Here we are concerned with a variety of inverse relations. Consider such equations as

$$K_i A_i = B \tag{2.82a}$$

$$K_{ij} A_j = B_i \tag{2.82b}$$

$$K_{ij} A_{jk} = B_{ik} \tag{2.82c}$$

$$K_{ijkl} A_{ij} = B_{kl} \tag{2.82d}$$

$$K_{ij} A_k = B_{ijk}. \tag{2.82e}$$

Inline with our restriction to Cartesian systems, we write all indices as subscripts and, unless specified otherwise, sum repeated indices.

In each of these expressions **A** and **B** are known tensors of rank indicated by the number of indices and **A** is arbitrary. In each case K is an unknown quantity. We wish to establish the transformation properties of K. The quotient rule asserts that if the equation of interest holds in all (rotated) Cartesian coordinate systems, K is a tensor of the indicated rank. The importance in physical theory is that the quotient rule can establish the tensor nature of quantities. Exercise 2.8.1 is a simple illustration of this. The quotient rule (Eq. (2.82b)) shows that the inertia matrix appearing in the angular momentum equation $\mathbf{L} = I\boldsymbol{\omega}$, Section 3.5, is a tensor.

In proving the quotient rule, we consider Eq. (2.82b) as a typical case. In our primed coordinate system

$$K'_{ij} A'_j = B'_i = a_{ik} B_k, \tag{2.83}$$

using the vector transformation properties of **B**. Since the equation holds in all rotated Cartesian coordinate systems,

$$a_{ik} B_k = a_{ik} (K_{kl} A_l). \tag{2.84}$$

Now, transforming **A** back into the primed coordinate system[11] (compare Eq. (2.62)), we have

$$K'_{ij} A'_j = a_{ik} K_{kl} a_{jl} A'_j. \tag{2.85}$$

Rearranging, we obtain

$$(K'_{ij} - a_{ik} a_{jl} K_{kl}) A'_j = 0. \tag{2.86}$$

[11] Note the order of the indices of the direction cosine a_{jl} in this **inverse** transformation. We have

$$A_l = \sum_j \frac{\partial x_l}{\partial x'_j} A'_j = \sum_j a_{jl} A'_j.$$

This must hold for each value of the index i and for every primed coordinate system. Since the A'_j is arbitrary,[12] we conclude

$$K'_{ij} = a_{ik}a_{jl}K_{kl}, \tag{2.87}$$

which is our definition of second-rank tensor.

The other equations may be treated similarly, giving rise to other forms of the quotient rule. One minor pitfall should be noted: The quotient rule does not necessarily apply if B is zero. The transformation properties of zero are indeterminate.

Example 2.8.1 EQUATIONS OF MOTION AND FIELD EQUATIONS

In classical mechanics, Newton's equations of motion $m\dot{\mathbf{v}} = \mathbf{F}$ tell us on the basis of the quotient rule that, if the mass is a scalar and the force a vector, then the acceleration $\mathbf{a} \equiv \dot{\mathbf{v}}$ is a vector. In other words, the vector character of the force as the driving term imposes its vector character on the acceleration, provided the scale factor m is scalar.

The wave equation of electrodynamics $\partial^2 A^\mu = J^\mu$ involves the four-dimensional version of the Laplacian $\partial^2 = \frac{\partial^2}{c^2\partial t^2} - \nabla^2$, a Lorentz scalar, and the external four-vector current J^μ as its driving term. From the quotient rule, we infer that the vector potential A^μ is a four-vector as well. If the driving current is a four-vector, the vector potential must be of rank 1 by the quotient rule. ■

The quotient rule is a substitute for the illegal division of tensors.

Exercises

2.8.1 The double summation $K_{ij}A_iB_j$ is invariant for any two vectors A_i and B_j. Prove that K_{ij} is a second-rank tensor.
Note. In the form ds^2 (invariant) $= g_{ij}\,dx^i\,dx^j$, this result shows that the matrix g_{ij} is a tensor.

2.8.2 The equation $K_{ij}A_{jk} = B_{ik}$ holds for all orientations of the coordinate system. If $\mathbf{A}$ and $\mathbf{B}$ are arbitrary second-rank tensors, show that $\mathbf{K}$ is a second-rank tensor also.

2.8.3 The exponential in a plane wave is $\exp[i(\mathbf{k}\cdot\mathbf{r} - \omega t)]$. We recognize $x^\mu = (ct, x_1, x_2, x_3)$ as a prototype vector in Minkowski space. If $\mathbf{k}\cdot\mathbf{r} - \omega t$ is a scalar under Lorentz transformations (Section 4.5), show that $k^\mu = (\omega/c, k_1, k_2, k_3)$ is a vector in Minkowski space.
Note. Multiplication by $\hbar$ yields $(E/c, \mathbf{p})$ as a vector in Minkowski space.

2.9 PSEUDOTENSORS, DUAL TENSORS

So far our coordinate transformations have been restricted to pure passive rotations. We now consider the effect of reflections or inversions.

[12] We might, for instance, take $A'_1 = 1$ and $A'_m = 0$ for $m \neq 1$. Then the equation $K'_{i1} = a_{ik}a_{1l}K_{kl}$ follows immediately. The rest of Eq. (2.87) comes from other special choices of the arbitrary A'_j.

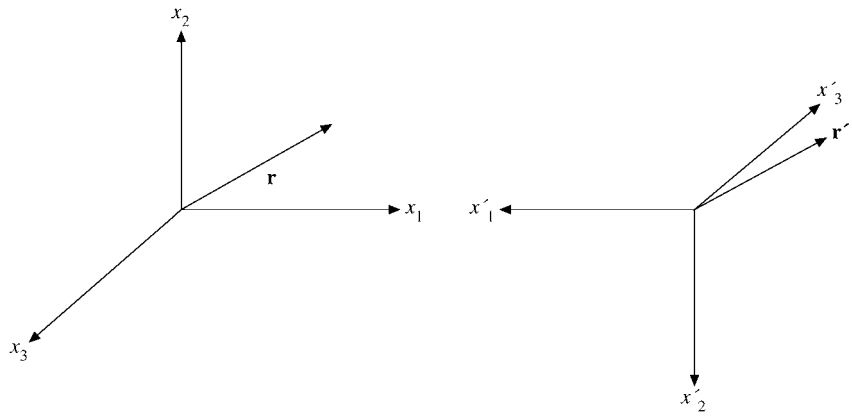

FIGURE 2.9 Inversion of Cartesian coordinates — polar vector.

If we have transformation coefficients $a_{ij} = -\delta_{ij}$, then by Eq. (2.60)

$$x^i = -x'^i, \tag{2.88}$$

which is an inversion or parity transformation. Note that this transformation changes our initial right-handed coordinate system into a left-handed coordinate system.[13] Our prototype vector $\mathbf{r}$ with components (x^1, x^2, x^3) transforms to

$$\mathbf{r}' = \left(x'^1, x'^2, x'^3\right) = \left(-x^1, -x^2, -x^3\right).$$

This new vector $\mathbf{r}'$ has negative components, relative to the new transformed set of axes. As shown in Fig. 2.9, reversing the directions of the coordinate axes and changing the signs of the components gives $\mathbf{r}' = \mathbf{r}$. The vector (an arrow in space) stays exactly as it was before the transformation was carried out. The position vector $\mathbf{r}$ and all other vectors whose components behave this way (reversing sign with a reversal of the coordinate axes) are called **polar vectors** and have odd parity.

A fundamental difference appears when we encounter a vector defined as the cross product of two polar vectors. Let $\mathbf{C} = \mathbf{A} \times \mathbf{B}$, where both $\mathbf{A}$ and $\mathbf{B}$ are polar vectors. From Eq. (1.33), the components of $\mathbf{C}$ are given by

$$C^1 = A^2 B^3 - A^3 B^2 \tag{2.89}$$

and so on. Now, when the coordinate axes are inverted, $A^i \to -A'^i$, $B_j \to -B'_j$, but from its definition $C^k \to +C'^k$; that is, our cross-product vector, vector $\mathbf{C}$, does **not** behave like a polar vector under inversion. To distinguish, we label it a pseudovector or axial vector (see Fig. 2.10) that has even parity. The term **axial vector** is frequently used because these cross products often arise from a description of rotation.

[13] This is an inversion of the coordinate system or coordinate axes, objects in the physical world remaining fixed.

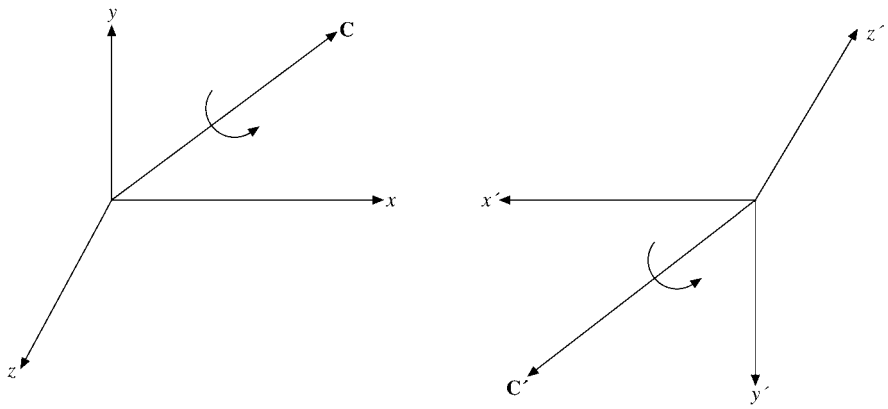

FIGURE 2.10 Inversion of Cartesian coordinates — axial vector.

Examples are

angular velocity,	$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$,
orbital angular momentum,	$\mathbf{L} = \mathbf{r} \times \mathbf{p}$,
torque, force $= \mathbf{F}$,	$\mathbf{N} = \mathbf{r} \times \mathbf{F}$,
magnetic induction field $\mathbf{B}$,	$\dfrac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}$.

In $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, the axial vector is the angular velocity $\boldsymbol{\omega}$, and $\mathbf{r}$ and $\mathbf{v} = d\mathbf{r}/dt$ are polar vectors. Clearly, axial vectors occur frequently in physics, although this fact is usually not pointed out. In a right-handed coordinate system an axial vector $\mathbf{C}$ has a sense of rotation associated with it given by a right-hand rule (compare Section 1.4). In the inverted left-handed system the sense of rotation is a left-handed rotation. This is indicated by the curved arrows in Fig. 2.10.

The distinction between polar and axial vectors may also be illustrated by a reflection. A polar vector reflects in a mirror like a real physical arrow, Fig. 2.11a. In Figs. 2.9 and 2.10 the coordinates are inverted; the physical world remains fixed. Here the coordinate axes remain fixed; the world is reflected — as in a mirror in the xz-plane. Specifically, in this representation we keep the axes fixed and associate a change of sign with the component of the vector. For a mirror in the xz-plane, $P_y \to -P_y$. We have

$$\mathbf{P} = (P_x, P_y, P_z)$$
$$\mathbf{P}' = (P_x, -P_y, P_z) \qquad \text{polar vector.}$$

An axial vector such as a magnetic field $\mathbf{H}$ or a magnetic moment $\boldsymbol{\mu}$ (= current × area of current loop) behaves quite differently under reflection. Consider the magnetic field $\mathbf{H}$ and magnetic moment $\boldsymbol{\mu}$ to be produced by an electric charge moving in a circular path (Exercise 5.8.4 and Example 12.5.3). Reflection reverses the sense of rotation of the charge.

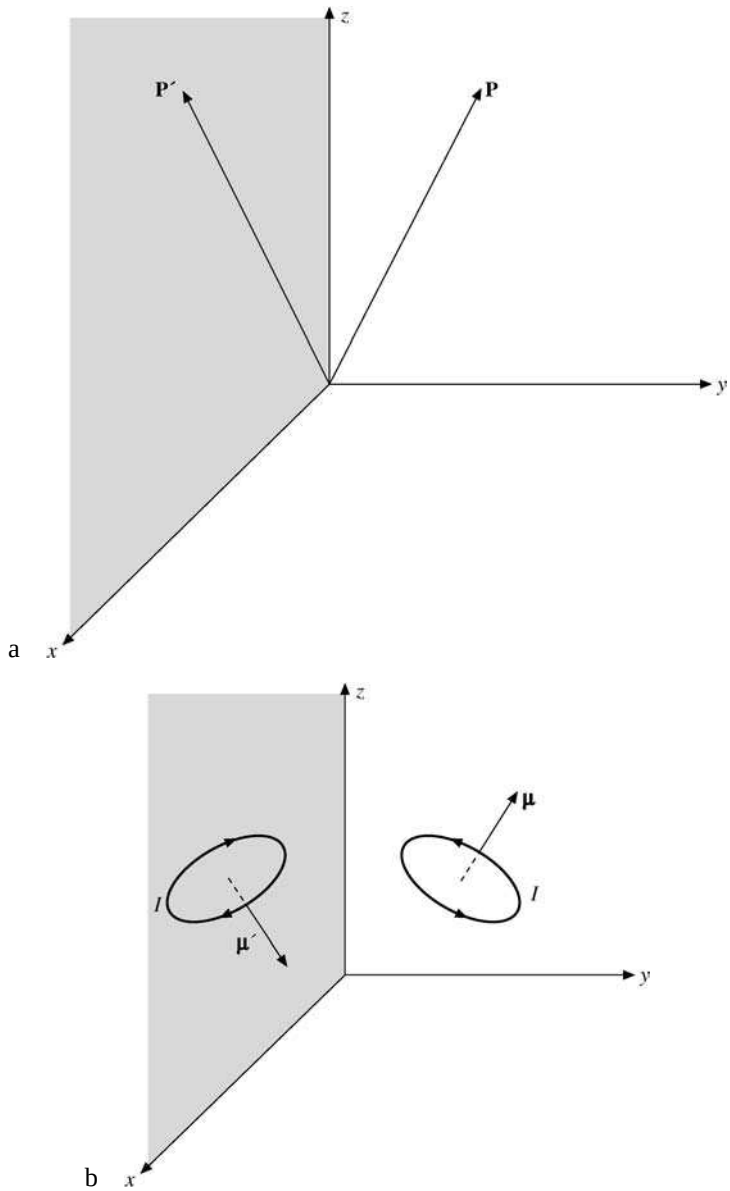

FIGURE 2.11 (a) Mirror in xz-plane; (b) mirror in xz-plane.

The two current loops and the resulting magnetic moments are shown in Fig. 2.11b. We have

$$\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z)$$

$$\boldsymbol{\mu}' = (-\mu_x, \mu_y, -\mu_z) \quad \text{reflected axial vector.}$$

If we agree that the universe does not care whether we use a right- or left-handed coordinate system, then it does not make sense to add an axial vector to a polar vector. In the vector equation $\mathbf{A} = \mathbf{B}$, both $\mathbf{A}$ and $\mathbf{B}$ are either polar vectors or axial vectors.[14] Similar restrictions apply to scalars and pseudoscalars and, in general, to the tensors and pseudotensors considered subsequently.

Usually, pseudoscalars, pseudovectors, and pseudotensors will transform as

$$S' = JS, \qquad C_i' = Ja_{ij}C_j, \qquad A_{ij}' = Ja_{ik}a_{jl}A_{kl}, \tag{2.90}$$

where J is the determinant[15] of the array of coefficients a_{mn}, the Jacobian of the parity transformation. In our inversion the Jacobian is

$$J = \begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix} = -1. \tag{2.91}$$

For a reflection of one axis, the x-axis,

$$J = \begin{vmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -1, \tag{2.92}$$

and again the Jacobian $J = -1$. On the other hand, for all pure rotations, the Jacobian J is always $+1$. Rotation matrices discussed further in Section 3.3.

In Chapter 1 the triple scalar product $S = \mathbf{A} \times \mathbf{B} \cdot \mathbf{C}$ was shown to be a scalar (under rotations). Now by considering the parity transformation given by Eq. (2.88), we see that $S \to -S$, proving that the triple scalar product is actually a pseudoscalar: This behavior was foreshadowed by the geometrical analogy of a volume. If all three parameters of the volume — length, depth, and height — change from positive distances to negative distances, the product of the three will be negative.

Levi-Civita Symbol

For future use it is convenient to introduce the three-dimensional Levi-Civita symbol ε_{ijk}, defined by

$$\begin{aligned} \varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1, \\ \varepsilon_{132} = \varepsilon_{213} = \varepsilon_{321} = -1, \\ \text{all other } \varepsilon_{ijk} = 0. \end{aligned} \tag{2.93}$$

Note that ε_{ijk} is antisymmetric with respect to all pairs of indices. Suppose now that we have a third-rank pseudotensor δ_{ijk}, which in one particular coordinate system is equal to ε_{ijk}. Then

$$\delta_{ijk}' = |a| a_{ip}a_{jq}a_{kr}\varepsilon_{pqr} \tag{2.94}$$

[14] The big exception to this is in beta decay, weak interactions. Here the universe distinguishes between right- and left-handed systems, and we add polar and axial vector interactions.

[15] Determinants are described in Section 3.1.

by definition of pseudotensor. Now,

$$a_{1p}a_{2q}a_{3r}\varepsilon_{pqr} = |a| \tag{2.95}$$

by direct expansion of the determinant, showing that $\delta'_{123} = |a|^2 = 1 = \varepsilon_{123}$. Considering the other possibilities one by one, we find

$$\delta'_{ijk} = \varepsilon_{ijk} \tag{2.96}$$

for rotations and reflections. Hence ε_{ijk} is a pseudotensor.[16,17] Furthermore, it is seen to be an isotropic pseudotensor with the same components in all rotated Cartesian coordinate systems.

Dual Tensors

With any **antisymmetric** second-rank tensor **C** (in three-dimensional space) we may associate a dual pseudovector C_i defined by

$$C_i = \frac{1}{2}\varepsilon_{ijk}C^{jk}. \tag{2.97}$$

Here the antisymmetric **C** may be written

$$\mathbf{C} = \begin{pmatrix} 0 & C^{12} & -C^{31} \\ -C^{12} & 0 & C^{23} \\ C^{31} & -C^{23} & 0 \end{pmatrix}. \tag{2.98}$$

We know that C_i must transform as a vector under rotations from the double contraction of the fifth-rank (pseudo) tensor $\varepsilon_{ijk}C_{mn}$ but that it is really a pseudovector from the pseudo nature of ε_{ijk}. Specifically, the components of **C** are given by

$$(C_1, C_2, C_3) = \left(C^{23}, C^{31}, C^{12}\right). \tag{2.99}$$

Notice the cyclic order of the indices that comes from the cyclic order of the components of ε_{ijk}. Eq. (2.99) means that our three-dimensional vector product may literally be taken to be either a pseudovector or an antisymmetric second-rank tensor, depending on how we choose to write it out.

If we take three (polar) vectors **A**, **B**, and **C**, we may define the direct product

$$V^{ijk} = A^i B^j C^k. \tag{2.100}$$

By an extension of the analysis of Section 2.6, V^{ijk} is a tensor of third rank. The dual quantity

$$V = \frac{1}{3!}\varepsilon_{ijk}V^{ijk} \tag{2.101}$$

[16]The usefulness of ε_{pqr} extends far beyond this section. For instance, the matrices M_k of Exercise 3.2.16 are derived from $(M_r)_{pq} = -i\varepsilon_{pqr}$. Much of elementary vector analysis can be written in a very compact form by using ε_{ijk} and the identity of Exercise 2.9.4 See A. A. Evett, Permutation symbol approach to elementary vector analysis. *Am. J. Phys.* **34**: 503 (1966).

[17]The numerical value of ε_{pqr} is given by the triple scalar product of coordinate unit vectors:

$$\hat{\mathbf{x}}_p \cdot \hat{\mathbf{x}}_q \times \hat{\mathbf{x}}_r.$$

From this point of view each element of ε_{pqr} is a pseudoscalar, but the ε_{pqr} collectively form a third-rank pseudotensor.

is clearly a pseudoscalar. By expansion it is seen that

$$V = \begin{vmatrix} A^1 & B^1 & C^1 \\ A^2 & B^2 & C^2 \\ A^3 & B^3 & C^3 \end{vmatrix} \tag{2.102}$$

is our familiar triple scalar product.

For use in writing Maxwell's equations in covariant form, Section 4.6, we want to extend this dual vector analysis to four-dimensional space and, in particular, to indicate that the four-dimensional volume element $dx^0\,dx^1\,dx^2\,dx^3$ is a pseudoscalar.

We introduce the Levi-Civita symbol ε_{ijkl}, the four-dimensional analog of ε_{ijk}. This quantity ε_{ijkl} is defined as totally antisymmetric in all four indices. If $(ijkl)$ is an even permutation[18] of (0, 1, 2, 3), then ε_{ijkl} is defined as $+1$; if it is an odd permutation, then ε_{ijkl} is -1, and 0 if any two indices are equal. The Levi-Civita ε_{ijkl} may be proved a pseudotensor of rank 4 by analysis similar to that used for establishing the tensor nature of ε_{ijk}. Introducing the direct product of four vectors as fourth-rank tensor with components

$$H^{ijkl} = A^i B^j C^k D^l, \tag{2.103}$$

built from the polar vectors **A**, **B**, **C**, and **D**, we may define the dual quantity

$$H = \frac{1}{4!}\varepsilon_{ijkl} H^{ijkl}, \tag{2.104}$$

a pseudoscalar due to the quadruple contraction with the pseudotensor ε_{ijkl}. Now we let **A**, **B**, **C**, and **D** be infinitesimal displacements along the four coordinate axes (Minkowski space),

$$\begin{aligned} \mathbf{A} &= \left(dx^0, 0, 0, 0\right) \\ \mathbf{B} &= \left(0, dx^1, 0, 0\right), \qquad \text{and so on,} \end{aligned} \tag{2.105}$$

and

$$H = dx^0\,dx^1\,dx^2\,dx^3. \tag{2.106}$$

The four-dimensional volume element is now identified as a pseudoscalar. We use this result in Section 4.6. This result could have been expected from the results of the special theory of relativity. The Lorentz–Fitzgerald contraction of $dx^1\,dx^2\,dx^3$ just balances the time dilation of dx^0.

We slipped into this four-dimensional space as a simple mathematical extension of the three-dimensional space and, indeed, we could just as easily have discussed 5-, 6-, or N-dimensional space. This is typical of the power of the component analysis. Physically, this four-dimensional space may be taken as Minkowski space,

$$\left(x^0, x^1, x^2, x^3\right) = (ct, x, y, z), \tag{2.107}$$

where t is time. This is the merger of space and time achieved in special relativity. The transformations that describe the rotations in four-dimensional space are the Lorentz transformations of special relativity. We encounter these Lorentz transformations in Section 4.6.

[18] A permutation is odd if it involves an odd number of interchanges of adjacent indices, such as (0 1 2 3) → (0 2 1 3). Even permutations arise from an even number of transpositions of adjacent indices. (Actually the word *adjacent* is unnecessary.) $\varepsilon_{0123} = +1$.

Irreducible Tensors

For some applications, particularly in the quantum theory of angular momentum, our Cartesian tensors are not particularly convenient. In mathematical language our general second-rank tensor A_{ij} is reducible, which means that it can be decomposed into parts of lower tensor rank. In fact, we have already done this. From Eq. (2.78),

$$A = A^i{}_i \tag{2.108}$$

is a scalar quantity, the trace of A_{ij}.[19]

The antisymmetric portion,

$$B_{ij} = \tfrac{1}{2}(A_{ij} - A_{ji}), \tag{2.109}$$

has just been shown to be equivalent to a (pseudo) vector, or

$$B_{ij} = C_k \qquad \text{cyclic permutation of } i, j, k. \tag{2.110}$$

By subtracting the scalar A and the vector C_k from our original tensor, we have an irreducible, symmetric, zero-trace second-rank tensor, S_{ij}, in which

$$S_{ij} = \tfrac{1}{2}(A_{ij} + A_{ji}) - \tfrac{1}{3}A\delta_{ij}, \tag{2.111}$$

with five independent components. Then, finally, our original Cartesian tensor may be written

$$A_{ij} = \tfrac{1}{3}A\delta_{ij} + C_k + S_{ij}. \tag{2.112}$$

The three quantities A, C_k, and S_{ij} form spherical tensors of rank 0, 1, and 2, respectively, transforming like the spherical harmonics Y_L^M (Chapter 12) for $L = 0$, 1, and 2. Further details of such spherical tensors and their uses will be found in Chapter 4 and the books by Rose and Edmonds cited there.

A specific example of the preceding reduction is furnished by the symmetric electric quadrupole tensor

$$Q_{ij} = \int \left(3x_i x_j - r^2\delta_{ij}\right)\rho(x_1, x_2, x_3)\, d^3x.$$

The $-r^2\delta_{ij}$ term represents a subtraction of the scalar trace (the three $i = j$ terms). The resulting Q_{ij} has zero trace.

Exercises

2.9.1 An antisymmetric square array is given by

$$\begin{pmatrix} 0 & C_3 & -C_2 \\ -C_3 & 0 & C_1 \\ C_2 & -C_1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & C^{12} & C^{13} \\ -C^{12} & 0 & C^{23} \\ -C^{13} & -C^{23} & 0 \end{pmatrix},$$

[19] An alternate approach, using matrices, is given in Section 3.3 (see Exercise 3.3.9).

where (C_1, C_2, C_3) form a pseudovector. Assuming that the relation

$$C_i = \frac{1}{2!}\varepsilon_{ijk}C^{jk}$$

holds in all coordinate systems, prove that C^{jk} is a tensor. (This is another form of the quotient theorem.)

2.9.2 Show that the vector product is unique to three-dimensional space; that is, only in three dimensions can we establish a one-to-one correspondence between the components of an antisymmetric tensor (second-rank) and the components of a vector.

2.9.3 Show that in $\mathbb{R}^3$

(a) $\delta_{ii} = 3$,
(b) $\delta_{ij}\varepsilon_{ijk} = 0$,
(c) $\varepsilon_{ipq}\varepsilon_{jpq} = 2\delta_{ij}$,
(d) $\varepsilon_{ijk}\varepsilon_{ijk} = 6$.

2.9.4 Show that in $\mathbb{R}^3$

$$\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}.$$

2.9.5 (a) Express the components of a cross-product vector $\mathbf{C}$, $\mathbf{C} = \mathbf{A} \times \mathbf{B}$, in terms of ε_{ijk} and the components of $\mathbf{A}$ and $\mathbf{B}$.
(b) Use the antisymmetry of ε_{ijk} to show that $\mathbf{A} \cdot \mathbf{A} \times \mathbf{B} = 0$.

ANS. (a) $C_i = \varepsilon_{ijk}A_jB_k$.

2.9.6 (a) Show that the inertia tensor (matrix) may be written

$$I_{ij} = m(x_i x_j \delta_{ij} - x_i x_j)$$

for a particle of mass m at (x_1, x_2, x_3).
(b) Show that

$$I_{ij} = -M_{il}M_{lj} = -m\varepsilon_{ilk}x_k\varepsilon_{ljm}x_m,$$

where $M_{il} = m^{1/2}\varepsilon_{ilk}x_k$. This is the contraction of two second-rank tensors and is identical with the matrix product of Section 3.2.

2.9.7 Write $\nabla \cdot \nabla \times \mathbf{A}$ and $\nabla \times \nabla\varphi$ in tensor (index) notation in $\mathbb{R}^3$ so that it becomes obvious that each expression vanishes.

ANS. $\nabla \cdot \nabla \times \mathbf{A} = \varepsilon_{ijk}\dfrac{\partial}{\partial x^i}\dfrac{\partial}{\partial x^j}A^k$,

$(\nabla \times \nabla\varphi)_i = \varepsilon_{ijk}\dfrac{\partial}{\partial x^j}\dfrac{\partial}{\partial x^k}\varphi$.

2.9.8 Expressing cross products in terms of Levi-Civita symbols (ε_{ijk}), derive the *BAC–CAB* rule, Eq. (1.55).
Hint. The relation of Exercise 2.9.4 is helpful.

2.9.9 Verify that each of the following fourth-rank tensors is isotropic, that is, that it has the same form independent of any rotation of the coordinate systems.

(a) $A_{ijkl} = \delta_{ij}\delta_{kl}$,
(b) $B_{ijkl} = \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}$,
(c) $C_{ijkl} = \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}$.

2.9.10 Show that the two-index Levi-Civita symbol ε_{ij} is a second-rank pseudotensor (in two-dimensional space). Does this contradict the uniqueness of δ_{ij} (Exercise 2.6.4)?

2.9.11 Represent ε_{ij} by a 2×2 matrix, and using the 2×2 rotation matrix of Section 3.3 show that ε_{ij} is invariant under orthogonal similarity transformations.

2.9.12 Given $A_k = \frac{1}{2}\varepsilon_{ijk}B^{ij}$ with $B^{ij} = -B^{ji}$, antisymmetric, show that

$$B^{mn} = \varepsilon^{mnk}A_k.$$

2.9.13 Show that the vector identity

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{A} \cdot \mathbf{D})(\mathbf{B} \cdot \mathbf{C})$$

(Exercise 1.5.12) follows directly from the description of a cross product with ε_{ijk} and the identity of Exercise 2.9.4.

2.9.14 Generalize the cross product of two vectors to n-dimensional space for $n = 4, 5, \ldots$. Check the consistency of your construction and discuss concrete examples. See Exercise 1.4.17 for the case $n = 2$.

2.10 General Tensors

The distinction between contravariant and covariant transformations was established in Section 2.6. Then, for convenience, we restricted our attention to Cartesian coordinates (in which the distinction disappears). Now in these two concluding sections we return to non-Cartesian coordinates and resurrect the contravariant and covariant dependence. As in Section 2.6, a superscript will be used for an index denoting contravariant and a subscript for an index denoting covariant dependence. The metric tensor of Section 2.1 will be used to relate contravariant and covariant indices.

The emphasis in this section is on differentiation, culminating in the construction of the **covariant derivative**. We saw in Section 2.7 that the derivative of a vector yields a second-rank tensor — in Cartesian coordinates. In non-Cartesian coordinate systems, it is the covariant derivative of a vector rather than the ordinary derivative that yields a second-rank tensor by differentiation of a vector.

Metric Tensor

Let us start with the transformation of vectors from one set of coordinates (q^1, q^2, q^3) to another $\mathbf{r} = (x^1, x^2, x^3)$. The new coordinates are (in general **nonlinear**) functions

$x^i(q^1, q^2, q^3)$ of the old, such as spherical polar coordinates (r, θ, ϕ). But their **differentials** obey the **linear** transformation law

$$dx^i = \frac{\partial x^i}{\partial q^j} dq^j, \tag{2.113a}$$

or

$$d\mathbf{r} = \boldsymbol{\varepsilon}_j dq^j \tag{2.113b}$$

in vector notation. For convenience we take the basis vectors $\boldsymbol{\varepsilon}_1 = (\frac{\partial x^1}{\partial q^1}, \frac{\partial x^1}{\partial q^2}, \frac{\partial x^1}{\partial q^3})$, $\boldsymbol{\varepsilon}_2$, and $\boldsymbol{\varepsilon}_3$ to form a right-handed set. These vectors are not necessarily orthogonal. Also, a limitation to three-dimensional space will be required only for the discussions of cross products and curls. Otherwise these $\boldsymbol{\varepsilon}_i$ may be in N-dimensional space, including the four-dimensional space–time of special and general relativity. The basis vectors $\boldsymbol{\varepsilon}_i$ may be expressed by

$$\boldsymbol{\varepsilon}_i = \frac{\partial \mathbf{r}}{\partial q^i}, \tag{2.114}$$

as in Exercise 2.2.3. Note, however, that the $\boldsymbol{\varepsilon}_i$ here do **not** necessarily have unit magnitude. From Exercise 2.2.3, the unit vectors are

$$\mathbf{e}_i = \frac{1}{h_i} \frac{\partial \mathbf{r}}{\partial q_i} \qquad \text{(no summation)},$$

and therefore

$$\boldsymbol{\varepsilon}_i = h_i \mathbf{e}_i \qquad \text{(no summation)}. \tag{2.115}$$

The $\boldsymbol{\varepsilon}_i$ are related to the unit vectors $\mathbf{e}_i$ by the scale factors h_i of Section 2.2. The $\mathbf{e}_i$ have no dimensions; the $\boldsymbol{\varepsilon}_i$ have the dimensions of h_i. In spherical polar coordinates, as a specific example,

$$\boldsymbol{\varepsilon}_r = \mathbf{e}_r = \hat{\mathbf{r}}, \qquad \boldsymbol{\varepsilon}_\theta = r\mathbf{e}_\theta = r\hat{\boldsymbol{\theta}}, \qquad \boldsymbol{\varepsilon}_\varphi = r\sin\theta \mathbf{e}_\varphi = r\sin\theta \hat{\boldsymbol{\varphi}}. \tag{2.116}$$

In Euclidean spaces, or in Minkowski space of special relativity, the partial derivatives in Eq. (2.113) are constants that define the new coordinates in terms of the old ones. We used them to define the transformation laws of vectors in Eq. (2.59) and (2.62) and tensors in Eq. (2.66). Generalizing, we define a **contravariant** vector V^i under **general** coordinate transformations if its components transform according to

$$V'^i = \frac{\partial x^i}{\partial q^j} V^j, \tag{2.117a}$$

or

$$\mathbf{V}' = V^j \boldsymbol{\varepsilon}_j \tag{2.117b}$$

in vector notation. For **covariant** vectors we inspect the transformation of the gradient operator

$$\frac{\partial}{\partial x^i} = \frac{\partial q^j}{\partial x^i} \frac{\partial}{\partial q^j} \tag{2.118}$$

using the chain rule. From

$$\frac{\partial x^i}{\partial q^j}\frac{\partial q^j}{\partial x^k} = \delta^i{}_k \tag{2.119}$$

it is clear that Eq. (2.118) is related to the **inverse** transformation of Eq. (2.113),

$$dq^j = \frac{\partial q^j}{\partial x^i} dx^i. \tag{2.120}$$

Hence we define a **covariant** vector V_i if

$$V_i' = \frac{\partial q^j}{\partial x^i} V_j \tag{2.121a}$$

holds or, in vector notation,

$$\mathbf{V}' = V_j \boldsymbol{\varepsilon}^j, \tag{2.121b}$$

where $\boldsymbol{\varepsilon}^j$ are the contravariant vectors $g^{ji}\boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}^j$.

Second-rank tensors are defined as in Eq. (2.66),

$$A'^{ij} = \frac{\partial x^i}{\partial q^k}\frac{\partial x^j}{\partial q^l} A^{kl}, \tag{2.122}$$

and tensors of higher rank similarly.

As in Section 2.1, we construct the square of a differential displacement

$$(ds)^2 = d\mathbf{r}\cdot d\mathbf{r} = \left(\boldsymbol{\varepsilon}_i\, dq^i\right)^2 = \boldsymbol{\varepsilon}_i \cdot \boldsymbol{\varepsilon}_j\, dq^i\, dq^j. \tag{2.123}$$

Comparing this with $(ds)^2$ of Section 2.1, Eq. (2.5), we identify $\boldsymbol{\varepsilon}_i \cdot \boldsymbol{\varepsilon}_j$ as the covariant metric tensor

$$\boldsymbol{\varepsilon}_i \cdot \boldsymbol{\varepsilon}_j = g_{ij}. \tag{2.124}$$

Clearly, g_{ij} is symmetric. The tensor nature of g_{ij} follows from the quotient rule, Exercise 2.8.1. We take the relation

$$g^{ik} g_{kj} = \delta^i{}_j \tag{2.125}$$

to define the corresponding contravariant tensor g^{ik}. Contravariant g^{ik} enters as the inverse[20] of covariant g_{kj}. We use this contravariant g^{ik} to raise indices, converting a covariant index into a contravariant index, as shown subsequently. Likewise the covariant g_{kj} will be used to lower indices. The choice of g^{ik} and g_{kj} for this raising–lowering operation is arbitrary. Any second-rank tensor (and its inverse) would do. Specifically, we have

$$\begin{aligned} g^{ij}\boldsymbol{\varepsilon}_j &= \boldsymbol{\varepsilon}^i && \text{relating covariant and contravariant basis vectors,} \\ g^{ij}F_j &= F^i && \text{relating covariant and contravariant vector components.} \end{aligned} \tag{2.126}$$

[20] If the tensor g_{kj} is written as a matrix, the tensor g^{ik} is given by the inverse matrix.

Then

$$\begin{aligned} g_{ij}\boldsymbol{\varepsilon}^j &= \boldsymbol{\varepsilon}_i \qquad \text{as the corresponding index} \\ g_{ij}F^j &= F_i \qquad \text{lowering relations.} \end{aligned} \tag{2.127}$$

It should be emphasized again that the $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}^j$ do **not** have unit magnitude. This may be seen in Eqs. (2.116) and in the metric tensor g_{ij} for spherical polar coordinates and its inverse g^{ij}:

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix} \qquad \left(g^{ij}\right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \dfrac{1}{r^2} & 0 \\ 0 & 0 & \dfrac{1}{r^2\sin^2\theta} \end{pmatrix}.$$

Christoffel Symbols

Let us form the differential of a scalar ψ,

$$d\psi = \frac{\partial\psi}{\partial q^i}\,dq^i. \tag{2.128}$$

Since the dq^i are the components of a contravariant vector, the partial derivatives $\partial\psi/\partial q^i$ must form a covariant vector — by the quotient rule. The gradient of a scalar becomes

$$\nabla\psi = \frac{\partial\psi}{\partial q^i}\boldsymbol{\varepsilon}^i. \tag{2.129}$$

Note that $\partial\psi/\partial q^i$ are not the gradient components of Section 2.2 — because $\boldsymbol{\varepsilon}^i \neq \mathbf{e}_i$ of Section 2.2.

Moving on to the derivatives of a vector, we find that the situation is much more complicated because the basis vectors $\boldsymbol{\varepsilon}_i$ are in general not constant. Remember, we are no longer restricting ourselves to Cartesian coordinates and the nice, convenient $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$! Direct differentiation of Eq. (2.117a) yields

$$\frac{\partial V'^k}{\partial q^j} = \frac{\partial x^k}{\partial q^i}\frac{\partial V^i}{\partial q^j} + \frac{\partial^2 x^k}{\partial q^j\partial q^i}V^i, \tag{2.130a}$$

or, in vector notation,

$$\frac{\partial\mathbf{V}'}{\partial q^j} = \frac{\partial V^i}{\partial q^j}\boldsymbol{\varepsilon}_i + V^i\frac{\partial\boldsymbol{\varepsilon}_i}{\partial q^j}. \tag{2.130b}$$

The right side of Eq. (2.130a) differs from the transformation law for a second-rank mixed tensor by the second term, which contains second derivatives of the coordinates x^k. The latter are nonzero for nonlinear coordinate transformations.

Now, $\partial\boldsymbol{\varepsilon}_i/\partial q^j$ will be some linear combination of the $\boldsymbol{\varepsilon}_k$, with the coefficient depending on the indices i and j from the partial derivative and index k from the base vector. We write

$$\frac{\partial\boldsymbol{\varepsilon}_i}{\partial q^j} = \Gamma^k_{ij}\boldsymbol{\varepsilon}_k. \tag{2.131a}$$

Multiplying by $\boldsymbol{\varepsilon}^m$ and using $\boldsymbol{\varepsilon}^m \cdot \boldsymbol{\varepsilon}_k = \delta_k^m$ from Exercise 2.10.2, we have

$$\Gamma_{ij}^m = \boldsymbol{\varepsilon}^m \cdot \frac{\partial \boldsymbol{\varepsilon}_i}{\partial q^j}. \tag{2.131b}$$

The Γ_{ij}^k is a Christoffel symbol of the **second kind**. It is also called a *coefficient of connection*. These Γ_{ij}^k are **not** third-rank tensors and the $\partial V^i/\partial q^j$ of Eq. (2.130a) are not second-rank tensors. Equations (2.131) should be compared with the results quoted in Exercise 2.2.3 (remembering that in general $\boldsymbol{\varepsilon}_i \neq \mathbf{e}_i$). In Cartesian coordinates, $\Gamma_{ij}^k = 0$ for all values of the indices i, j, and k. These Christoffel three-index symbols may be computed by the techniques of Section 2.2. This is the topic of Exercise 2.10.8. Equation (2.138) offers an easier method. Using Eq. (2.114), we obtain

$$\frac{\partial \boldsymbol{\varepsilon}_i}{\partial q^j} = \frac{\partial^2 \mathbf{r}}{\partial q^j \, \partial q^i} = \frac{\partial \boldsymbol{\varepsilon}_j}{\partial q^i} = \Gamma_{ji}^k \boldsymbol{\varepsilon}_k. \tag{2.132}$$

Hence these Christoffel symbols are symmetric in the two lower indices:

$$\Gamma_{ij}^k = \Gamma_{ji}^k. \tag{2.133}$$

Christoffel Symbols as Derivatives of the Metric Tensor

It is often convenient to have an explicit expression for the Christoffel symbols in terms of derivatives of the metric tensor. As an initial step, we define the Christoffel symbol of the **first kind** $[ij, k]$ by

$$[ij, k] \equiv g_{mk} \Gamma_{ij}^m, \tag{2.134}$$

from which the symmetry $[ij, k] = [ji, k]$ follows. Again, this $[ij, k]$ is not a third-rank tensor. From Eq. (2.131b),

$$\begin{aligned} [ij, k] &= g_{mk} \boldsymbol{\varepsilon}^m \cdot \frac{\partial \boldsymbol{\varepsilon}_i}{\partial q^j} \\ &= \boldsymbol{\varepsilon}_k \cdot \frac{\partial \boldsymbol{\varepsilon}_i}{\partial q^j}. \end{aligned} \tag{2.135}$$

Now we differentiate $g_{ij} = \boldsymbol{\varepsilon}_i \cdot \boldsymbol{\varepsilon}_j$, Eq. (2.124):

$$\begin{aligned} \frac{\partial g_{ij}}{\partial q^k} &= \frac{\partial \boldsymbol{\varepsilon}_i}{\partial q^k} \cdot \boldsymbol{\varepsilon}_j + \boldsymbol{\varepsilon}_i \cdot \frac{\partial \boldsymbol{\varepsilon}_j}{\partial q^k} \\ &= [ik, j] + [jk, i] \end{aligned} \tag{2.136}$$

by Eq. (2.135). Then

$$[ij, k] = \frac{1}{2} \left\{ \frac{\partial g_{ik}}{\partial q^j} + \frac{\partial g_{jk}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^k} \right\}, \tag{2.137}$$

and

$$\begin{aligned} \Gamma_{ij}^s &= g^{ks} [ij, k] \\ &= \frac{1}{2} g^{ks} \left\{ \frac{\partial g_{ik}}{\partial q^j} + \frac{\partial g_{jk}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^k} \right\}. \end{aligned} \tag{2.138}$$

These Christoffel symbols are applied in the next section.

Covariant Derivative

With the Christoffel symbols, Eq. (2.130b) may be rewritten

$$\frac{\partial \mathbf{V}'}{\partial q^j} = \frac{\partial V^i}{\partial q^j}\boldsymbol{\varepsilon}_i + V^i \Gamma^k_{ij}\boldsymbol{\varepsilon}_k. \tag{2.139}$$

Now, i and k in the last term are dummy indices. Interchanging i and k (in this one term), we have

$$\frac{\partial \mathbf{V}'}{\partial q^j} = \left(\frac{\partial V^i}{\partial q^j} + V^k \Gamma^i_{kj}\right)\boldsymbol{\varepsilon}_i. \tag{2.140}$$

The quantity in parenthesis is labeled a **covariant derivative**, $V^i_{;j}$. We have

$$V^i_{;j} \equiv \frac{\partial V^i}{\partial q^j} + V^k \Gamma^i_{kj}. \tag{2.141}$$

The $;j$ subscript indicates differentiation with respect to q^j. The differential $d\mathbf{V}'$ becomes

$$d\mathbf{V}' = \frac{\partial \mathbf{V}'}{\partial q^j}\, dq^j = [V^i_{;j}\, dq^j]\boldsymbol{\varepsilon}_i. \tag{2.142}$$

A comparison with Eq. (2.113) or (2.122) shows that the quantity in square brackets is the ith contravariant component of a vector. Since dq^j is the jth contravariant component of a vector (again, Eq. (2.113)), $V^i_{;j}$ must be the ijth component of a (mixed) second-rank tensor (quotient rule). The covariant derivatives of the contravariant components of a vector form a mixed second-rank tensor, $V^i_{;j}$.

Since the Christoffel symbols vanish in Cartesian coordinates, the covariant derivative and the ordinary partial derivative coincide:

$$\frac{\partial V^i}{\partial q^j} = V^i_{;j} \qquad \text{(Cartesian coordinates).} \tag{2.143}$$

The covariant derivative of a covariant vector V_i is given by (Exercise 2.10.9)

$$V_{i;j} = \frac{\partial V_i}{\partial q^j} - V_k \Gamma^k_{ij}. \tag{2.144}$$

Like $V^i_{;j}$, $V_{i;j}$ is a second-rank tensor.

The physical importance of the covariant derivative is that "A consistent replacement of regular partial derivatives by covariant derivatives carries the laws of physics (in component form) from flat space–time into the curved (Riemannian) space–time of general relativity. Indeed, this substitution may be taken as a mathematical statement of Einstein's principle of equivalence."[21]

[21] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*. San Francisco: W. H. Freeman (1973), p. 387.

Geodesics, Parallel Transport

The covariant derivative of vectors, tensors, and the Christoffel symbols may also be approached from geodesics. A geodesic in Euclidean space is a straight line. In general, it is the curve of shortest length between two points and the curve along which a freely falling particle moves. The ellipses of planets are geodesics around the sun, and the moon is in free fall around the Earth on a geodesic. Since we can throw a particle in any direction, a geodesic can have any direction through a given point. Hence the geodesic equation can be obtained from Fermat's variational principle of optics (see Chapter 17 for Euler's equation),

$$\delta \int ds = 0, \tag{2.145}$$

where ds^2 is the metric, Eq. (2.123), of our space. Using the variation of ds^2,

$$2\,ds\,\delta\,ds = dq^i\,dq^j\,\delta\,g_{ij} + g_{ij}\,dq^i\,\delta\,dq^j + g_{ij}\,dq^j\,\delta\,dq^i \tag{2.146}$$

in Eq. (2.145) yields

$$\frac{1}{2}\int\left[\frac{dq^i}{ds}\frac{dq^j}{ds}\delta g_{ij} + g_{ij}\frac{dq^i}{ds}\frac{d}{ds}\delta\,dq^j + g_{ij}\frac{dq^j}{ds}\frac{d}{ds}\delta\,dq^i\right]ds = 0, \tag{2.147}$$

where ds measures the length on the geodesic. Expressing the variations

$$\delta g_{ij} = \frac{\partial g_{ij}}{\partial q^k}\delta\,dq^k \equiv (\partial_k g_{ij})\delta\,dq^k$$

in terms of the **independent variations** $\delta\,dq^k$, shifting their derivatives in the other two terms of Eq. (2.147) upon integrating by parts, and renaming dummy summation indices, we obtain

$$\frac{1}{2}\int\left[\frac{dq^i}{ds}\frac{dq^j}{ds}\partial_k g_{ij} - \frac{d}{ds}\left(g_{ik}\frac{dq^i}{ds} + g_{kj}\frac{dq^j}{ds}\right)\right]\delta\,dq^k\,ds = 0. \tag{2.148}$$

The integrand of Eq. (2.148), set equal to zero, is the geodesic equation. It is the Euler equation of our variational problem. Upon expanding

$$\frac{dg_{ik}}{ds} = (\partial_j g_{ik})\frac{dq^j}{ds}, \qquad \frac{dg_{kj}}{ds} = (\partial_i g_{kj})\frac{dq^i}{ds} \tag{2.149}$$

along the geodesic we find

$$\frac{1}{2}\frac{dq^i}{ds}\frac{dq^j}{ds}(\partial_k g_{ij} - \partial_j g_{ik} - \partial_i g_{kj}) - g_{ik}\frac{d^2q^i}{ds^2} = 0. \tag{2.150}$$

Multiplying Eq. (2.150) with g^{kl} and using Eq. (2.125), we find the **geodesic** equation

$$\frac{d^2q^l}{ds^2} + \frac{dq^i}{ds}\frac{dq^j}{ds}\frac{1}{2}g^{kl}(\partial_i g_{kj} + \partial_j g_{ik} - \partial_k g_{ij}) = 0, \tag{2.151}$$

where the coefficient of the velocities is the Christoffel symbol Γ^l_{ij} of Eq. (2.138).

Geodesics are curves that are independent of the choice of coordinates. They can be drawn through any point in space in various directions. Since the length ds measured along

the geodesic is a scalar, the velocities dq^i/ds (of a freely falling particle along the geodesic, for example) form a contravariant vector. Hence $V_k\,dq^k/ds$ is a well-defined scalar on any geodesic, which we can differentiate in order to define the covariant derivative of any covariant vector V_k. Using Eq. (2.151) we obtain from the scalar

$$\begin{aligned}\frac{d}{ds}\left(V_k\frac{dq^k}{ds}\right) &= \frac{dV_k}{ds}\frac{dq^k}{ds} + V_k\frac{d^2q^k}{ds^2}\\ &= \frac{\partial V_k}{\partial q^i}\frac{dq^i}{ds}\frac{dq^k}{ds} - V_k\Gamma^k_{ij}\frac{dq^i}{ds}\frac{dq^j}{ds}\\ &= \frac{dq^i}{ds}\frac{dq^k}{ds}\left(\frac{\partial V_k}{\partial q^i} - \Gamma^l_{ik}V_l\right).\end{aligned} \tag{2.152}$$

When the quotient theorem is applied to Eq. (2.152) it tells us that

$$V_{k;i} = \frac{\partial V_k}{\partial q^i} - \Gamma^l_{ik}V_l \tag{2.153}$$

is a covariant tensor that defines the covariant derivative of V_k, consistent with Eq. (2.144). Similarly, higher-order tensors may be derived.

The second term in Eq. (2.153) defines the **parallel transport or displacement**,

$$\delta V_k = \Gamma^l_{ki}V_l\delta q^i, \tag{2.154}$$

of the covariant vector V_k from the point with coordinates q^i to $q^i + \delta q^i$. The parallel transport, δU^k, of a contravariant vector U^k may be found from the invariance of the scalar product U^kV_k under parallel transport,

$$\delta(U^kV_k) = \delta U^kV_k + U^k\delta V_k = 0, \tag{2.155}$$

in conjunction with the quotient theorem.

In summary, when we shift a vector to a neighboring point, parallel transport prevents it from sticking out of our space. This can be clearly seen on the surface of a sphere in spherical geometry, where a tangent vector is supposed to remain a tangent upon translating it along some path on the sphere. This explains why the covariant derivative of a vector or tensor is naturally defined by translating it along a geodesic in the desired direction.

Exercises

2.10.1 Equations (2.115) and (2.116) use the scale factor h_i, citing Exercise 2.2.3. In Section 2.2 we had restricted ourselves to orthogonal coordinate systems, yet Eq. (2.115) holds for nonorthogonal systems. Justify the use of Eq. (2.115) for nonorthogonal systems.

2.10.2 (a) Show that $\boldsymbol{\varepsilon}^i \cdot \boldsymbol{\varepsilon}_j = \delta^i_j$.
(b) From the result of part (a) show that

$$F^i = \mathbf{F}\cdot\boldsymbol{\varepsilon}^i \qquad \text{and} \qquad F_i = \mathbf{F}\cdot\boldsymbol{\varepsilon}_i.$$

CHAPTER 4

GROUP THEORY

Disciplined judgment, about what is neat and symmetrical and elegant has time and time again proved an excellent guide to how nature works

MURRAY GELL-MANN

4.1 INTRODUCTION TO GROUP THEORY

In classical mechanics the **symmetry** of a physical system leads to **conservation laws**. Conservation of angular momentum is a direct consequence of rotational symmetry, which means **invariance** under spatial rotations. In the first third of the 20th century, Wigner and others realized that invariance was a key concept in understanding the new quantum phenomena and in developing appropriate theories. Thus, in quantum mechanics the concept of angular momentum and spin has become even more central. Its generalizations, **isospin** in nuclear physics and the **flavor symmetry** in particle physics, are indispensable tools in building and solving theories. Generalizations of the concept of **gauge invariance** of classical electrodynamics to the isospin symmetry lead to the electroweak gauge theory.

In each case the set of these symmetry operations forms a group. Group theory is the mathematical tool to treat invariants and symmetries. It brings unification and formalization of principles, such as spatial reflections, or parity, angular momentum, and geometry, that are widely used by physicists.

In geometry the fundamental role of group theory was recognized more than a century ago by mathematicians (e.g., Felix Klein's Erlanger Program). In Euclidean geometry the distance between two points, the scalar product of two vectors or metric, does not change under rotations or translations. These symmetries are characteristic of this geometry. In special relativity the metric, or scalar product of four-vectors, differs from that of

Euclidean geometry in that it is no longer positive definite and is invariant under Lorentz transformations.

For a crystal the symmetry group contains only a finite number of rotations at discrete values of angles or reflections. The theory of such **discrete** or **finite** groups, developed originally as a branch of pure mathematics, now is a useful tool for the development of crystallography and condensed matter physics. A brief introduction to this area appears in Section 4.7. When the rotations depend on continuously varying angles (the Euler angles of Section 3.3) the rotation groups have an infinite number of elements. Such continuous (or **Lie**[1]) **groups** are the topic of Sections 4.2–4.6. In Section 4.8 we give an introduction to differential forms, with applications to Maxwell's equations and topics of Chapters 1 and 2, which allows seeing these topics from a different perspective.

Definition of a Group

A group G may be defined as a set of objects or operations, rotations, transformations, called the elements of G, that may be combined, or "multiplied," to form a well-defined product in G, denoted by a *, that satisfies the following four conditions.

1. If a and b are any two elements of G, then the product $a * b$ is also an element of G, where b acts before a; or $(a, b) \to a * b$ associates (or maps) an element $a * b$ of G with the pair (a, b) of elements of G. This property is known as "G is closed under multiplication of its own elements."
2. This multiplication is associative: $(a * b) * c = a * (b * c)$.
3. There is a unit element[2] 1 in G such that $1 * a = a * 1 = a$ for every element a in G. The unit is unique: $1 = 1' * 1 = 1'$.
4. There is an inverse, or reciprocal, of each element a of G, labeled a^{-1}, such that $a * a^{-1} = a^{-1} * a = 1$. The inverse is unique: If a^{-1} and a'^{-1} are both inverses of a, then $a'^{-1} = a'^{-1} * (a * a'^{-1}) = (a'^{-1} * a) * a^{-1} = a^{-1}$.

Since the * for multiplication is tedious to write, it is customary to drop it and simply let it be understood. From now on, we write ab instead of $a * b$.

- If a subset G' of G is closed under multiplication, it is a group and called a **subgroup** of G; that is, G' is closed under the multiplication of G. The unit of G always forms a subgroup of G.
- If $gg'g^{-1}$ is an element of G' for any g of G and g' of G', then G' is called an **invariant subgroup** of G. The subgroup consisting of the unit is invariant. If the group elements are square matrices, then $gg'g^{-1}$ corresponds to a similarity transformation (see Eq. (3.100)).
- If $ab = ba$ for all a, b of G, the group is called **abelian**, that is, the order in products does not matter; commutative multiplication is often denoted by a $+$ sign. Examples are vector spaces whose unit is the zero vector and $-a$ is the inverse of a for all elements a in G.

[1] After the Norwegian mathematician Sophus Lie.

[2] Following E. Wigner, the unit element of a group is often labeled E, from the German **Einheit**, that is, unit, or just 1, or I for identity.

Example 4.1.1 ORTHOGONAL AND UNITARY GROUPS

Orthogonal $n \times n$ matrices form the group $\mathsf{O}(n)$, and $\mathsf{SO}(n)$ if their determinants are $+1$ (S stands for "special"). If $\tilde{\mathsf{O}}_i = \mathsf{O}_i^{-1}$ for $i = 1$ and 2 (see Section 3.3 for orthogonal matrices) are elements of $\mathsf{O}(n)$, then the product

$$\widetilde{\mathsf{O}_1\mathsf{O}_2} = \tilde{\mathsf{O}}_2\tilde{\mathsf{O}}_1 = \mathsf{O}_2^{-1}\mathsf{O}_1^{-1} = (\mathsf{O}_1\mathsf{O}_2)^{-1}$$

is also an orthogonal matrix in $\mathsf{O}(n)$, thus proving closure under (matrix) multiplication. The inverse is the transpose (orthogonal) matrix. The unit of the group is the n-dimensional unit matrix 1_n. A real orthogonal $n \times n$ matrix has $n(n-1)/2$ independent parameters. For $n = 2$, there is only one parameter: one angle. For $n = 3$, there are three independent parameters: the three Euler angles of Section 3.3.

If $\tilde{\mathsf{O}}_i = \mathsf{O}_i^{-1}$ (for $i = 1$ and 2) are elements of $\mathsf{SO}(n)$, then closure requires proving in addition that their product has determinant $+1$, which follows from the product theorem in Chapter 3.

Likewise, unitary $n \times n$ matrices form the group $\mathsf{U}(n)$, and $\mathsf{SU}(n)$ if their determinants are $+1$. If $\mathsf{U}_i^\dagger = \mathsf{U}_i^{-1}$ (see Section 3.4 for unitary matrices) are elements of $\mathsf{U}(n)$, then

$$(\mathsf{U}_1\mathsf{U}_2)^\dagger = \mathsf{U}_2^\dagger\mathsf{U}_1^\dagger = \mathsf{U}_2^{-1}\mathsf{U}_1^{-1} = (\mathsf{U}_1\mathsf{U}_2)^{-1},$$

so the product is unitary and an element of $\mathsf{U}(n)$, thus proving closure under multiplication. Each unitary matrix has an inverse (its Hermitian adjoint), which again is unitary.

If $\mathsf{U}_i^\dagger = \mathsf{U}_i^{-1}$ are elements of $\mathsf{SU}(n)$, then closure requires us to prove that their product also has determinant $+1$, which follows from the product theorem in Chapter 3. ■

- Orthogonal groups are called **Lie groups**; that is, they depend on continuously varying parameters (the Euler angles and their generalization for higher dimensions); they are **compact** because the angles vary over closed, finite intervals (containing the limit of any converging sequence of angles). Unitary groups are also compact. Translations form a noncompact group because the limit of translations with distance $d \to \infty$ is not part of the group. The Lorentz group is not compact either.

Homomorphism, Isomorphism

There may be a correspondence between the elements of two groups: one-to-one, two-to-one, or many-to-one. If this correspondence preserves the group multiplication, we say that the two groups are **homomorphic**. A most important homomorphic correspondence between the rotation group $\mathsf{SO}(3)$ and the unitary group $\mathsf{SU}(2)$ is developed in Section 4.2. If the correspondence is one-to-one, still preserving the group multiplication,[3] then the groups are **isomorphic**.

- If a group G is homomorphic to a group of matrices G', then G' is called a **representation** of G. If G and G' are isomorphic, the representation is called **faithful**. There are many representations of groups; they are not unique.

[3] Suppose the elements of one group are labeled g_i, the elements of a second group h_i. Then $g_i \leftrightarrow h_i$ is a one-to-one correspondence for all values of i. If $g_i g_j = g_k$ and $h_i h_j = h_k$, then g_k and h_k must be the corresponding group elements.

Example 4.1.2 ROTATIONS

Another instructive example for a group is the set of counterclockwise coordinate rotations of three-dimensional Euclidean space about its z-axis. From Chapter 3 we know that such a rotation is described by a linear transformation of the coordinates involving a 3×3 matrix made up of three rotations depending on the Euler angles. If the z-axis is fixed, the linear transformation is through an angle φ of the xy-coordinate system to a new orientation in Eq. (1.8), Fig. 1.6, and Section 3.3:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathsf{R}_z(\varphi) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \equiv \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{4.1}$$

involves only one angle of the rotation about the z-axis. As shown in Chapter 3, the linear transformation of two successive rotations involves the product of the matrices corresponding to the sum of the angles. The product corresponds to two rotations, $\mathsf{R}_z(\varphi_1)\mathsf{R}_z(\varphi_2)$, and is defined by rotating first by the angle φ_2 and then by φ_1. According to Eq. (3.29), this corresponds to the product of the orthogonal 2×2 submatrices,

$$\begin{aligned} &\begin{pmatrix} \cos\varphi_1 & \sin\varphi_1 \\ -\sin\varphi_1 & \cos\varphi_1 \end{pmatrix} \begin{pmatrix} \cos\varphi_2 & \sin\varphi_2 \\ -\sin\varphi_2 & \cos\varphi_2 \end{pmatrix} \\ &\quad = \begin{pmatrix} \cos(\varphi_1+\varphi_2) & \sin(\varphi_1+\varphi_2) \\ -\sin(\varphi_1+\varphi_2) & \cos(\varphi_1+\varphi_2) \end{pmatrix}, \end{aligned} \tag{4.2}$$

using the addition formulas for the trigonometric functions. The unity in the lower right-hand corner of the matrix in Eq. (4.1) is also reproduced upon multiplication. The product is clearly a rotation, represented by the orthogonal matrix with angle $\varphi_1 + \varphi_2$. The associative group multiplication corresponds to the associative matrix multiplication. It is **commutative**, or **abelian**, because the order in which these rotations are performed does not matter. The inverse of the rotation with angle φ is that with angle $-\varphi$. The unit corresponds to the angle $\varphi = 0$. Striking off the coordinate vectors in Eq. (4.1), we can associate the matrix of the linear transformation with each rotation, which is a group multiplication preserving one-to-one mapping, an isomorphism: The matrices form a faithful representation of the rotation group. The unity in the right-hand corner is superfluous as well, like the coordinate vectors, and may be deleted. This defines another isomorphism and representation by the 2×2 submatrices:

$$\mathsf{R}_z(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \mathsf{R}(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}. \tag{4.3}$$

The group's name is $\mathsf{SO}(2)$, if the angle φ varies continuously from 0 to 2π; $\mathsf{SO}(2)$ has infinitely many elements and is compact.

The group of rotations R_z is obviously isomorphic to the group of rotations in Eq. (4.3). The unity with angle $\varphi = 0$ and the rotation with $\varphi = \pi$ form a finite subgroup. The finite subgroups with angles $2\pi m/n$, n an integer and $m = 0, 1, \ldots, n-1$ are **cyclic**; that is, the rotations $\mathsf{R}(2\pi m/n) = \mathsf{R}(2\pi/n)^m$. ■

In the following we shall discuss only the rotation groups $\mathsf{SO}(n)$ and unitary groups $\mathsf{SU}(n)$ among the classical Lie groups. (More examples of finite groups will be given in Section 4.7.)

Representations — Reducible and Irreducible

The representation of group elements by matrices is a very powerful technique and has been almost universally adopted by physicists. The use of matrices imposes no significant restriction. It can be shown that the elements of any finite group and of the continuous groups of Sections 4.2–4.4 may be represented by matrices. Examples are the rotations described in Eq. (4.3).

To illustrate how matrix representations arise from a symmetry, consider the stationary Schrödinger equation (or some other eigenvalue equation, such as $\mathsf{I}v_i = I_i v_i$ for the principal moments of inertia of a rigid body in classical mechanics, say),

$$H\psi = E\psi. \tag{4.4}$$

Let us assume that the Hamiltonian H stays invariant under a group G of transformations R in G (coordinate rotations, for example, for a central potential $V(r)$ in the Hamiltonian H); that is,

$$H_R = \mathsf{R}H\mathsf{R}^{-1} = H, \qquad \mathsf{R}H = H\mathsf{R}. \tag{4.5}$$

Now take a solution ψ of Eq. (4.4) and "rotate" it: $\psi \to \mathsf{R}\psi$. Then $\mathsf{R}\psi$ has the **same energy E** because multiplying Eq. (4.4) by R and using Eq. (4.5) yields

$$\boxed{\mathsf{R}H\psi = E(\mathsf{R}\psi) = \left(\mathsf{R}H\mathsf{R}^{-1}\right)R\psi = H(\mathsf{R}\psi).} \tag{4.6}$$

In other words, all rotated solutions $\mathsf{R}\psi$ are **degenerate** in energy or form what physicists call a **multiplet**. For example, the spin-up and -down states of a bound electron in the ground state of hydrogen form a doublet, and the states with projection quantum numbers $m = -l, -l+1, \ldots, l$ of orbital angular momentum l form a multiplet with $2l+1$ basis states.

Let us assume that this vector space V_ψ of transformed solutions has a finite dimension n. Let $\psi_1, \psi_2, \ldots, \psi_n$ be a basis. Since $\mathsf{R}\psi_j$ is a member of the multiplet, we can expand it in terms of its basis,

$$\mathsf{R}\psi_j = \sum_k r_{jk}\psi_k. \tag{4.7}$$

Thus, with each R in G we can associate a matrix (r_{jk}). Just as in Example 4.1.2, two successive rotations correspond to the product of their matrices, so this map $\mathsf{R} \to (r_{jk})$ is a representation of G. It is necessary for a representation to be **irreducible** that we can take any element of V_ψ and, by rotating with **all** elements R of G, transform it into **all** other elements of V_ψ. If not all elements of V_ψ are reached, then V_ψ splits into a direct sum of two or more vector subspaces, $V_\psi = V_1 \oplus V_2 \oplus \cdots$, which are mapped into themselves by rotating their elements. For example, the $2s$ state and $2p$ states of principal quantum number $n = 2$ of the hydrogen atom have the same energy (that is, are degenerate) and form

a reducible representation, because the $2s$ state cannot be rotated into the $2p$ states, and vice versa (angular momentum is conserved under rotations). In this case the representation is called **reducible**. Then we can find a basis in V_ψ (that is, there is a unitary matrix U) so that

$$\mathsf{U}(r_{jk})\mathsf{U}^\dagger = \begin{pmatrix} \mathbf{r}_1 & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{r}_2 & \cdots \\ \vdots & \vdots & \end{pmatrix} \tag{4.8}$$

for **all** R of G, and **all** matrices (r_{jk}) have **similar** block-diagonal shape. Here $\mathbf{r}_1, \mathbf{r}_2, \ldots$ are matrices of lower dimension than (r_{jk}) that are lined up along the diagonal and the $\mathbf{0}$'s are matrices made up of zeros. We may say that the representation has been decomposed into $\mathbf{r}_1 + \mathbf{r}_2 + \cdots$ along with $V_\psi = V_1 \oplus V_2 \oplus \cdots$.

The irreducible representations play a role in group theory that is roughly analogous to the unit vectors of vector analysis. They are the simplest representations; all others can be built from them. (See Section 4.4 on Clebsch–Gordan coefficients and Young tableaux.)

Exercises

4.1.1 Show that an $n \times n$ orthogonal matrix has $n(n-1)/2$ independent parameters.
Hint. The orthogonality condition, Eq. (3.71), provides constraints.

4.1.2 Show that an $n \times n$ unitary matrix has $n^2 - 1$ independent parameters.
Hint. Each element may be complex, doubling the number of possible parameters. Some of the constraint equations are likewise complex and count as two constraints.

4.1.3 The special linear group $\mathsf{SL}(2)$ consists of all 2×2 matrices (with complex elements) having a determinant of $+1$. Show that such matrices form a group.
Note. The $\mathsf{SL}(2)$ group can be related to the full Lorentz group in Section 4.4, much as the $\mathsf{SU}(2)$ group is related to $\mathsf{SO}(3)$.

4.1.4 Show that the rotations about the z-axis form a subgroup of $\mathsf{SO}(3)$. Is it an invariant subgroup?

4.1.5 Show that if R, S, T are elements of a group G so that $RS = T$ and $R \to (r_{ik})$, $S \to (s_{ik})$ is a representation according to Eq. (4.7), then

$$(r_{ik})(s_{ik}) = \left(t_{ik} = \sum_n r_{in} s_{nk}\right),$$

that is, group multiplication translates into matrix multiplication for any group representation.

4.2 Generators of Continuous Groups

A characteristic property of continuous groups known as Lie groups is that the parameters of a product element are analytic functions[4] of the parameters of the factors. The analytic

[4] *Analytic* here means having derivatives of all orders.

nature of the functions (differentiability) allows us to develop the concept of generator and to reduce the study of the whole group to a study of the group elements in the neighborhood of the identity element.

Lie's essential idea was to study elements R in a group G that are infinitesimally close to the unity of G. Let us consider the $\mathsf{SO}(2)$ group as a simple example. The 2×2 rotation matrices in Eq. (4.2) can be written in exponential form using the Euler identity, Eq. (3.170a), as

$$\mathsf{R}(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} = 1_2 \cos\varphi + i\sigma_2 \sin\varphi = \exp(i\sigma_2\varphi). \tag{4.9}$$

From the exponential form it is obvious that multiplication of these matrices is equivalent to addition of the arguments

$$\mathsf{R}(\varphi_2)\mathsf{R}(\varphi_1) = \exp(i\sigma_2\varphi_2)\exp(i\sigma_2\varphi_1) = \exp\big(i\sigma_2(\varphi_1+\varphi_2)\big) = \mathsf{R}(\varphi_1+\varphi_2).$$

Rotations close to 1 have small angle $\varphi \approx 0$.

This suggests that we look for an exponential representation

$$\mathsf{R} = \exp(i\varepsilon\mathsf{S}) = 1 + i\varepsilon\mathsf{S} + \mathsf{O}\big(\varepsilon^2\big), \qquad \varepsilon \to 0, \tag{4.10}$$

for group elements R in G close to the unity 1. The infinitesimal transformations are $\varepsilon\mathsf{S}$, and the S are called generators of G. They form a linear space because multiplication of the group elements R translates into addition of generators S. The dimension of this vector space (over the complex numbers) is the **order** of G, that is, the number of linearly independent generators of the group.

If R is a rotation, it does not change the volume element of the coordinate space that it rotates, that is, $\det(\mathsf{R}) = 1$, and we may use Eq. (3.171) to see that

$$\det(\mathsf{R}) = \exp\big(\mathrm{trace}(\ln\mathsf{R})\big) = \exp\big(i\varepsilon\,\mathrm{trace}(\mathsf{S})\big) = 1$$

implies $\varepsilon\,\mathrm{trace}(\mathsf{S}) = 0$ and, upon dividing by the small but nonzero parameter ε, that **generators are traceless**,

$$\mathrm{trace}(\mathsf{S}) = 0. \tag{4.11}$$

This is the case not only for the rotation groups $\mathsf{SO}(n)$ but also for unitary groups $\mathsf{SU}(n)$.

If R of G in Eq. (4.10) is unitary, then $S^\dagger = S$ is Hermitian, which is also the case for $\mathsf{SO}(n)$ and $\mathsf{SU}(n)$. This explains why the extra i has been inserted in Eq. (4.10).

Next we go around the unity in four steps, similar to parallel transport in differential geometry. We expand the group elements

$$\begin{aligned} \mathsf{R}_i &= \exp(i\varepsilon_i\mathsf{S}_i) = 1 + i\varepsilon_i\mathsf{S}_i - \tfrac{1}{2}\varepsilon_i^2\mathsf{S}_i^2 + \cdots, \\ \mathsf{R}_i^{-1} &= \exp(-i\varepsilon_i\mathsf{S}_i) = 1 - i\varepsilon_i\mathsf{S}_i - \tfrac{1}{2}\varepsilon_i^2\mathsf{S}_i^2 + \cdots, \end{aligned} \tag{4.12}$$

to second order in the small group parameter ε_i because the linear terms and several quadratic terms all cancel in the product (Fig. 4.1)

$$\begin{aligned} \mathsf{R}_i^{-1}\mathsf{R}_j^{-1}\mathsf{R}_i\mathsf{R}_j &= 1 + \varepsilon_i\varepsilon_j[\mathsf{S}_j, \mathsf{S}_i] + \cdots, \\ &= 1 + \varepsilon_i\varepsilon_j \sum_k c_{ji}^k \mathsf{S}_k + \cdots, \end{aligned} \tag{4.13}$$

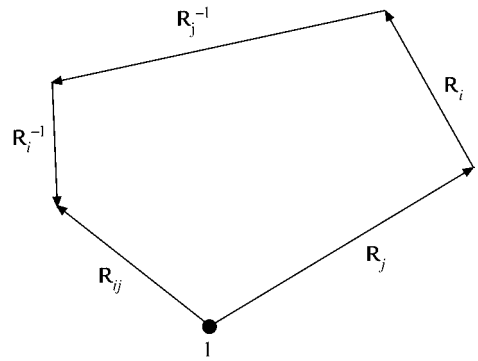

FIGURE 4.1 Illustration of Eq. (4.13).

when Eq. (4.12) is substituted into Eq. (4.13). The last line holds because the product in Eq. (4.13) is again a group element, R_{ij}, close to the unity in the group G. Hence its exponent must be a linear combination of the generators S_k, and its infinitesimal group parameter has to be proportional to the product $\varepsilon_i \varepsilon_j$. Comparing both lines in Eq. (4.13) we find the **closure** relation of the generators of the Lie group G,

$$[\mathsf{S}_i, \mathsf{S}_j] = \sum_k c_{ij}^k \mathsf{S}_k. \tag{4.14}$$

The coefficients c_{ij}^k are the structure constants of the group G. Since the commutator in Eq. (4.14) is antisymmetric in i and j, so are the structure constants in the lower indices,

$$c_{ij}^k = -c_{ji}^k. \tag{4.15}$$

If the commutator in Eq. (4.14) is taken as a multiplication law of generators, we see that the vector space of generators becomes an algebra, the **Lie algebra** $\mathcal{G}$ of the group G. An **algebra** has two group structures, a commutative product denoted by a $+$ symbol (this is the addition of infinitesimal generators of a Lie group) and a multiplication (the commutator of generators). Often an algebra is a vector space with a multiplication, such as a ring of square matrices. For $\mathsf{SU}(l+1)$ the Lie algebra is called $\mathcal{A}_l$, for $\mathsf{SO}(2l+1)$ it is $\mathcal{B}_l$, and for $\mathsf{SO}(2l)$ it is $\mathcal{D}_l$, where $l = 1, 2, \ldots$ is a positive integer, later called the **rank** of the Lie group G or of its algebra $\mathcal{G}$.

Finally, the **Jacobi identity** holds for all double commutators

$$\big[[\mathsf{S}_i, \mathsf{S}_j], \mathsf{S}_k\big] + \big[[\mathsf{S}_j, \mathsf{S}_k], \mathsf{S}_i\big] + \big[[\mathsf{S}_k, \mathsf{S}_i], \mathsf{S}_j\big] = 0, \tag{4.16}$$

which is easily verified using the definition of any commutator $[A, B] \equiv AB - BA$. When Eq. (4.14) is substituted into Eq. (4.16) we find another constraint on structure constants,

$$\sum_m \big\{c_{ij}^m [\mathsf{S}_m, \mathsf{S}_k] + c_{jk}^m [\mathsf{S}_m, \mathsf{S}_i] + c_{ki}^m [\mathsf{S}_m, \mathsf{S}_j]\big\} = 0. \tag{4.17}$$

Upon inserting Eq. (4.14) again, Eq. (4.17) implies that

$$\sum_{mn} \big\{c_{ij}^m c_{mk}^n \mathsf{S}_n + c_{jk}^m c_{mi}^n \mathsf{S}_n + c_{ki}^m c_{mj}^n \mathsf{S}_n\big\} = 0, \tag{4.18}$$

where the common factor S_n (and the sum over n) may be dropped because the generators are linearly independent. Hence

$$\sum_m \left\{ c_{ij}^m c_{mk}^n + c_{jk}^m c_{mi}^n + c_{ki}^m c_{mj}^n \right\} = 0. \tag{4.19}$$

The relations (4.14), (4.15), and (4.19) form the basis of Lie algebras from which finite elements of the Lie group near its unity can be reconstructed.

Returning to Eq. (4.5), the inverse of R is $\mathsf{R}^{-1} = \exp(-i\varepsilon \mathsf{S})$. We expand H_R according to the Baker–Hausdorff formula, Eq. (3.172),

$$H = H_R = \exp(i\varepsilon \mathsf{S}) H \exp(-i\varepsilon \mathsf{S}) = H + i\varepsilon[\mathsf{S}, H] - \tfrac{1}{2}\varepsilon^2 \big[\mathsf{S}[\mathsf{S}, H]\big] + \cdots \tag{4.20}$$

We drop H from Eq. (4.20), divide by the small (but nonzero), ε, and let $\varepsilon \to 0$. Then Eq. (4.20) implies that the commutator

$$\boxed{[\mathsf{S}, H] = 0.} \tag{4.21}$$

If S and H are Hermitian matrices, Eq. (4.21) implies that S and H can be simultaneously diagonalized and have common eigenvectors (for matrices, see Section 3.5; for operators, see Schur's lemma in Section 4.3). If S and H are differential operators like the Hamiltonian and orbital angular momentum in quantum mechanics, then Eq. (4.21) implies that S and H have common eigenfunctions and that the degenerate eigenvalues of H can be distinguished by the eigenvalues of the generators S. These eigenfunctions and eigenvalues, s, are solutions of separate differential equations, $\mathsf{S}\psi_s = s\psi_s$, so group theory (that is, symmetries) leads to a separation of variables for a partial differential equation that is invariant under the transformations of the group.

For example, let us take the single-particle Hamiltonian

$$H = -\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{\hbar^2}{2mr^2}\mathbf{L}^2 + V(r)$$

that is invariant under $\mathsf{SO}(3)$ and, therefore, a function of the radial distance r, the radial gradient, and the rotationally invariant operator $\mathbf{L}^2$ of $\mathsf{SO}(3)$. Upon replacing the orbital angular momentum operator $\mathbf{L}^2$ by its eigenvalue $l(l+1)$ we obtain the radial Schrödinger equation (ODE),

$$H R_l(r) = \left[-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr} r^2 \frac{d}{dr} + \frac{\hbar^2 l(l+1)}{2mr^2} + V(r) \right] R_l(r) = E_l R_l(r),$$

where $R_l(r)$ is the radial wave function.

For cylindrical symmetry, the invariance of H under rotations about the z-axis would require H to be independent of the rotation angle φ, leading to the ODE

$$H R_m(z, \rho) = E_m R_m(z, \rho),$$

with m the eigenvalue of $L_z = -i\,\partial/\partial\varphi$, the z-component of the orbital angular momentum operator. For more examples, see the separation of variables method for partial differential equations in Section 9.3 and special functions in Chapter 12. This is by far the most important application of group theory in quantum mechanics.

In the next subsections we shall study orthogonal and unitary groups as examples to understand better the general concepts of this section.

Rotation Groups SO(2) and SO(3)

For SO(2) as defined by Eq. (4.3) there is only one linearly independent generator, σ_2, and the order of SO(2) is 1. We get σ_2 from Eq. (4.9) by differentiation at the unity of SO(2), that is, $\varphi = 0$,

$$-i\,d\mathsf{R}(\varphi)/d\varphi|_{\varphi=0} = -i\begin{pmatrix} -\sin\varphi & \cos\varphi \\ -\cos\varphi & -\sin\varphi \end{pmatrix}\Bigg|_{\varphi=0} = -i\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \sigma_2. \tag{4.22}$$

For the rotations $\mathsf{R}_z(\varphi)$ about the z-axis described by Eq. (4.1), the generator is given by

$$-i\,d\mathsf{R}_z(\varphi)/d\varphi|_{\varphi=0} = \mathsf{S}_z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.23}$$

where the factor i is inserted to make S_z Hermitian. The rotation $\mathsf{R}_z(\delta\varphi)$ through an infinitesimal angle $\delta\varphi$ may then be expanded to first order in the small $\delta\varphi$ as

$$\mathsf{R}_z(\delta\varphi) = 1_3 + i\delta\varphi\mathsf{S}_z. \tag{4.24}$$

A finite rotation $\mathsf{R}(\varphi)$ may be compounded of successive infinitesimal rotations

$$\mathsf{R}_z(\delta\varphi_1 + \delta\varphi_2) = (1 + i\delta\varphi_1\mathsf{S}_z)(1 + i\delta\varphi_2\mathsf{S}_z). \tag{4.25}$$

Let $\delta\varphi = \varphi/N$ for N rotations, with $N \to \infty$. Then

$$\mathsf{R}_z(\varphi) = \lim_{N\to\infty}\left[1 + (i\varphi/N)\mathsf{S}_z\right]^N = \exp(i\varphi\mathsf{S}_z). \tag{4.26}$$

This form identifies S_z as the generator of the group R_z, an abelian subgroup of SO(3), the group of rotations in three dimensions with determinant $+1$. Each 3×3 matrix $\mathsf{R}_z(\varphi)$ is orthogonal, hence unitary, and trace$(\mathsf{S}_z) = 0$, in accord with Eq. (4.11).

By differentiation of the coordinate rotations

$$\mathsf{R}_x(\psi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{pmatrix}, \qquad \mathsf{R}_y(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}, \tag{4.27}$$

we get the generators

$$\mathsf{S}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \qquad \mathsf{S}_y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \tag{4.28}$$

of $\mathsf{R}_x(\mathsf{R}_y)$, the subgroup of rotations about the x- (y-)axis.

Rotation of Functions and Orbital Angular Momentum

In the foregoing discussion the group elements are matrices that rotate the coordinates. Any physical system being described is held fixed. Now let us hold the coordinates fixed and rotate a function $\psi(x, y, z)$ relative to our fixed coordinates. With R to rotate the coordinates,

$$\mathbf{x}' = \mathsf{R}\mathbf{x}, \tag{4.29}$$

we define R on ψ by

$$\mathsf{R}\psi(x, y, z) = \psi'(x, y, z) \equiv \psi(\mathbf{x}'). \tag{4.30}$$

In words, R operates on the function ψ, creating a **new function** ψ' that is numerically equal to $\psi(\mathbf{x}')$, where $\mathbf{x}'$ are the coordinates rotated by R. If R rotates the coordinates counterclockwise, the effect of R is to rotate the pattern of the function ψ clockwise.

Returning to Eqs. (4.30) and (4.1), consider an infinitesimal rotation again, $\varphi \to \delta\varphi$. Then, using R_z Eq. (4.1), we obtain

$$\mathsf{R}_z(\delta\varphi)\psi(x, y, z) = \psi(x + y\delta\varphi, y - x\delta\varphi, z). \tag{4.31}$$

The right side may be expanded to first order in the small $\delta\varphi$ to give

$$\begin{aligned}\mathsf{R}_z(\delta\varphi)\psi(x, y, z) &= \psi(x, y, z) - \delta\varphi\{x\partial\psi/\partial y - y\partial\psi/\partial x\} + O(\delta\varphi)^2 \\ &= (1 - i\delta\varphi L_z)\psi(x, y, z),\end{aligned} \tag{4.32}$$

the differential expression in curly brackets being the orbital angular momentum iL_z (Exercise 1.8.7). Since a rotation of first φ and then $\delta\varphi$ about the z-axis is given by

$$\mathsf{R}_z(\varphi + \delta\varphi)\psi = \mathsf{R}_z(\delta\varphi)\mathsf{R}_z(\varphi)\psi = (1 - i\delta\varphi L_z)\mathsf{R}_z(\varphi)\psi, \tag{4.33}$$

we have (as an operator equation)

$$\frac{d\mathsf{R}_z}{d\varphi} = \lim_{\delta\varphi\to 0} \frac{\mathsf{R}_z(\varphi + \delta\varphi) - \mathsf{R}_z(\varphi)}{\delta\varphi} = -iL_z\mathsf{R}_z(\varphi). \tag{4.34}$$

In this form Eq. (4.34) integrates immediately to

$$\mathsf{R}_z(\varphi) = \exp(-i\varphi L_z). \tag{4.35}$$

Note that $\mathsf{R}_z(\varphi)$ rotates functions (clockwise) relative to fixed coordinates and that L_z is the z component of the orbital angular momentum $\mathbf{L}$. The constant of integration is fixed by the boundary condition $\mathsf{R}_z(0) = 1$.

As suggested by Eq. (4.32), L_z is connected to S_z by

$$L_z = (x, y, z)\mathsf{S}_z \begin{pmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{pmatrix} = -i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right), \tag{4.36}$$

so L_x, L_y, and L_z satisfy the same commutation relations,

$$\boxed{[L_i, L_j] = i\varepsilon_{ijk}L_k,} \tag{4.37}$$

as S_x, S_y, and S_z and yield the same structure constants $i\varepsilon_{ijk}$ of $\mathsf{SO}(3)$.

SU(2) — SO(3) Homomorphism

Since unitary 2×2 matrices transform complex two-dimensional vectors preserving their norm, they represent the most general transformations of (a basis in the Hilbert space of) spin $\frac{1}{2}$ wave functions in nonrelativistic quantum mechanics. The basis states of this system are conventionally chosen to be

$$|\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

corresponding to spin $\frac{1}{2}$ up and down states, respectively. We can show that the **special unitary** group SU(2) of unitary 2×2 matrices with determinant $+1$ has all three Pauli matrices σ_i as generators (while the rotations of Eq. (4.3) form a one-dimensional abelian subgroup). So SU(2) is of order 3 and depends on three real continuous parameters ξ, η, ζ, which are often called the **Cayley–Klein** parameters. To construct its general element, we start with the observation that orthogonal 2×2 matrices are real unitary matrices, so they form a subgroup of SU(2). We also see that

$$\begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix}$$

is unitary for real angle α with determinant $+1$. So these simple and manifestly unitary matrices form another subgroup of SU(2) from which we can obtain all elements of SU(2), that is, the general 2×2 unitary matrix of determinant $+1$. For a two-component spin $\frac{1}{2}$ wave function of quantum mechanics this diagonal unitary matrix corresponds to multiplication of the spin-up wave function with a phase factor $e^{i\alpha}$ and the spin-down component with the inverse phase factor. Using the real angle η instead of φ for the rotation matrix and then multiplying by the diagonal unitary matrices, we construct a 2×2 unitary matrix that depends on three parameters and clearly is a more general element of SU(2):

$$\begin{aligned}
&\begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix} \begin{pmatrix} \cos\eta & \sin\eta \\ -\sin\eta & \cos\eta \end{pmatrix} \begin{pmatrix} e^{i\beta} & 0 \\ 0 & e^{-i\beta} \end{pmatrix} \\
&\quad = \begin{pmatrix} e^{i\alpha}\cos\eta & e^{i\alpha}\sin\eta \\ -e^{-i\alpha}\sin\eta & e^{-i\alpha}\cos\eta \end{pmatrix} \begin{pmatrix} e^{i\beta} & 0 \\ 0 & e^{-i\beta} \end{pmatrix} \\
&\quad = \begin{pmatrix} e^{i(\alpha+\beta)}\cos\eta & e^{i(\alpha-\beta)}\sin\eta \\ -e^{-i(\alpha-\beta)}\sin\eta & e^{-i(\alpha+\beta)}\cos\eta \end{pmatrix}.
\end{aligned}$$

Defining $\alpha + \beta \equiv \xi, \alpha - \beta \equiv \zeta$, we have in fact constructed the general element of SU(2):

$$\mathsf{U}(\xi, \eta, \zeta) = \begin{pmatrix} e^{i\xi}\cos\eta & e^{i\zeta}\sin\eta \\ -e^{-i\zeta}\sin\eta & e^{-i\xi}\cos\eta \end{pmatrix} = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}. \tag{4.38}$$

To see this, we write the general SU(2) element as $\mathsf{U} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with complex numbers a, b, c, d so that $\det(\mathsf{U}) = 1$. Writing unitarity, $\mathsf{U}^\dagger = \mathsf{U}^{-1}$, and using Eq. (3.50) for the

inverse we obtain

$$\begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

implying $c = -b^*, d = a^*$, as shown in Eq. (4.38). It is easy to check that the determinant $\det(\mathsf{U}) = 1$ and that $\mathsf{U}^\dagger\mathsf{U} = 1 = \mathsf{U}\mathsf{U}^\dagger$ hold.

To get the generators, we differentiate (and drop irrelevant overall factors):

$$-i\,\partial\mathsf{U}/\partial\xi_{|\xi=0,\eta=0} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \sigma_3, \tag{4.39a}$$

$$-i\,\partial\mathsf{U}/\partial\eta_{|\eta=0,\zeta=0} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \sigma_2. \tag{4.39b}$$

To avoid a factor $1/\sin\eta$ for $\eta \to 0$ upon differentiating with respect to ζ, we use instead the right-hand side of Eq. (4.38) for U for pure imaginary $b = i\beta$ with $\beta \to 0$, so $a = \sqrt{1-\beta^2}$ from $|a|^2 + |b|^2 = a^2 + \beta^2 = 1$. Differentiating such a U, we get the third generator,

$$-i\,\frac{\partial}{\partial\beta}\begin{pmatrix} \sqrt{1-\beta^2} & i\beta \\ i\beta & \sqrt{1-\beta^2} \end{pmatrix}\Bigg|_{\beta=0} = -i\begin{pmatrix} -\frac{\beta}{\sqrt{1-\beta^2}} & i \\ -i & \frac{\beta}{\sqrt{1-\beta^2}} \end{pmatrix}\Bigg|_{\beta=0} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \sigma_1. \tag{4.39c}$$

The Pauli matrices are all traceless and Hermitian.

With the Pauli matrices as generators, the elements $\mathsf{U}_1, \mathsf{U}_2, \mathsf{U}_3$ of $\mathsf{SU}(2)$ may be generated by

$$\mathsf{U}_1 = \exp(ia_1\sigma_1/2), \qquad \mathsf{U}_2 = \exp(ia_2\sigma_2/2), \qquad \mathsf{U}_3 = \exp(ia_3\sigma_3/2). \tag{4.40}$$

The three parameters a_i are real. The extra factor $1/2$ is present in the exponents to make $S_i = \sigma_i/2$ satisfy the same commutation relations,

$$\boxed{[\mathsf{S}_i, \mathsf{S}_j] = i\varepsilon_{ijk}\mathsf{S}_k,} \tag{4.41}$$

as the angular momentum in Eq. (4.37).

To connect and compare our results, Eq. (4.3) gives a rotation operator for rotating the Cartesian coordinates in the three-space $\mathbb{R}^3$. Using the angular momentum matrix S_3, we have as the corresponding rotation operator in two-dimensional (complex) space $\mathsf{R}_z(\varphi) = \exp(i\varphi\sigma_3/2)$. For rotating the two-component vector wave function (spinor) or a spin $1/2$ particle relative to fixed coordinates, the corresponding rotation operator is $\mathsf{R}_z(\varphi) = \exp(-i\varphi\sigma_3/2)$ according to Eq. (4.35).

More generally, using in Eq. (4.40) the Euler identity, Eq. (3.170a), we obtain

$$\mathsf{U}_j = \cos\left(\frac{a_j}{2}\right) + i\sigma_j \sin\left(\frac{a_j}{2}\right). \tag{4.42}$$

Here the parameter a_j appears as an angle, the coefficient of an angular momentum matrix-like φ in Eq. (4.26). The selection of Pauli matrices corresponds to the Euler angle rotations described in Section 3.3.

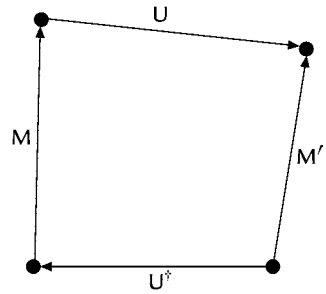

FIGURE 4.2 Illustration of $\mathsf{M}' = \mathsf{U}\mathsf{M}\mathsf{U}^\dagger$ in Eq. (4.43).

As just seen, the elements of SU(2) describe rotations in a two-dimensional complex space that leave $|z_1|^2 + |z_2|^2$ invariant. The determinant is $+1$. There are three independent real parameters. Our real orthogonal group SO(3) clearly describes rotations in ordinary three-dimensional space with the important characteristic of leaving $x^2 + y^2 + z^2$ **invariant**. Also, there are three independent real parameters. The rotation interpretations and the equality of numbers of parameters suggest the existence of some correspondence between the groups SU(2) and SO(3). Here we develop this correspondence.

The operation of SU(2) on a matrix is given by a unitary transformation, Eq. (4.5), with $\mathsf{R} = \mathsf{U}$ and Fig. 4.2:

$$\mathsf{M}' = \mathsf{U}\mathsf{M}\mathsf{U}^\dagger. \tag{4.43}$$

Taking M to be a 2×2 matrix, we note that any 2×2 matrix may be written as a linear combination of the unit matrix and the three Pauli matrices of Section 3.4. Let M be the zero-trace matrix,

$$\mathsf{M} = x\sigma_1 + y\sigma_2 + z\sigma_3 = \begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix}, \tag{4.44}$$

the unit matrix not entering. Since the trace is invariant under a unitary similarity transformation (Exercise 3.3.9), M' must have the same form,

$$\mathsf{M}' = x'\sigma_1 + y'\sigma_2 + z'\sigma_3 = \begin{pmatrix} z' & x' - iy' \\ x' + iy' & -z' \end{pmatrix}. \tag{4.45}$$

The determinant is also invariant under a unitary transformation (Exercise 3.3.10). Therefore

$$-\left(x^2 + y^2 + z^2\right) = -\left(x'^2 + y'^2 + z'^2\right), \tag{4.46}$$

or $x^2 + y^2 + z^2$ is invariant under this operation of SU(2), just as with SO(3). Operations of SU(2) on M must produce rotations of the coordinates x, y, z appearing therein. This suggests that SU(2) and SO(3) may be isomorphic or at least homomorphic.

We approach the problem of what this operation of SU(2) corresponds to by considering special cases. Returning to Eq. (4.38), let $a = e^{i\xi}$ and $b = 0$, or

$$\mathsf{U}_3 = \begin{pmatrix} e^{i\xi} & 0 \\ 0 & e^{-i\xi} \end{pmatrix}. \tag{4.47}$$

In anticipation of Eq. (4.51), this U is given a subscript 3.

Carrying out a unitary similarity transformation, Eq. (4.43), on each of the three Pauli σ's of SU(2), we have

$$\begin{aligned} \mathsf{U}_3\sigma_1\mathsf{U}_3^\dagger &= \begin{pmatrix} e^{i\xi} & 0 \\ 0 & e^{-i\xi} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} e^{-i\xi} & 0 \\ 0 & e^{i\xi} \end{pmatrix} \\ &= \begin{pmatrix} 0 & e^{2i\xi} \\ e^{-2i\xi} & 0 \end{pmatrix}. \end{aligned} \tag{4.48}$$

We reexpress this result in terms of the Pauli σ_i, as in Eq. (4.44), to obtain

$$\mathsf{U}_3 x\sigma_1\mathsf{U}_3^\dagger = x\sigma_1\cos 2\xi - x\sigma_2\sin 2\xi. \tag{4.49}$$

Similarly,

$$\begin{aligned} \mathsf{U}_3 y\sigma_2\mathsf{U}_3^\dagger &= y\sigma_1\sin 2\xi + y\sigma_2\cos 2\xi, \\ \mathsf{U}_3 z\sigma_3\mathsf{U}_3^\dagger &= z\sigma_3. \end{aligned} \tag{4.50}$$

From these double angle expressions we see that we should start with a **halfangle**: $\xi = \alpha/2$. Then, adding Eqs. (4.49) and (4.50) and comparing with Eqs. (4.44) and (4.45), we obtain

$$\begin{aligned} x' &= x\cos\alpha + y\sin\alpha \\ y' &= -x\sin\alpha + y\cos\alpha \\ z' &= z. \end{aligned} \tag{4.51}$$

The 2×2 unitary transformation using $\mathsf{U}_3(\alpha)$ is equivalent to the rotation operator $\mathsf{R}(\alpha)$ of Eq. (4.3).

The correspondence of

$$\mathsf{U}_2(\beta) = \begin{pmatrix} \cos\beta/2 & \sin\beta/2 \\ -\sin\beta/2 & \cos\beta/2 \end{pmatrix} \tag{4.52}$$

and $\mathsf{R}_y(\beta)$ and of

$$\mathsf{U}_1(\varphi) = \begin{pmatrix} \cos\varphi/2 & i\sin\varphi/2 \\ i\sin\varphi/2 & \cos\varphi/2 \end{pmatrix} \tag{4.53}$$

and $\mathsf{R}_1(\varphi)$ follow similarly. Note that $\mathsf{U}_k(\psi)$ has the general form

$$\mathsf{U}_k(\psi) = 1_2\cos\psi/2 + i\sigma_k\sin\psi/2, \tag{4.54}$$

where $k = 1, 2, 3$.

The correspondence

$$\mathsf{U}_3(\alpha)=\begin{pmatrix} e^{i\alpha/2} & 0 \\ 0 & e^{-i\alpha/2}\end{pmatrix}\leftrightarrow\begin{pmatrix}\cos\alpha & \sin\alpha & 0\\ -\sin\alpha & \cos\alpha & 0\\ 0 & 0 & 1\end{pmatrix}=\mathsf{R}_z(\alpha) \tag{4.55}$$

is not a simple one-to-one correspondence. Specifically, as α in R_z ranges from 0 to 2π, the parameter in U_3, $\alpha/2$, goes from 0 to π. We find

$$\begin{aligned}\mathsf{R}_z(\alpha+2\pi)&=\mathsf{R}_z(\alpha)\\ \mathsf{U}_3(\alpha+2\pi)&=\begin{pmatrix}-e^{i\alpha/2} & 0\\ 0 & -e^{-i\alpha/2}\end{pmatrix}=-\mathsf{U}_3(\alpha).\end{aligned} \tag{4.56}$$

Therefore **both** $\mathsf{U}_3(\alpha)$ and $\mathsf{U}_3(\alpha+2\pi)=-\mathsf{U}_3(\alpha)$ correspond to $\mathsf{R}_z(\alpha)$. The correspondence is 2 to 1, or SU(2) and SO(3) are **homomorphic**. This establishment of the correspondence between the representations of SU(2) and those of SO(3) means that the known representations of SU(2) automatically provide us with the representations of SO(3).

Combining the various rotations, we find that a unitary transformation using

$$\mathsf{U}(\alpha,\beta,\gamma)=\mathsf{U}_3(\gamma)\mathsf{U}_2(\beta)\mathsf{U}_3(\alpha) \tag{4.57}$$

corresponds to the general Euler rotation $\mathsf{R}_z(\gamma)\mathsf{R}_y(\beta)\mathsf{R}_z(\alpha)$. By direct multiplication,

$$\begin{aligned}\mathsf{U}(\alpha,\beta,\gamma)&=\begin{pmatrix}e^{i\gamma/2} & 0\\ 0 & e^{-i\gamma/2}\end{pmatrix}\begin{pmatrix}\cos\beta/2 & \sin\beta/2\\ -\sin\beta/2 & \cos\beta/2\end{pmatrix}\begin{pmatrix}e^{i\alpha/2} & 0\\ 0 & e^{-i\alpha/2}\end{pmatrix}\\ &=\begin{pmatrix}e^{i(\gamma+\alpha)/2}\cos\beta/2 & e^{i(\gamma-\alpha)/2}\sin\beta/2\\ -e^{-i(\gamma-\alpha)/2}\sin\beta/2 & e^{-i(\gamma+\alpha)/2}\cos\beta/2\end{pmatrix}.\end{aligned} \tag{4.58}$$

This is our alternate general form, Eq. (4.38), with

$$\xi=(\gamma+\alpha)/2,\qquad \eta=\beta/2,\qquad \zeta=(\gamma-\alpha)/2. \tag{4.59}$$

Thus, from Eq. (4.58) we may identify the parameters of Eq. (4.38) as

$$\begin{aligned}a&=e^{i(\gamma+\alpha)/2}\cos\beta/2\\ b&=e^{i(\gamma-\alpha)/2}\sin\beta/2.\end{aligned} \tag{4.60}$$

SU(2)-Isospin and SU(3)-Flavor Symmetry

The application of group theory to "elementary" particles has been labeled by Wigner the third stage of group theory and physics. The first stage was the search for the 32 crystallographic point groups and the 230 space groups giving crystal symmetries — Section 4.7. The second stage was a search for representations such as of SO(3) and SU(2) — Section 4.2. Now in this stage, physicists are back to a search for groups.

In the 1930s to 1960s the study of strongly interacting particles of nuclear and high-energy physics led to the SU(2) isospin group and the SU(3) flavor symmetry. In the 1930s, after the neutron was discovered, Heisenberg proposed that the nuclear forces were charge

Table 4.1 Baryons with Spin $\frac{1}{2}$ Even Parity

		Mass (MeV)	Y	I	I_3
	Ξ^-	1321.32			$-\frac{1}{2}$
Ξ			-1	$\frac{1}{2}$	
	Ξ^0	1314.9			$+\frac{1}{2}$
	Σ^-	1197.43			-1
Σ	Σ^0	1192.55	0	1	0
	Σ^+	1189.37			$+1$
Λ	Λ	1115.63	0	0	0
	n	939.566			$-\frac{1}{2}$
N			1	$\frac{1}{2}$	
	p	938.272			$+\frac{1}{2}$

independent. The neutron mass differs from that of the proton by only 1.6%. If this tiny mass difference is ignored, the neutron and proton may be considered as two charge (or isospin) states of a doublet, called the **nucleon**. The isospin $\mathbf{I}$ has z-projection $I_3 = 1/2$ for the proton and $I_3 = -1/2$ for the neutron. Isospin has nothing to do with spin (the particle's intrinsic angular momentum), but the two-component isospin state obeys the same mathematical relations as the spin $1/2$ state. For the nucleon, $\mathbf{I} = \boldsymbol{\tau}/2$ are the usual Pauli matrices and the $\pm 1/2$ isospin states are eigenvectors of the Pauli matrix $\tau_3 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Similarly, the three charge states of the pion (π^+, π^0, π^-) form a triplet. The pion is the lightest of all strongly interacting particles and is the carrier of the nuclear force at long distances, much like the photon is that of the electromagnetic force. The strong interaction treats alike members of these particle families, or multiplets, and conserves isospin. The symmetry is the SU(2) isospin group.

By the 1960s particles produced as resonances by accelerators had proliferated. The eight shown in Table 4.1 attracted particular attention.[5] The relevant conserved quantum numbers that are analogs and generalizations of L_z and $\mathbf{L}^2$ from SO(3) are I_3 and I^2 for isospin and Y for **hypercharge**. Particles may be grouped into charge or isospin multiplets. Then the hypercharge may be taken as twice the average charge of the multiplet. For the nucleon, that is, the neutron–proton doublet, $Y = 2 \cdot \frac{1}{2}(0 + 1) = 1$. The hypercharge and isospin values are listed in Table 4.1 for baryons like the nucleon and its (approximately degenerate) partners. They form an octet, as shown in Fig. 4.3, after which the corresponding symmetry is called the **eightfold way**. In 1961 Gell-Mann, and independently Ne'eman, suggested that the strong interaction should be (approximately) invariant under a three-dimensional special unitary group, SU(3), that is, has SU(3) **flavor symmetry**.

The choice of SU(3) was based first on the two conserved and independent quantum numbers, $H_1 = I_3$ and $H_2 = Y$ (that is, generators with $[I_3, Y] = 0$, not Casimir invariants; see the summary in Section 4.3) that call for a group of rank 2. Second, the group had to have an eight-dimensional representation to account for the nearly degenerate baryons and four similar octets for the mesons. In a sense, SU(3) is the simplest generalization of SU(2) isospin. Three of its generators are zero-trace Hermitian 3×3 matrices that contain

[5] All masses are given in energy units, 1 MeV $= 10^6$ eV.

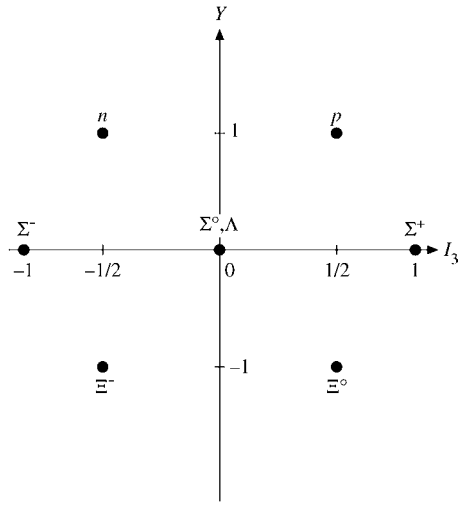

FIGURE 4.3 Baryon octet weight diagram for SU(3).

the 2×2 isospin Pauli matrices τ_i in the upper left corner,

$$\lambda_i = \begin{pmatrix} \tau_i & & 0 \\ & & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad i = 1, 2, 3. \tag{4.61a}$$

Thus, the SU(2)-isospin group is a subgroup of SU(3)-flavor with $I_3 = \lambda_3/2$. Four other generators have the off-diagonal 1's of τ_1, and $-i, i$ of τ_2 in all other possible locations to form zero-trace Hermitian 3×3 matrices,

$$\begin{aligned} \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \\ \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \end{aligned} \tag{4.61b}$$

The second diagonal generator has the two-dimensional unit matrix 1_2 in the upper left corner, which makes it clearly independent of the SU(2)-isospin subgroup because of its nonzero trace in that subspace, and -2 in the third diagonal place to make it traceless,

$$\lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{4.61c}$$

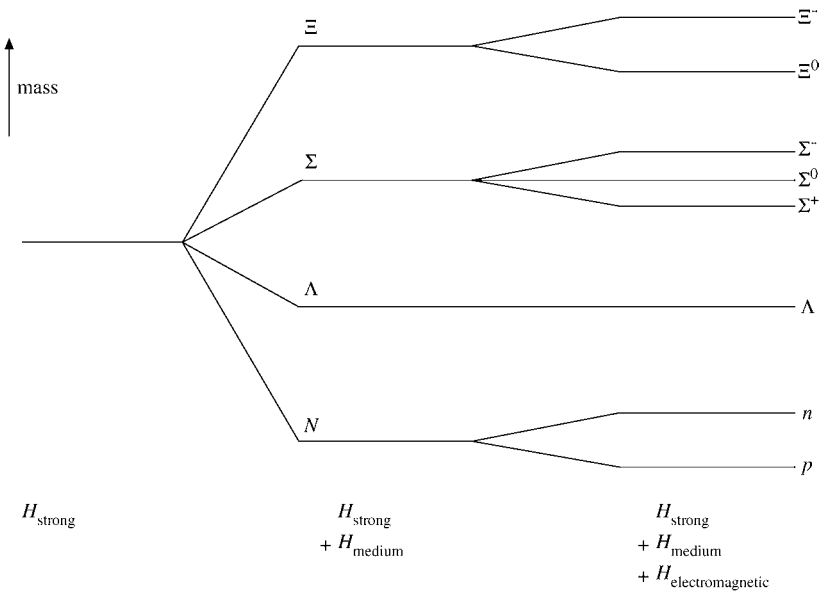

FIGURE 4.4 Baryon mass splitting.

Altogether there are $3^2 - 1 = 8$ generators for SU(3), which has order 8. From the commutators of these generators the structure constants of SU(3) can easily be obtained.

Returning to the SU(3) flavor symmetry, we imagine the Hamiltonian for our eight baryons to be composed of three parts:

$$H = H_{\text{strong}} + H_{\text{medium}} + H_{\text{electromagnetic}}. \tag{4.62}$$

The first part, H_{strong}, has the SU(3) symmetry and leads to the eightfold degeneracy. Introduction of the symmetry-breaking term, H_{medium}, removes part of the degeneracy, giving the four isospin multiplets (Ξ^-, Ξ^0), $(\Sigma^-, \Sigma^0, \Sigma^+)$, Λ, and $N = (p, n)$ different masses. These are still multiplets because H_{medium} has SU(2)-isospin symmetry. Finally, the presence of charge-dependent forces splits the isospin multiplets and removes the last degeneracy. This imagined sequence is shown in Fig. 4.4.

The octet representation is not the simplest SU(3) representation. The simplest representations are the triangular ones shown in Fig. 4.5, from which all others can be generated by generalized angular momentum coupling (see Section 4.4 on Young tableaux). The **fundamental representation** in Fig. 4.5a contains the u (up), d (down), and s (strange) quarks, and Fig. 4.5b contains the corresponding antiquarks. Since the meson octets can be obtained from the quark representations as $q\bar{q}$, with $3^2 = 8 + 1$ states, this suggests that mesons contain quarks (and antiquarks) as their constituents (see Exercise 4.4.3). The resulting quark model gives a successful description of hadronic spectroscopy. The resolution of its problem with the Pauli exclusion principle eventually led to the SU(3)-color gauge theory of the **strong interaction** called **quantum chromodynamics** (QCD).

To keep group theory and its very real accomplishment in proper perspective, we should emphasize that group theory identifies and formalizes symmetries. It classifies (and sometimes predicts) particles. But aside from saying that one part of the Hamiltonian has SU(2)

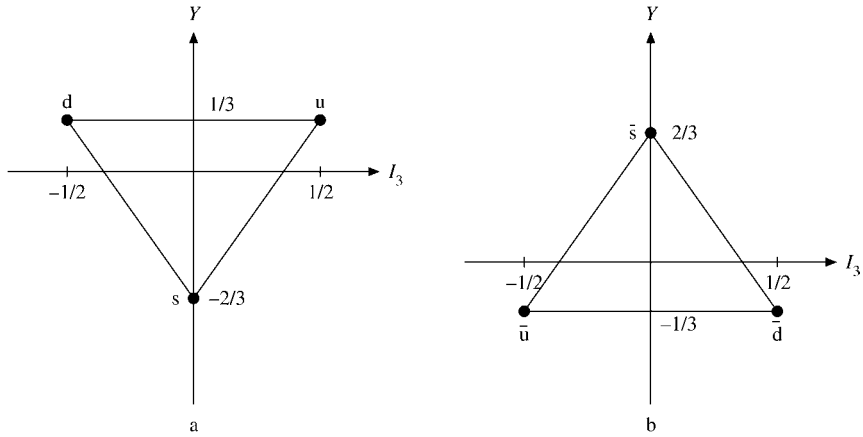

FIGURE 4.5 (a) Fundamental representation of SU(3), the weight diagram for the u, d, s quarks; (b) weight diagram for the antiquarks $\bar{u}, \bar{d}, \bar{s}$.

symmetry and another part has SU(3) symmetry, group theory says nothing about the particle interaction. Remember that the statement that the atomic potential is spherically symmetric tells us nothing about the radial dependence of the potential or of the wave function. In contrast, in a gauge theory the interaction is mediated by vector bosons (like the photon in quantum electrodynamics) and uniquely determined by the gauge covariant derivative (see Section 1.13).

Exercises

4.2.1 (i) Show that the Pauli matrices are the generators of SU(2) without using the parameterization of the general unitary 2×2 matrix in Eq. (4.38). (ii) Derive the eight independent generators λ_i of SU(3) similarly. Normalize them so that $\text{tr}(\lambda_i \lambda_j) = 2\delta_{ij}$. Then determine the structure constants of SU(3).
Hint. The λ_i are traceless and Hermitian 3×3 matrices.
(iii) Construct the quadratic Casimir invariant of SU(3).
Hint. Work by analogy with $\sigma_1^2 + \sigma_2^2 + \sigma_3^2$ of SU(2) or $\mathbf{L}^2$ of SO(3).

4.2.2 Prove that the general form of a 2×2 unitary, unimodular matrix is

$$\mathsf{U} = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}$$

with $a^*a + b^*b = 1$.

4.2.3 Determine three SU(2) subgroups of SU(3).

4.2.4 A **translation** operator $T(a)$ converts $\psi(x)$ to $\psi(x + a)$,

$$T(a)\psi(x) = \psi(x + a).$$

In terms of the (quantum mechanical) linear momentum operator $p_x = -id/dx$, show that $T(a) = \exp(iap_x)$, that is, p_x is the generator of translations.
Hint. Expand $\psi(x + a)$ as a Taylor series.

4.2.5 Consider the general SU(2) element Eq. (4.38) to be built up of three Euler rotations: (i) a rotation of $a/2$ about the z-axis, (ii) a rotation of $b/2$ about the new x-axis, and (iii) a rotation of $c/2$ about the new z-axis. (All rotations are counterclockwise.) Using the Pauli $\boldsymbol{\sigma}$ generators, show that these rotation angles are determined by

$$\begin{aligned} a &= \xi - \zeta + \tfrac{\pi}{2} = \alpha + \tfrac{\pi}{2} \\ b &= 2\eta \qquad\quad = \beta \\ c &= \xi + \zeta - \tfrac{\pi}{2} = \gamma - \tfrac{\pi}{2}. \end{aligned}$$

Note. The angles a and b here are not the a and b of Eq. (4.38).

4.2.6 Rotate a nonrelativistic wave function $\tilde{\psi} = (\psi_\uparrow, \psi_\downarrow)$ of spin $1/2$ about the z-axis by a small angle $d\theta$. Find the corresponding generator.

4.3 ORBITAL ANGULAR MOMENTUM

The classical concept of angular momentum, $\mathbf{L}_{\text{class}} = \mathbf{r} \times \mathbf{p}$, is presented in Section 1.4 to introduce the cross product. Following the usual Schrödinger representation of quantum mechanics, the classical linear momentum $\mathbf{p}$ is replaced by the operator $-i\nabla$. The quantum mechanical orbital angular momentum **operator** becomes[6]

$$\mathbf{L}_{QM} = -i\mathbf{r} \times \nabla. \tag{4.63}$$

This is used repeatedly in Sections 1.8, 1.9, and 2.4 to illustrate vector differential operators. From Exercise 1.8.8 the angular momentum components satisfy the commutation relations

$$[L_i, L_j] = i\varepsilon_{ijk} L_k. \tag{4.64}$$

The ε_{ijk} is the Levi-Civita symbol of Section 2.9. A summation over the index k is understood.

The differential operator corresponding to the square of the angular momentum

$$\mathbf{L}^2 = \mathbf{L} \cdot \mathbf{L} = L_x^2 + L_y^2 + L_z^2 \tag{4.65}$$

may be determined from

$$\mathbf{L} \cdot \mathbf{L} = (\mathbf{r} \times \mathbf{p}) \cdot (\mathbf{r} \times \mathbf{p}), \tag{4.66}$$

which is the subject of Exercises 1.9.9 and 2.5.17(b). Since $\mathbf{L}^2$ as a scalar product is invariant under rotations, that is, a rotational scalar, we expect $[\mathbf{L}^2, L_i] = 0$, which can also be verified directly.

Equation (4.64) presents the basic commutation relations of the components of the quantum mechanical angular momentum. Indeed, within the framework of quantum mechanics and group theory, these commutation relations define an angular momentum operator. We shall use them now to construct the angular momentum eigenstates and find the eigenvalues. For the orbital angular momentum these are the spherical harmonics of Section 12.6.

[6] For simplicity, $\hbar$ is set equal to 1. This means that the angular momentum is measured in units of $\hbar$.

Ladder Operator Approach

Let us start with a general approach, where the angular momentum $\mathbf{J}$ we consider may represent an orbital angular momentum $\mathbf{L}$, a spin $\boldsymbol{\sigma}/2$, or a total angular momentum $\mathbf{L}+\boldsymbol{\sigma}/2$, etc. We assume that

1. $\mathbf{J}$ is an Hermitian operator whose components satisfy the commutation relations
$$[J_i, J_j] = i\varepsilon_{ijk}J_k, \qquad \left[\mathbf{J}^2, J_i\right] = 0. \tag{4.67}$$
Otherwise $\mathbf{J}$ is arbitrary. (See Exercise 4.3.l.)
2. $|\lambda M\rangle$ is simultaneously a normalized eigenfunction (or eigenvector) of J_z with eigenvalue M and an eigenfunction[7] of $\mathbf{J}^2$,
$$J_z|\lambda M\rangle = M|\lambda M\rangle, \qquad \mathbf{J}^2|\lambda M\rangle = \lambda|\lambda M\rangle, \qquad \langle\lambda M|\lambda M\rangle = 1. \tag{4.68}$$

We shall show that $\lambda = J(J+1)$ and then find other properties of the $|\lambda M\rangle$. The treatment will illustrate the generality and power of operator techniques, particularly the use of ladder operators.[8]

The **ladder operators** are defined as
$$J_+ = J_x + iJ_y, \qquad J_- = J_x - iJ_y. \tag{4.69}$$
In terms of these operators $\mathbf{J}^2$ may be rewritten as
$$\mathbf{J}^2 = \tfrac{1}{2}(J_+J_- + J_-J_+) + J_z^2. \tag{4.70}$$
From the commutation relations, Eq. (4.67), we find
$$\boxed{[J_z, J_+] = +J_+, \qquad [J_z, J_-] = -J_-, \qquad [J_+, J_-] = 2J_z.} \tag{4.71}$$
Since J_+ commutes with $\mathbf{J}^2$ (Exercise 4.3.1),
$$\mathbf{J}^2\big(J_+|\lambda M\rangle\big) = J_+\big(\mathbf{J}^2|\lambda M\rangle\big) = \lambda\big(J_+|\lambda M\rangle\big). \tag{4.72}$$
Therefore, $J_+|\lambda M\rangle$ is still an eigenfunction of $\mathbf{J}^2$ with eigenvalue λ, and similarly for $J_-|\lambda M\rangle$. But from Eq. (4.71),
$$J_zJ_+ = J_+(J_z+1), \tag{4.73}$$
or
$$J_z\big(J_+|\lambda M\rangle\big) = J_+(J_z+1)|\lambda M\rangle = (M+1)J_+|\lambda M\rangle. \tag{4.74}$$

[7] That $|\lambda M\rangle$ can be an eigenfunction of **both** J_z and $\mathbf{J}^2$ follows from $[J_z, \mathbf{J}^2] = 0$ in Eq. (4.67). For SU(2), $\langle\lambda M|\lambda M\rangle$ is the scalar product (of the bra and ket vector or spinors) in the bra-ket notation introduced in Section 3.1. For SO(3), $|\lambda M\rangle$ is a function $Y(\theta,\varphi)$ and $|\lambda M'\rangle$ is a function $Y'(\theta,\varphi)$ and the matrix element $\langle\lambda M|\lambda M'\rangle \equiv \int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\pi} Y^*(\theta,\varphi)Y'(\theta,\varphi)\sin\theta\,d\theta\,d\varphi$ is their overlap. However, in our algebraic approach only the norm in Eq. (4.68) is used and matrix elements of the angular momentum operators are reduced to the norm by means of the eigenvalue equation for J_z, Eq. (4.68), and Eqs. (4.83) and (4.84).

[8] Ladder operators can be developed for other mathematical functions. Compare the next subsection, on other Lie groups, and Section 13.1, for Hermite polynomials.

Therefore, $J_+|\lambda M\rangle$ is still an eigenfunction of J_z but with eigenvalue $M+1$. J_+ has raised the eigenvalue by 1 and so is called a **raising operator**. Similarly, J_- lowers the eigenvalue by 1 and is called a **lowering operator**.

Taking expectation values and using $J_x^\dagger = J_x$, $J_y^\dagger = J_y$, we get

$$\langle \lambda M|\mathbf{J}^2 - J_z^2|\lambda M\rangle = \langle \lambda M|J_x^2 + J_y^2|\lambda M\rangle = \big|J_x|\lambda M\rangle\big|^2 + \big|J_y|\lambda M\rangle\big|^2$$

and see that $\lambda - M^2 \geq 0$, so M is bounded. Let J be the **largest** M. Then $J_+|\lambda J\rangle = 0$, which implies $J_-J_+|\lambda J\rangle = 0$. Hence, combining Eqs. (4.70) and (4.71) to get

$$\mathbf{J}^2 = J_-J_+ + J_z(J_z+1), \tag{4.75}$$

we find from Eq. (4.75) that

$$0 = J_-J_+|\lambda J\rangle = \big(\mathbf{J}^2 - J_z^2 - J_z\big)|\lambda J\rangle = \big(\lambda - J^2 - J\big)|\lambda J\rangle.$$

Therefore

$$\lambda = J(J+1) \geq 0, \tag{4.76}$$

with nonnegative J. We now relabel the states $|\lambda M\rangle \equiv |JM\rangle$. Similarly, let J' be the **smallest** M. Then $J_-|JJ'\rangle = 0$. From

$$\mathbf{J}^2 = J_+J_- + J_z(J_z-1), \tag{4.77}$$

we see that

$$0 = J_+J_-|JJ'\rangle = \big(\mathbf{J}^2 + J_z - J_z^2\big)|JJ'\rangle = \big(\lambda + J' - J'^2\big)|JJ'\rangle. \tag{4.78}$$

Hence

$$\lambda = J(J+1) = J'(J'-1) = (-J)(-J-1).$$

So $J' = -J$, and M runs in **integer steps** from $-J$ to $+J$,

$$-J \leq M \leq J. \tag{4.79}$$

Starting from $|JJ\rangle$ and applying J_- repeatedly, we reach all other states $|JM\rangle$. Hence the $|JM\rangle$ form an irreducible representation of SO(3) or SU(2); M varies and J is fixed.

Then using Eqs. (4.67), (4.75), and (4.77) we obtain

$$\begin{aligned} J_-J_+|JM\rangle &= \big[J(J+1) - M(M+1)\big]|JM\rangle = (J-M)(J+M+1)|JM\rangle, \\ J_+J_-|JM\rangle &= \big[J(J+1) - M(M-1)\big]|JM\rangle = (J+M)(J-M+1)|JM\rangle. \end{aligned} \tag{4.80}$$

Because J_+ and J_- are Hermitian conjugates,[9]

$$J_+^\dagger = J_-, \qquad J_-^\dagger = J_+, \tag{4.81}$$

the eigenvalues in Eq. (4.80) must be positive or zero.[10] Examples of Eq. (4.81) are provided by the matrices of Exercise 3.2.13 (spin 1/2), 3.2.15 (spin 1), and 3.2.18 (spin 3/2).

[9]The Hermitian conjugation or adjoint operation is defined for matrices in Section 3.5, and for operators in general in Section 10.1.

[10]For an excellent discussion of adjoint operators and Hilbert space see A. Messiah, *Quantum Mechanics*. New York: Wiley 1961, Chapter 7.

For the orbital angular momentum ladder operators, L_+, and L_-, explicit forms are given in Exercises 2.5.14 and 12.6.7. You can now show (see also Exercise 12.7.2) that

$$\langle JM|J_-\big(J_+|JM\rangle\big) = \big(J_+|JM\rangle\big)^\dagger J_+|JM\rangle. \tag{4.82}$$

Since J_+ raises the eigenvalue M to $M+1$, we relabel the resultant eigenfunction $|JM+1\rangle$. The normalization is given by Eq. (4.80) as

$$J_+|JM\rangle = \sqrt{(J-M)(J+M+1)}|JM+1\rangle = \sqrt{J(J+1)-M(M+1)}|JM+1\rangle, \tag{4.83}$$

taking the positive square root and not introducing any phase factor. By the same arguments,

$$J_-|JM\rangle = \sqrt{(J+M)(J-M+1)}|JM-1\rangle = \sqrt{(J(J+1)-M(M-1)}|JM-1\rangle. \tag{4.84}$$

Applying J_+ to Eq. (4.84), we obtain the second line of Eq. (4.80) and verify that Eq. (4.84) is consistent with Eq. (4.83).

Finally, since M ranges from $-J$ to $+J$ in unit steps, $2J$ must be an integer; J is either an integer or half of an odd integer. As seen later, if **J** is an orbital angular momentum **L**, the set $|LM\rangle$ for all M is a basis defining a representation of SO(3) and L will then be integral. In spherical polar coordinates θ, φ, the functions $|LM\rangle$ become the spherical harmonics $Y_L^M(\theta,\varphi)$ of Section 12.6. The sets of $|JM\rangle$ states with half-integral J define representations of SU(2) that are not representations of SO(3); we get $J = 1/2, 3/2, 5/2, \ldots$. Our angular momentum is quantized, essentially as a result of the commutation relations. All these representations are irreducible, as an application of the raising and lowering operators suggests.

Summary of Lie Groups and Lie Algebras

The general commutation relations, Eq. (4.14) in Section 4.2, for a classical Lie group [SO(n) and SU(n) in particular] can be simplified to look more like Eq. (4.71) for SO(3) and SU(2) in this section. Here we merely review and, as a rule, do not provide proofs for various theorems that we explain.

First we choose linearly independent and mutually commuting generators H_i which are generalizations of J_z for SO(3) and SU(2). Let l be the maximum number of such H_i with

$$[H_i, H_k] = 0. \tag{4.85}$$

Then l is called the **rank** of the Lie group G or its Lie algebra $\mathcal{G}$. The rank and dimension, or order, of some Lie groups are given in Table 4.2. All other generators E_α can be shown to be raising and lowering operators with respect to all the H_i, so

$$[H_i, E_\alpha] = \alpha_i E_\alpha, \qquad i = 1, 2, \ldots, l. \tag{4.86}$$

The set of so-called **root vectors** $(\alpha_1, \alpha_2, \ldots, \alpha_l)$ form the **root diagram** of $\mathcal{G}$.

When the H_i commute, they can be simultaneously diagonalized (for symmetric (or Hermitian) matrices see Chapter 3; for operators see Chapter 10). The H_i provide us with a set of eigenvalues $m_1, m_2, \ldots, m_l$ [projection or additive quantum numbers generalizing

Table 4.2 Rank and Order of Unitary and Rotational Groups

Lie algebra	$\mathcal{A}_l$	$\mathcal{B}_l$	$\mathcal{D}_l$
Lie group	$\mathsf{SU}(l+1)$	$\mathsf{SO}(2l+1)$	$\mathsf{SO}(2l)$
Rank	l	l	l
Order	$l(l+2)$	$l(2l+1)$	$l(2l-1)$

M of J_z in $\mathsf{SO}(3)$ and $\mathsf{SU}(2)$]. The set of so-called **weight vectors** $(m_1, m_2, \ldots, m_l)$ for an irreducible representation (multiplet) form a **weight diagram**.

There are l invariant operators C_i, called **Casimir** operators, that commute with all generators and are generalizations of $\mathbf{J}^2$,

$$[C_i, H_j] = 0, \qquad [C_i, E_\alpha] = 0, \qquad i = 1, 2, \ldots, l. \tag{4.87}$$

The first one, C_1, is a quadratic function of the generators; the others are more complicated. Since the C_j commute with all H_j, they can be simultaneously diagonalized with the H_j. Their eigenvalues $c_1, c_2, \ldots, c_l$ characterize irreducible representations and stay constant while the weight vector varies over any particular irreducible representation. Thus the general eigenfunction may be written as

$$\big|(c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l\big\rangle, \tag{4.88}$$

generalizing the multiplet $|JM\rangle$ of $\mathsf{SO}(3)$ and $\mathsf{SU}(2)$. Their eigenvalue equations are

$$H_i\big|(c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l\big\rangle = m_i\big|(c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l\big\rangle \tag{4.89a}$$

$$C_i\big|(c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l\big\rangle = c_i\big|(c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l\big\rangle. \tag{4.89b}$$

We can now show that $E_\alpha|(c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l\rangle$ has the weight vector $(m_1 + \alpha_1, m_2 + \alpha_2, \ldots, m_l + \alpha_l)$ using the commutation relations, Eq. (4.86), in conjunction with Eqs. (4.89a) and (4.89b):

$$\begin{aligned} &H_i E_\alpha\big|(c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l\big\rangle \\ &\quad = \big(E_\alpha H_i + [H_i, E_\alpha]\big)\big|(c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l\big\rangle \\ &\quad = (m_i + \alpha_i)E_\alpha\big|(c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l\big\rangle. \end{aligned} \tag{4.90}$$

Therefore

$$E_\alpha\big|(c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l\big\rangle \sim \big|(c_1, \ldots, c_l)m_1 + \alpha_1, \ldots, m_l + \alpha_l\big\rangle,$$

the generalization of Eqs. (4.83) and (4.84) from $\mathsf{SO}(3)$. These changes of eigenvalues by the operator E_α are called its **selection rules** in quantum mechanics. They are displayed in the root diagram of a Lie algebra.

Examples of root diagrams are given in Fig. 4.6 for $\mathsf{SU}(2)$ and $\mathsf{SU}(3)$. If we attach the roots denoted by arrows in Fig. 4.6b to a weight in Figs. 4.3 or 4.5a, b, we can reach any other state (represented by a dot in the weight diagram).

Here **Schur's lemma** applies: An operator H that commutes with all group operators, and therefore with all generators H_i of a (classical) Lie group G in particular, has as eigenvectors all states of a multiplet and is degenerate with the multiplet. As a consequence, such an operator commutes with all Casimir invariants, $[H, C_i] = 0$.

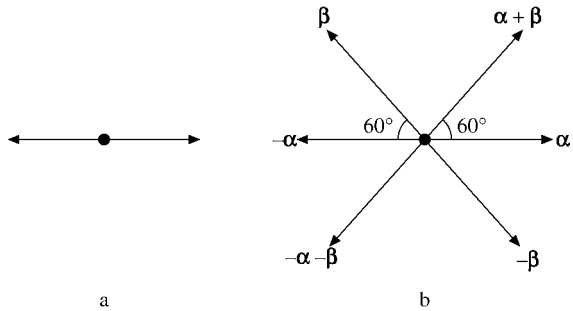

FIGURE 4.6 Root diagram for (a) SU(2) and (b) SU(3).

The last result is clear because the Casimir invariants are constructed from the generators and raising and lowering operators of the group. To prove the rest, let ψ be an eigenvector, $H\psi = E\psi$. Then, for any rotation R of G, we have $HR\psi = ER\psi$, which says that $R\psi$ is an eigenstate with the same eigenvalue E along with ψ. Since $[H, C_i] = 0$, all Casimir invariants can be diagonalized simultaneously with H and an eigenstate of H is an eigenstate of all the C_i. Since $[H_i, C_i] = 0$, the rotated eigenstates $R\psi$ are eigenstates of C_i, along with ψ belonging to the same multiplet characterized by the eigenvalues c_i of C_i.

Finally, such an operator H cannot induce transitions between different multiplets of the group because

$$\langle (c'_1, c'_2, \ldots, c'_l) m'_1, m'_2, \ldots, m'_l | H | (c_1, c_2, \ldots, c_l) m_1, m_2, \ldots, m_l \rangle = 0.$$

Using $[H, C_j] = 0$ (for any j) we have

$$\begin{aligned} 0 &= \langle (c'_1, c'_2, \ldots, c'_l) m'_1, m'_2, \ldots, m'_l | [H, C_j] | (c_1, c_2, \ldots, c_l) m_1, m_2, \ldots, m_l \rangle \\ &= (c_j - c'_j) \langle (c'_1, c'_2, \ldots, c'_l) m'_1, m'_2, \ldots, m'_l | H | (c_1, c_2, \ldots, c_l) m_1, m_2, \ldots, m_l \rangle. \end{aligned}$$

If $c'_j \neq c_j$ for some j, then the previous equation follows.

Exercises

4.3.1 Show that (a) $[J_+, \mathbf{J}^2] = 0$, (b) $[J_-, \mathbf{J}^2] = 0$.

4.3.2 Derive the root diagram of SU(3) in Fig. 4.6b from the generators λ_i in Eq. (4.61). *Hint*. Work out first the SU(2) case in Fig. 4.6a from the Pauli matrices.

4.4 ANGULAR MOMENTUM COUPLING

In many-body systems of classical mechanics, the total angular momentum is the sum $\mathbf{L} = \sum_i \mathbf{L}_i$ of the individual orbital angular momenta. Any isolated particle has conserved angular momentum. In quantum mechanics, conserved angular momentum arises when particles move in a central potential, such as the Coulomb potential in atomic physics, a shell model potential in nuclear physics, or a confinement potential of a quark model in

particle physics. In the relativistic Dirac equation, orbital angular momentum is no longer conserved, but $\mathbf{J} = \mathbf{L} + \mathbf{S}$ is conserved, the total angular momentum of a particle consisting of its orbital and intrinsic angular momentum, called spin $\mathbf{S} = \boldsymbol{\sigma}/2$, in units of $\hbar$.

It is readily shown that the sum of angular momentum operators obeys the same commutation relations in Eq. (4.37) or (4.41) as the individual angular momentum operators, provided those from different particles commute.

Clebsch–Gordan Coefficients: SU(2)–SO(3)

Clearly, combining two commuting angular momenta $\mathbf{J}_i$ to form their sum

$$\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2, \qquad [J_{1i}, J_{2i}] = 0, \tag{4.91}$$

occurs often in applications, and $\mathbf{J}$ satisfies the angular momentum commutation relations

$$[J_j, J_k] = [J_{1j} + J_{2j}, J_{1k} + J_{2k}] = [J_{1j}, J_{1k}] + [J_{2j}, J_{2k}] = i\varepsilon_{jkl}(J_{1l} + J_{2l}) = i\varepsilon_{jkl}J_l.$$

For a single particle with spin $1/2$, for example, an electron or a quark, the total angular momentum is a sum of orbital angular momentum and spin. For two spinless particles their total orbital angular momentum $\mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2$. For $\mathbf{J}^2$ and J_z of Eq. (4.91) to be both diagonal, $[\mathbf{J}^2, J_z] = 0$ has to hold. To show this we use the obvious commutation relations $[J_{iz}, \mathbf{J}_j^2] = 0$, and

$$\mathbf{J}^2 = \mathbf{J}_1^2 + \mathbf{J}_2^2 + 2\mathbf{J}_1 \cdot \mathbf{J}_2 = \mathbf{J}_1^2 + \mathbf{J}_2^2 + J_{1+}J_{2-} + J_{1-}J_{2+} + 2J_{1z}J_{2z} \tag{4.91$'$}$$

in conjunction with Eq. (4.71), for both $\mathbf{J}_i$, to obtain

$$\begin{aligned}\left[\mathbf{J}^2, J_z\right] &= [J_{1-}J_{2+} + J_{1+}J_{2-}, J_{1z} + J_{2z}] \\ &= [J_{1-}, J_{1z}]J_{2+} + J_{1-}[J_{2+}, J_{2z}] + [J_{1+}, J_{1z}]J_{2-} + J_{1+}[J_{2-}, J_{2z}] \\ &= J_{1-}J_{2+} - J_{1-}J_{2+} - J_{1+}J_{2-} + J_{1+}J_{2-} = 0.\end{aligned}$$

Similarly $[\mathbf{J}^2, \mathbf{J}_i^2] = 0$ is proved. Hence the eigenvalues of $\mathbf{J}_i^2, \mathbf{J}^2, J_z$ can be used to label the total angular momentum states $|J_1 J_2 J M\rangle$.

The product states $|J_1 m_1\rangle|J_2 m_2\rangle$ obviously satisfy the eigenvalue equations

$$\begin{aligned}J_z|J_1 m_1\rangle|J_2 m_2\rangle &= (J_{1z} + J_{2z})|J_1 m_1\rangle|J_2 m_2\rangle = (m_1 + m_2)|J_1 m_1\rangle|J_2 m_2\rangle \\ &= M|J_1 m_1\rangle|J_2 m_2\rangle, \\ \mathbf{J}_i^2|J_1 m_1\rangle|J_2 m_2\rangle &= J_i(J_i + 1)|J_1 m_1\rangle|J_2 m_2\rangle,\end{aligned} \tag{4.92}$$

but will not have diagonal $\mathbf{J}^2$ except for the maximally stretched states with $M = \pm(J_1 + J_2)$ and $J = J_1 + J_2$ (see Fig. 4.7a). To see this we use Eq. (4.91$'$) again in conjunction with Eqs. (4.83) and (4.84) in

$$\begin{aligned}\mathbf{J}^2|J_1 m_1\rangle J_2 m_2\rangle &= \left\{J_1(J_1 + 1) + J_2(J_2 + 1) + 2m_1 m_2\right\}|J_1 m_1\rangle|J_2 m_2\rangle \\ &+ \left\{J_1(J_1 + 1) - m_1(m_1 + 1)\right\}^{1/2}\left\{J_2(J_2 + 2) - m_2(m_2 - 1)\right\}^{1/2} \\ &\times |J_1 m_1 + 1\rangle|J_2 m_2 - 1\rangle + \left\{J_1(J_1 + 1) - m_1(m_1 - 1)\right\}^{1/2} \\ &\times \left\{J_2(J_2 + 1) - m_2(m_2 + 1)\right\}^{1/2}|J_1 m_1 - 1\rangle|J_2 m_2 + 1\rangle.\end{aligned} \tag{4.93}$$

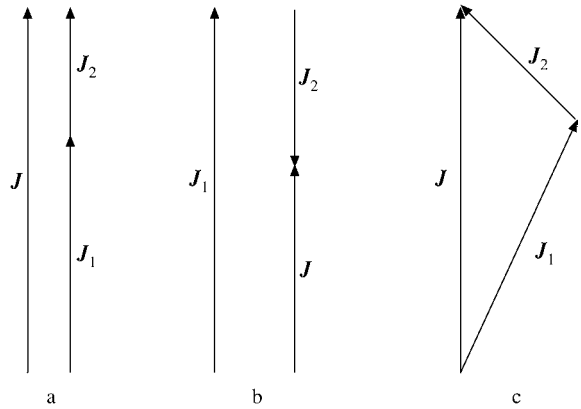

FIGURE 4.7 Coupling of two angular momenta: (a) parallel stretched, (b) antiparallel, (c) general case.

The last two terms in Eq. (4.93) vanish only when $m_1 = J_1$ and $m_2 = J_2$ or $m_1 = -J_1$ and $m_2 = -J_2$. In both cases $J = J_1 + J_2$ follows from the first line of Eq. (4.93). In general, therefore, we have to form appropriate linear combinations of product states

$$|J_1 J_2 J M\rangle = \sum_{m_1, m_2} C\big(J_1 J_2 J | m_1 m_2 M\big) |J_1 m_1\rangle |J_2 m_2\rangle, \tag{4.94}$$

so that $\mathbf{J}^2$ has eigenvalue $J(J+1)$. The quantities $C(J_1 J_2 J | m_1 m_2 M)$ in Eq. (4.94) are called **Clebsch–Gordan coefficients**. From Eq. (4.92) we see that they vanish unless $M = m_1 + m_2$, reducing the double sum to a single sum. Applying $J_\pm$ to $|JM\rangle$ shows that the eigenvalues M of J_z satisfy the usual inequalities $-J \le M \le J$.

Clearly, the maximal $J_{\max} = J_1 + J_2$ (see Fig. 4.7a). In this case Eq. (4.93) reduces to a pure product state

$$|J_1 J_2 J = J_1 + J_2 M = J_1 + J_2\rangle = |J_1 J_1\rangle |J_2 J_2\rangle, \tag{4.95a}$$

so the Clebsch–Gordan coefficient

$$C(J_1 J_2 J = J_1 + J_2 | J_1 J_2 J_1 + J_2) = 1. \tag{4.95b}$$

The minimal $J = J_1 - J_2$ (if $J_1 > J_2$, see Fig. 4.7b) and $J = J_2 - J_1$ for $J_2 > J_1$ follow if we keep in mind that there are just as many product states as $|JM\rangle$ states; that is,

$$\begin{aligned} \sum_{J=J_{\min}}^{J_{\max}} (2J+1) &= (J_{\max} - J_{\min} + 1)(J_{\max} + J_{\min} + 1) \\ &= (2J_1 + 1)(2J_2 + 1). \end{aligned} \tag{4.96}$$

This condition holds because the $|J_1 J_2 J M\rangle$ states merely rearrange all product states into irreducible representations of total angular momentum. It is equivalent to the **triangle rule**:

$$\begin{aligned} &\Delta(J_1 J_2 J) = 1, \qquad \text{if } |J_1 - J_2| \le J \le J_1 + J_2; \\ &\Delta(J_1 J_2 J) = 0, \qquad \text{else.} \end{aligned} \tag{4.97}$$

This indicates that one complete multiplet of each J value from $J_{\min}$ to $J_{\max}$ accounts for all the states and that all the $|JM\rangle$ states are necessarily orthogonal. In other words, Eq. (4.94) defines a unitary transformation from the orthogonal basis set of products of single-particle states $|J_1 m_1; J_2 m_2\rangle = |J_1 m_1\rangle|J_2 m_2\rangle$ to the two-particle states $|J_1 J_2 JM\rangle$. The Clebsch–Gordan coefficients are just the overlap matrix elements

$$C(J_1 J_2 J|m_1 m_2 M) \equiv \langle J_1 J_2 JM|J_1 m_1; J_2 m_2\rangle. \tag{4.98}$$

The explicit construction in what follows shows that they are all real. The states in Eq. (4.94) are orthonormalized, provided that the constraints

$$\begin{aligned}\sum_{m_1,m_2,\ m_1+m_2=M} & C(J_1 J_2 J|m_1 m_2 M)C(J_1 J_2 J'|m_1 m_2 M')\\ &= \langle J_1 J_2 JM|J_1 J_2 J'M'\rangle = \delta_{JJ'}\delta_{MM'}\end{aligned} \tag{4.99a}$$

$$\begin{aligned}\sum_{J,M} & C(J_1 J_2 J|m_1 m_2 M)C(J_1 J_2 J|m_1' m_2' M)\\ &= \langle J_1 m_1|J_1 m_1'\rangle\langle J_2 m_2|J_2 m_2'\rangle = \delta_{m_1 m_1'}\delta_{m_2 m_2'}\end{aligned} \tag{4.99b}$$

hold.

Now we are ready to construct more directly the total angular momentum states starting from $|J_{\max} = J_1 + J_2\ M = J_1 + J_2\rangle$ in Eq. (4.95a) and using the lowering operator $J_- = J_{1-} + J_{2-}$ repeatedly. In the first step we use Eq. (4.84) for

$$J_{i-}|J_i J_i\rangle = \left\{J_i(J_i+1) - J_i(J_i-1)\right\}^{1/2}|J_i J_i - 1\rangle = (2J_i)^{1/2}|J_i J_i - 1\rangle,$$

which we substitute into $(J_{1-} + J_{2-})|J_1 J_1\rangle|J_2 J_2\rangle$. Normalizing the resulting state with $M = J_1 + J_2 - 1$ properly to 1, we obtain

$$\begin{aligned}|J_1 J_2 J_1 + J_2 J_1 + J_2 - 1\rangle = & \left\{J_1/(J_1+J_2)\right\}^{1/2}|J_1 J_1 - 1\rangle|J_2 J_2\rangle\\ & + \left\{J_2/(J_1+J_2)\right\}^{1/2}|J_1 J_1\rangle|J_2 J_2 - 1\rangle.\end{aligned} \tag{4.100}$$

Equation (4.100) yields the Clebsch–Gordan coefficients

$$\begin{aligned}C(J_1 J_2 J_1 + J_2|J_1 - 1\ J_2\ J_1 + J_2 - 1) &= \left\{J_1/(J_1+J_2)\right\}^{1/2},\\ C(J_1 J_2 J_1 + J_2|J_1\ J_2 - 1\ J_1 + J_2 - 1) &= \left\{J_2/(J_1+J_2)\right\}^{1/2}.\end{aligned} \tag{4.101}$$

Then we apply J_- again and normalize the states obtained until we reach $|J_1 J_2\ J_1 + J_2 M\rangle$ with $M = -(J_1 + J_2)$. The Clebsch–Gordan coefficients $C(J_1 J_2 J_1 + J_2|m_1 m_2 M)$ may thus be calculated step by step, and they are all real.

The next step is to realize that the only other state with $M = J_1 + J_2 - 1$ is the top of the next lower tower of $|J_1 + J_2 - 1 M\rangle$ states. Since $|J_1 + J_2 - 1\ J_1 + J_2 - 1\rangle$ is orthogonal to $|J_1 + J_2 J_1 + J_2 - 1\rangle$ in Eq. (4.100), it must be the other linear combination with a relative minus sign,

$$\begin{aligned}|J_1 + J_2 - 1\ J_1 + J_2 - 1\rangle = & -\left\{J_2/(J_1+J_2)\right\}^{1/2}|J_1 J_1 - 1\rangle|J_2 J_2\rangle\\ & + \left\{J_1/(J_1+J_2)\right\}^{1/2}|J_1 J_1\rangle|J_2 J_2 - 1\rangle,\end{aligned} \tag{4.102}$$

up to an overall sign.

Hence we have determined the Clebsch–Gordan coefficients (for $J_2 \geq J_1$)

$$\begin{aligned} C(J_1 J_2\ J_1+J_2-1|J_1-1\ J_2\ J_1+J_2-1) &= -\left\{J_2/(J_1+J_2)\right\}^{1/2}, \\ C(J_1 J_2\ J_1+J_2-1|J_1\ J_2-1\ J_1+J_2-1) &= \left\{J_1/(J_1+J_2)\right\}^{1/2}. \end{aligned} \tag{4.103}$$

Again we continue using J_- until we reach $M=-(J_1+J_2-1)$, and we keep normalizing the resulting states $|J_1+J_2-1M\rangle$ of the $J=J_1+J_2-1$ tower.

In order to get to the top of the next tower, $|J_1+J_2-2M\rangle$ with $M=J_1+J_2-2$, we remember that we have already constructed two states with that M. Both $|J_1+J_2 J_1+J_2-2\rangle$ and $|J_1+J_2-1\ J_1+J_2-2\rangle$ are known linear combinations of the three product states $|J_1 J_1\rangle|J_2 J_2-2\rangle$, $|J_1 J_1-1\rangle\times|J_2 J_2-1\rangle$, and $|J_1 J_1-2\rangle|J_2 J_2\rangle$. The third linear combination is easy to find from orthogonality to these two states, up to an overall phase, which is chosen by the **Condon–Shortley phase conventions**[11] so that the coefficient $C(J_1 J_2\ J_1+J_2-2|J_1\ J_2-2\ J_1+J_2-2)$ of the last product state is positive for $|J_1 J_2\ J_1+J_2-2\ J_1+J_2-2\rangle$. It is straightforward, though a bit tedious, to determine the rest of the Clebsch–Gordan coefficients.

Numerous recursion relations can be derived from matrix elements of various angular momentum operators, for which we refer to the literature.[12]

The symmetry properties of Clebsch–Gordan coefficients are best displayed in the more symmetric Wigner's $3j$-symbols, which are tabulated:[12]

$$\begin{pmatrix} J_1 J_2 J_3 \\ m_1 m_2 m_3 \end{pmatrix} = \frac{(-1)^{J_1-J_2-m_3}}{(2J_3+1)^{1/2}} C(J_1 J_2 J_3|m_1 m_2,-m_3), \tag{4.104a}$$

obeying the symmetry relations

$$\begin{pmatrix} J_1 J_2 J_3 \\ m_1 m_2 m_3 \end{pmatrix} = (-1)^{J_1+J_2+J_3} \begin{pmatrix} J_k J_l J_n \\ m_k m_l m_n \end{pmatrix} \tag{4.104b}$$

for (k,l,n) an odd permutation of $(1,2,3)$. One of the most important places where Clebsch–Gordan coefficients occur is in matrix elements of tensor operators, which are governed by the Wigner–Eckart theorem discussed in the next section, on spherical tensors. Another is coupling of operators or state vectors to total angular momentum, such as spin-orbit coupling. Recoupling of operators and states in matrix elements leads to $6j$- and $9j$-symbols.[12] Clebsch–Gordan coefficients can and have been calculated for other Lie groups, such as SU(3).

[11] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra*. Cambridge, UK: Cambridge University Press (1935).

[12] There is a rich literature on this subject, e.g., A. R. Edmonds, *Angular Momentum in Quantum Mechanics*. Princeton, NJ: Princeton University Press (1957); M. E. Rose, *Elementary Theory of Angular Momentum*. New York: Wiley (1957); A. de-Shalit and I. Talmi, *Nuclear Shell Model*. New York: Academic Press (1963); Dover (2005). Clebsch–Gordan coefficients are tabulated in M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Wooten, Jr., *The 3j- and 6j-Symbols*. Cambridge, MA: Massachusetts Institute of Technology Press (1959).

Spherical Tensors

In Chapter 2 the properties of Cartesian tensors are defined using the group of nonsingular general linear transformations, which contains the three-dimensional rotations as a subgroup. A tensor of a given rank that is irreducible with respect to the full group may well become reducible for the rotation group SO(3). To explain this point, consider the second-rank tensor with components $T_{jk} = x_j y_k$ for $j, k = 1, 2, 3$. It contains the symmetric tensor $S_{jk} = (x_j y_k + x_k y_j)/2$ and the antisymmetric tensor $A_{jk} = (x_j y_k - x_k y_j)/2$, so $T_{jk} = S_{jk} + A_{jk}$. This reduces T_{jk} in SO(3). However, under rotations the scalar product $\mathbf{x} \cdot \mathbf{y}$ is invariant and is therefore irreducible in SO(3). Thus, S_{jk} can be reduced by subtraction of the multiple of $\mathbf{x} \cdot \mathbf{y}$ that makes it traceless. This leads to the SO(3)-irreducible tensor

$$S'_{jk} = \tfrac{1}{2}(x_j y_k + x_k y_j) - \tfrac{1}{3}\mathbf{x} \cdot \mathbf{y}\delta_{jk}.$$

Tensors of higher rank may be treated similarly. When we form tensors from products of the components of the coordinate vector $\mathbf{r}$ then, in polar coordinates that are tailored to SO(3) symmetry, we end up with the spherical harmonics of Chapter 12.

The form of the ladder operators for SO(3) in Section 4.3 leads us to introduce the **spherical components** (note the different normalization and signs, though, prescribed by the Y_{lm}) of a vector $\mathbf{A}$:

$$A_{+1} = -\tfrac{1}{\sqrt{2}}(A_x + iA_y), \qquad A_{-1} = \tfrac{1}{\sqrt{2}}(A_x - iA_y), \qquad A_0 = A_z. \tag{4.105}$$

Then we have for the coordinate vector $\mathbf{r}$ in polar coordinates,

$$r_{+1} = -\tfrac{1}{\sqrt{2}} r \sin\theta e^{i\varphi} = r\sqrt{\tfrac{4\pi}{3}} Y_{11}, \qquad r_{-1} = \tfrac{1}{\sqrt{2}} r \sin\theta e^{-i\varphi} = r\sqrt{\tfrac{4\pi}{3}} Y_{1,-1},$$
$$r_0 = r\sqrt{\tfrac{4\pi}{3}} Y_{10}, \tag{4.106}$$

where $Y_{lm}(\theta, \varphi)$ are the spherical harmonics of Chapter 12. Again, the spherical jm components of tensors T_{jm} of higher rank j may be introduced similarly.

An irreducible **spherical tensor operator** T_{jm} of rank j has $2j + 1$ components, just as for spherical harmonics, and m runs from $-j$ to $+j$. Under a rotation $\mathsf{R}(\alpha)$, where α stands for the Euler angles, the Y_{lm} transform as

$$Y_{lm}(\hat{\mathbf{r}}') = \sum_{m'} Y_{lm'}(\hat{\mathbf{r}}) D^l_{m'm}(\mathsf{R}), \tag{4.107a}$$

where $\hat{\mathbf{r}}' = (\theta', \varphi')$ are obtained from $\hat{\mathbf{r}} = (\theta, \varphi)$ by the rotation R and are the angles of the same point in the rotated frame, and

$$\mathsf{D}^J_{m'm}(\alpha, \beta, \gamma) = \langle Jm | \exp(i\alpha J_z) \exp(i\beta J_y) \exp(i\gamma J_z) | Jm' \rangle$$

are the rotation matrices. So, for the operator T_{jm}, we define

$$\mathsf{R} T_{jm} \mathsf{R}^{-1} = \sum_{m'} T_{jm'} D^j_{m'm}(\alpha). \tag{4.107b}$$

For an infinitesimal rotation (see Eq. (4.20) in Section 4.2 on generators) the left side of Eq. (4.107b) simplifies to a commutator and the right side to the matrix elements of **J**, the infinitesimal generator of the rotation R:

$$[J_n, T_{jm}] = \sum_{m'} T_{jm'} \langle jm'|J_n|jm\rangle. \tag{4.108}$$

If we substitute Eqs. (4.83) and (4.84) for the matrix elements of J_m we obtain the alternative transformation laws of a tensor operator,

$$[J_0, T_{jm}] = mT_{jm}, \qquad [J_\pm, T_{jm}] = T_{jm\pm1}\big\{(j-m)(j\pm m+1)\big\}^{1/2}. \tag{4.109}$$

We can use the Clebsch–Gordan coefficients of the previous subsection to couple two tensors of given rank to another rank. An example is the cross or vector product of two vectors **a** and **b** from Chapter 1. Let us write both vectors in spherical components, a_m and b_m. Then we verify that the tensor C_m of rank 1 defined as

$$C_m \equiv \sum_{m_1 m_2} C(111|m_1 m_2 m) a_{m_1} b_{m_2} = \frac{i}{\sqrt{2}}(\mathbf{a}\times\mathbf{b})_m. \tag{4.110}$$

Since C_m is a spherical tensor of rank 1 that is linear in the components of **a** and **b**, it must be proportional to the cross product, $C_m = N(\mathbf{a}\times\mathbf{b})_m$. The constant N can be determined from a special case, $\mathbf{a} = \hat{\mathbf{x}}, \mathbf{b} = \hat{\mathbf{y}}$, essentially writing $\hat{\mathbf{x}}\times\hat{\mathbf{y}} = \hat{\mathbf{z}}$ in spherical components as follows. Using

$$(\hat{\mathbf{z}})_0 = 1; \qquad (\hat{\mathbf{x}})_1 = -1/\sqrt{2}, \qquad (\hat{\mathbf{x}})_{-1} = 1/\sqrt{2};$$
$$(\hat{\mathbf{y}})_1 = -i/\sqrt{2}, \qquad (\hat{\mathbf{y}})_{-1} = -i/\sqrt{2},$$

Eq. (4.110) for $m = 0$ becomes

$$C(111|1,-1,0)\big[(\hat{\mathbf{x}})_1(\hat{\mathbf{y}})_{-1} - (\hat{\mathbf{x}})_{-1}(\hat{\mathbf{y}})_1\big] = N\big((\hat{\mathbf{z}})_0\big) = N$$
$$= \frac{1}{\sqrt{2}}\left[-\frac{1}{\sqrt{2}}\left(-\frac{i}{\sqrt{2}}\right) - \frac{1}{\sqrt{2}}\left(-\frac{i}{\sqrt{2}}\right)\right] = \frac{i}{\sqrt{2}},$$

where we have used $C(111|101) = \frac{1}{\sqrt{2}}$ from Eq. (4.103) for $J_1 = 1 = J_2$, which implies $C(111|1,-1,0) = \frac{1}{\sqrt{2}}$ using Eqs. (4.104a,b):

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix} = -\frac{1}{\sqrt{3}} C(111|101) = -\frac{1}{6} = -\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} = -\frac{1}{\sqrt{3}} C(111|1,-1,0).$$

A bit simpler is the usual scalar product of two vectors in Chapter 1, in which **a** and **b** are coupled to zero angular momentum:

$$\mathbf{a}\cdot\mathbf{b} \equiv -(\mathbf{ab})_0\sqrt{3} \equiv -\sqrt{3}\sum_m C(110|m,-m,0) a_m b_{-m}. \tag{4.111}$$

Again, the rank zero of our tensor product implies $\mathbf{a}\cdot\mathbf{b} = n(\mathbf{ab})_0$. The constant n can be determined from a special case, essentially writing $\hat{\mathbf{z}}^2 = 1$ in spherical components: $\hat{\mathbf{z}}^2 = 1 = nC(110|000) = -\frac{n}{\sqrt{3}}$.

Another often-used application of tensors is the **recoupling** that involves **6j-symbols** for three operators and $9j$ for four operators.[12] An example is the following scalar product, for which it can be shown[12] that

$$\boldsymbol{\sigma}_1 \cdot \mathbf{r}\boldsymbol{\sigma}_2 \cdot \mathbf{r} = \frac{1}{3}\mathbf{r}^2 \boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2 + (\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)_2 \cdot (\mathbf{rr})_2, \tag{4.112}$$

but which can also be rearranged by elementary means. Here the tensor operators are defined as

$$(\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)_{2m} = \sum_{m_1 m_2} C(112|m_1 m_2 m)\sigma_{1m_1}\sigma_{2m_2}, \tag{4.113}$$

$$(\mathbf{rr})_{2m} = \sum_m C(112|m_1 m_2 m) r_{m_1} r_{m_2} = \sqrt{\frac{8\pi}{15}}\mathbf{r}^2 Y_{2m}(\hat{\mathbf{r}}), \tag{4.114}$$

and the scalar product of tensors of rank 2 as

$$(\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)_2 \cdot (\mathbf{rr})_2 = \sum_m (-1)^m (\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)_{2m}(\mathbf{rr})_{2,-m} = \sqrt{5}\big((\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)_2(\mathbf{rr})_2\big)_0. \tag{4.115}$$

One of the most important applications of spherical tensor operators is the **Wigner–Eckart theorem**. It says that a matrix element of a spherical tensor operator T_{km} of rank k between states of angular momentum j and j' factorizes into a Clebsch–Gordan coefficient and a so-called **reduced matrix element**, denoted by double bars, that no longer has any dependence on the projection quantum numbers m, m', n:

$$\langle j'm'|T_{kn}|jm\rangle = C(kjj'|nmm')(-1)^{k-j+j'}\langle j'\|T_k\|j\rangle/\sqrt{(2j'+1)}. \tag{4.116}$$

In other words, such a matrix element factors into a dynamic part, the reduced matrix element, and a geometric part, the Clebsch–Gordan coefficient that contains the rotational properties (expressed by the projection quantum numbers) from the SO(3) invariance. To see this we couple T_{kn} with the initial state to total angular momentum j':

$$|j'm'\rangle_0 \equiv \sum_{nm} C(kjj'|nmm')T_{kn}|jm\rangle. \tag{4.117}$$

Under rotations the state $|j'm'\rangle_0$ transforms just like $|j'm'\rangle$. Thus, the overlap matrix element $\langle j'm'|j'm'\rangle_0$ is a rotational scalar that has no m' dependence, so we can average over the projections,

$$\langle JM|j'm'\rangle_0 = \frac{\delta_{Jj'}\delta_{Mm'}}{2j'+1}\sum_\mu \langle j'\mu|j'\mu\rangle_0. \tag{4.118}$$

Next we substitute our definition, Eq. (4.117), into Eq. (4.118) and invert the relation Eq. (4.117) using orthogonality, Eq. (4.99b), to find that

$$\langle JM|T_{kn}|jm\rangle = \sum_{j'm'} C(kjj'|nmm')\frac{\delta_{Jj'}\delta_{Mm'}}{2J+1}\sum_\mu \langle J\mu|J\mu\rangle_0, \tag{4.119}$$

which proves the Wigner–Eckart theorem, Eq. (4.116).[13]

[13]The extra factor $(-1)^{k-j+j'}/\sqrt{(2j'+1)}$ in Eq. (4.116) is just a convention that varies in the literature.

As an application, we can write the Pauli matrix elements in terms of Clebsch–Gordan coefficients. We apply the Wigner–Eckart theorem to

$$\left\langle \tfrac{1}{2}\gamma \middle| \sigma_\alpha \middle| \tfrac{1}{2}\beta \right\rangle = (\sigma_\alpha)_{\gamma\beta} = -\tfrac{1}{\sqrt{2}} C\left(1\tfrac{1}{2}\tfrac{1}{2} \middle| \alpha\beta\gamma\right) \left\langle \tfrac{1}{2} \middle\| \sigma \middle\| \tfrac{1}{2} \right\rangle. \tag{4.120}$$

Since $\langle \frac{1}{2}\frac{1}{2}|\sigma_0|\frac{1}{2}\frac{1}{2}\rangle = 1$ with $\sigma_0 = \sigma_3$ and $C(1\frac{1}{2}\frac{1}{2} \mid 0\frac{1}{2}\frac{1}{2}) = -1/\sqrt{3}$, we find

$$\left\langle \tfrac{1}{2} \middle\| \sigma \middle\| \tfrac{1}{2} \right\rangle = \sqrt{6}, \tag{4.121}$$

which, substituted into Eq. (4.120), yields

$$(\sigma_\alpha)_{\gamma\beta} = -\sqrt{3} C\left(1\tfrac{1}{2}\tfrac{1}{2} \middle| \alpha\beta\gamma\right). \tag{4.122}$$

Note that the $\alpha = \pm 1, 0$ denote the spherical components of the Pauli matrices.

Young Tableaux for SU(n)

Young tableaux (YT) provide a powerful and elegant method for decomposing products of SU(n) group representations into sums of irreducible representations. The YT provide the dimensions and symmetry types of the irreducible representations in this so-called **Clebsch–Gordan series**, though not the Clebsch–Gordan coefficients by which the product states are coupled to the quantum numbers of each irreducible representation of the series (see Eq. (4.94)).

Products of representations correspond to multiparticle states. In this context, permutations of particles are important when we deal with several identical particles. Permutations of n identical objects form the **symmetric group** S_n. A close connection between irreducible representations of S_n, which are the YT, and those of SU(n) is provided by this **theorem**: Every N-particle state of S_n that is made up of single-particle states of the fundamental n-dimensional SU(n) multiplet belongs to an irreducible SU(n) representation. A proof is in Chapter 22 of Wybourne.[14]

For SU(2) the fundamental representation is a box that stands for the spin $+\frac{1}{2}$ (up) and $-\frac{1}{2}$ (down) states and has dimension 2. For SU(3) the box comprises the three quark states in the triangle of Fig. 4.5a; it has dimension 3.

An array of boxes shown in Fig. 4.8 with λ_1 boxes in the first row, λ_2 boxes in the second row, ..., and λ_{n-1} boxes in the last row is called a Young tableau (YT), denoted by $[\lambda_1, \ldots, \lambda_{n-1}]$, and represents an irreducible representation of SU(n) if and only if

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1}. \tag{4.123}$$

Boxes in the same row are symmetric representations; those in the same column are antisymmetric. A YT consisting of one row is totally symmetric. A YT consisting of a single column is totally antisymmetric.

There are at most $n-1$ rows for SU(n) YT because a column of n boxes is the totally antisymmetric (**Slater determinant** of single-particle states) singlet representation that may be struck from the YT.

An array of N boxes is an N-particle state whose boxes may be labeled by positive integers so that the (particle labels or) numbers in one row of the YT do not decrease from

[14] B. G. Wybourne, *Classical Groups for Physicists*. New York: Wiley (1974).

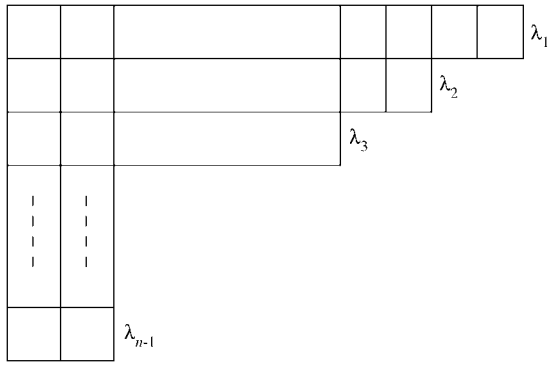

FIGURE 4.8 Young tableau (YT) for SU(n).

left to right and those in any one column increase from top to bottom. In contrast to the possible repetitions of row numbers, the numbers in any column must be different because of the antisymmetry of these states.

The product of a YT with a single box, [1], is the sum of YT formed when the box is put at the end of each row of the YT, provided the resulting YT is legitimate, that is, obeys Eq. (4.123). For SU(2) the product of two boxes, spin $1/2$ representations of dimension 2, generates

$$[1] \otimes [1] = [2] \oplus [1, 1], \tag{4.124}$$

the symmetric spin 1 representation of dimension 3 and the antisymmetric singlet of dimension 1 mentioned earlier.

The column of $n-1$ boxes is the conjugate representation of the fundamental representation; its product with a single box contains the column of n boxes, which is the singlet. For SU(3) the conjugate representation of the single box, [1] or fundamental quark representation, is the inverted triangle in Fig. 4.5b, [1, 1], which represents the three antiquarks $\bar{u}, \bar{d}, \bar{s}$, obviously of dimension 3 as well.

The dimension of a YT is given by the ratio

$$\dim \mathrm{YT} = \frac{N}{D}. \tag{4.125}$$

The numerator N is obtained by writing an n in all boxes of the YT along the diagonal, $(n+1)$ in all boxes immediately above the diagonal, $(n-1)$ immediately below the diagonal, etc. N is the product of all the numbers in the YT. An example is shown in Fig. 4.9a for the octet representation of SU(3), where $N = 2 \cdot 3 \cdot 4 = 24$. There is a closed formula that is equivalent to Eq. (4.125).[15] The denominator D is the product of all **hooks**.[16] A hook is drawn through each box of the YT by starting a horizontal line from the right to the box in question and then continuing it vertically out of the YT. The number of boxes encountered by the hook-line is the hook-number of the box. D is the product of all hook-numbers of

[15] See, for example, M. Hamermesh, *Group Theory and Its Application to Physical Problems*. Reading, MA: Addison-Wesley (1962).

[16] F. Close, *Introduction to Quarks and Partons*. New York: Academic Press (1979).

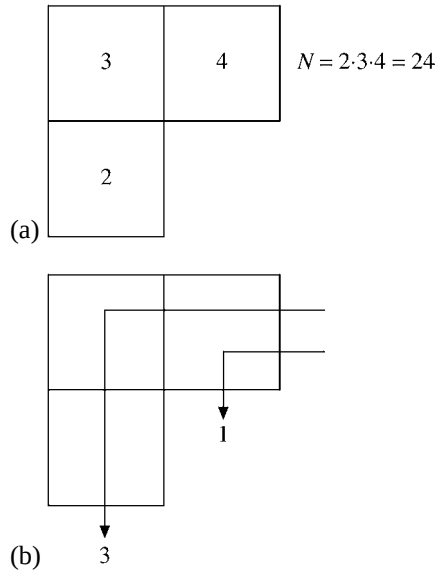

FIGURE 4.9 Illustration of (a) N and (b) D in Eq. (4.125) for the octet Young tableau of SU(3).

the YT. An example is shown in Fig. 4.9b for the octet of SU(3), whose hook-number is $D = 1 \cdot 3 \cdot 1 = 3$. Hence the dimension of the SU(3) octet is $24/3 = 8$, whence its name.

Now we can calculate the dimensions of the YT in Eq. (4.124). For SU(2) they are $2 \times 2 = 3 + 1 = 4$. For SU(3) they are $3 \cdot 3 = 3 \cdot 4/(1 \cdot 2) + 3 \cdot 2/(2 \cdot 1) = 6 + 3 = 9$. For the product of the quark times antiquark YT of SU(3) we get

$$[1, 1] \otimes [1] = [2, 1] \oplus [1, 1, 1], \tag{4.126}$$

that is, octet and singlet, which are precisely the meson multiplets considered in the subsection on the eightfold way, the SU(3) flavor symmetry, which suggest mesons are bound states of a quark and an antiquark, $q\bar{q}$ configurations. For the product of three quarks we get

$$\big([1] \otimes [1]\big) \otimes [1] = \big([2] \oplus [1, 1]\big) \otimes [1] = [3] \oplus 2[2, 1] \oplus [1, 1, 1], \tag{4.127}$$

that is, decuplet, octet, and singlet, which are the observed multiplets for the baryons, which suggests they are bound states of three quarks, q^3 configurations.

As we have seen, YT describe the decomposition of a product of SU(n) *irreducible representations into irreducible representations of* SU(n)*, which is called the Clebsch–Gordan series, while the Clebsch–Gordan coefficients considered earlier allow construction of the individual states in this series.*

Exercises

4.4.1 Derive recursion relations for Clebsch–Gordan coefficients. Use them to calculate $C(11J|m_1m_2M)$ for $J = 0, 1, 2$.
Hint. Use the known matrix elements of $J_+ = J_{1+} + J_{2+}$, J_{i+}, and $\mathbf{J}^2 = (\mathbf{J}_1 + \mathbf{J}_2)^2$, etc.

4.4.2 Show that $(Y_l\chi)^J_M = \sum C(l\tfrac{1}{2}J|m_l m_s M)Y_{lm_l}\chi_{m_s}$, where $\chi_{\pm 1/2}$ are the spin up and down eigenfunctions of $\sigma_3 = \sigma_z$, transforms like a spherical tensor of rank J.

4.4.3 When the spin of quarks is taken into account, the SU(3) flavor symmetry is replaced by the SU(6) symmetry. Why? Obtain the Young tableau for the antiquark configuration $\bar{q}$. Then decompose the product $q\bar{q}$. Which SU(3) representations are contained in the nontrivial SU(6) representation for mesons?
Hint. Determine the dimensions of all YT.

4.4.4 For $l = 1$, Eq. (4.107a) becomes

$$Y_1^m(\theta', \varphi') = \sum_{m'=-1}^{1} D^1_{m'm}(\alpha, \beta, \gamma)Y_1^{m'}(\theta, \varphi).$$

Rewrite these spherical harmonics in Cartesian form. Show that the resulting Cartesian coordinate equations are equivalent to the Euler rotation matrix $\mathsf{A}(\alpha, \beta, \gamma)$, Eq. (3.94), rotating the coordinates.

4.4.5 Assuming that $D^j(\alpha, \beta, \gamma)$ is unitary, show that

$$\sum_{m=-l}^{l} Y_l^{m*}(\theta_1, \varphi_1)Y_l^m(\theta_2, \varphi_2)$$

is a scalar quantity (invariant under rotations). This is a spherical tensor analog of a scalar product of vectors.

4.4.6 (a) Show that the α and γ dependence of $\mathsf{D}^j(\alpha, \beta, \gamma)$ may be factored out such that

$$\mathsf{D}^j(\alpha, \beta, \gamma) = \mathsf{A}^j(\alpha)\mathsf{d}^j(\beta)\mathsf{C}^j(\gamma).$$

(b) Show that $\mathsf{A}^j(\alpha)$ and $\mathsf{C}^j(\gamma)$ are diagonal. Find the explicit forms.
(c) Show that $\mathsf{d}^j(\beta) = \mathsf{D}^j(0, \beta, 0)$.

4.4.7 The angular momentum–exponential form of the Euler angle rotation operators is

$$\begin{aligned}\mathsf{R} &= \mathsf{R}_{z''}(\gamma)\mathsf{R}_{y'}(\beta)\mathsf{R}_z(\alpha)\\ &= \exp(-i\gamma J_{z''})\exp(-i\beta J_{y'})\exp(-i\alpha J_z).\end{aligned}$$

Show that in terms of the original axes

$$\mathsf{R} = \exp(i\alpha J_z)\exp(-i\beta J_y)\exp(-i\gamma J_z).$$

Hint. The R operators transform as matrices. The rotation about the y'-axis (second Euler rotation) may be referred to the original y-axis by

$$\exp(-i\beta J_{y'}) = \exp(-i\alpha J_z)\exp(-i\beta J_y)\exp(i\alpha J_z).$$

4.4.8 Using the Wigner–Eckart theorem, prove the decomposition theorem for a spherical vector operator $\langle j'm'|T_{1m}|jm\rangle = \frac{\langle jm'|\mathbf{J}\cdot\mathbf{T}_1|jm\rangle}{j(j+1)}\delta_{jj'}$.

4.4.9 Using the Wigner–Eckart theorem, prove the factorization $\langle j'm'|J_M\mathbf{J}\cdot\mathbf{T_1}|jm\rangle = \langle jm'|J_M|jm\rangle\delta_{j'j}\langle jm|\mathbf{J}\cdot\mathbf{T_1}|jm\rangle$.

4.5 HOMOGENEOUS LORENTZ GROUP

Generalizing the approach to vectors of Section 1.2, in special relativity we demand that our physical laws be covariant[17] under

a. space and time translations,
b. rotations in real, three-dimensional space, and
c. Lorentz transformations.

The demand for covariance under translations is based on the homogeneity of space and time. Covariance under rotations is an assertion of the isotropy of space. The requirement of Lorentz covariance follows from special relativity. All three of these transformations together form the inhomogeneous Lorentz group or the Poincaré group. When we exclude translations, the space rotations and the Lorentz transformations together form a group — the homogeneous Lorentz group.

We first generate a subgroup, the Lorentz transformations in which the relative velocity $\mathbf{v}$ is along the $x = x^1$-axis. The generator may be determined by considering space–time reference frames moving with a relative velocity δv, an infinitesimal.[18] The relations are similar to those for rotations in real space, Sections 1.2, 2.6, and 3.3, except that here the angle of rotation is pure imaginary (compare Section 4.6).

Lorentz transformations are linear not only in the space coordinates x_i but in the time t as well. They originate from Maxwell's equations of electrodynamics, which are invariant under Lorentz transformations, as we shall see later. Lorentz transformations leave the quadratic form $c^2t^2 - x_1^2 - x_2^2 - x_3^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2$ invariant, where $x_0 = ct$. We see this if we switch on a light source at the origin of the coordinate system. At time t light has traveled the distance $ct = \sqrt{\sum x_i^2}$, so $c^2t^2 - x_1^2 - x_2^2 - x_3^2 = 0$. Special relativity requires that in all (inertial) frames that move with velocity $v \le c$ in any direction relative to the x_i-system and have the same origin at time $t = 0$, $c^2t'^2 - x_1'^2 - x_2'^2 - x_3'^2 = 0$ holds also. Four-dimensional space–time with the metric $x \cdot x = x^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2$ is called Minkowski space, with the scalar product of two four-vectors defined as $a \cdot b = a_0b_0 - \mathbf{a}\cdot\mathbf{b}$. Using the metric tensor

$$(g_{\mu\nu}) = \left(g^{\mu\nu}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{4.128}$$

[17] To be covariant means to have the same form in different coordinate systems so that there is no preferred reference system (compare Sections 1.2 and 2.6).

[18] This derivation, with a slightly different metric, appears in an article by J. L. Strecker, *Am. J. Phys.* **35**: 12 (1967).

we can raise and lower the indices of a four-vector, such as the coordinates $x^\mu = (x_0, \mathbf{x})$, so that $x_\mu = g_{\mu\nu}x^\nu = (x_0, -\mathbf{x})$ and $x^\mu g_{\mu\nu}x^\nu = x_0^2 - \mathbf{x}^2$, Einstein's summation convention being understood. For the gradient, $\partial^\mu = (\partial/\partial x_0, -\nabla) = \partial/\partial x_\mu$ and $\partial_\mu = (\partial/\partial x_0, \nabla)$, so $\partial^2 = \partial^\mu\partial_\mu = (\partial/\partial x_0)^2 - \nabla^2$ is a Lorentz scalar, just like the metric $x^2 = x_0^2 - \mathbf{x}^2$.

For $v \ll c$, in the nonrelativistic limit, a Lorentz transformation must be Galilean. Hence, to derive the form of a Lorentz transformation along the x_1-axis, we start with a Galilean transformation for infinitesimal relative velocity δv:

$$x'^1 = x^1 - \delta v t = x^1 - x^0\delta\beta. \tag{4.129}$$

Here, $\beta = v/c$. By symmetry we also write

$$x'^0 = x^0 + a\delta\beta x^1, \tag{4.129$'$}$$

with the parameter a chosen so that $x_0^2 - x_1^2$ is invariant,

$$x_0'^2 - x_1'^2 = x_0^2 - x_1^2. \tag{4.130}$$

Remember, $x^\mu = (x^0, \mathbf{x})$ is the prototype four-dimensional vector in Minkowski space. Thus Eq. (4.130) is simply a statement of the invariance of the square of the magnitude of the "distance" vector under Lorentz transformation in Minkowski space. Here is where the special relativity is brought into our transformation. Squaring and subtracting Eqs. (4.129) and (4.129$'$) and discarding terms of order $(\delta\beta)^2$, we find $a = -1$. Equations (4.129) and (4.129$'$) may be combined as a matrix equation,

$$\begin{pmatrix} x'^0 \\ x'^1 \end{pmatrix} = (1_2 - \delta\beta\sigma_1)\begin{pmatrix} x^0 \\ x^1 \end{pmatrix}; \tag{4.131}$$

σ_1 happens to be the Pauli matrix, σ_1, and the parameter $\delta\beta$ represents an infinitesimal change. Using the same techniques as in Section 4.2, we repeat the transformation N times to develop a finite transformation with the velocity parameter $\rho = N\delta\beta$. Then

$$\begin{pmatrix} x'^0 \\ x'^1 \end{pmatrix} = \left(1_2 - \frac{\rho\sigma_1}{N}\right)^N \begin{pmatrix} x^0 \\ x^1 \end{pmatrix}. \tag{4.132}$$

In the limit as $N \to \infty$,

$$\lim_{N\to\infty}\left(1_2 - \frac{\rho\sigma_1}{N}\right)^N = \exp(-\rho\sigma_1). \tag{4.133}$$

As in Section 4.2, the exponential is interpreted by a Maclaurin expansion,

$$\exp(-\rho\sigma_1) = 1_2 - \rho\sigma_1 + \frac{1}{2!}(\rho\sigma_1)^2 - \frac{1}{3!}(\rho\sigma_1)^3 + \cdots. \tag{4.134}$$

Noting that $(\sigma_1)^2 = 1_2$,

$$\exp(-\rho\sigma_1) = 1_2\cosh\rho - \sigma_1\sinh\rho. \tag{4.135}$$

Hence our finite Lorentz transformation is

$$\begin{pmatrix} x'^0 \\ x'^1 \end{pmatrix} = \begin{pmatrix} \cosh\rho & -\sinh\rho \\ -\sinh\rho & \cosh\rho \end{pmatrix}\begin{pmatrix} x^0 \\ x^1 \end{pmatrix}. \tag{4.136}$$

σ_1 has generated the representations of this pure Lorentz transformation. The quantities $\cosh\rho$ and $\sinh\rho$ may be identified by considering the origin of the primed coordinate system, $x'^1 = 0$, or $x^1 = vt$. Substituting into Eq. (4.136), we have

$$0 = x^1 \cosh\rho - x^0 \sinh\rho. \tag{4.137}$$

With $x^1 = vt$ and $x^0 = ct$,

$$\tanh\rho = \beta = \frac{v}{c}.$$

Note that the **rapidity** $\rho \neq v/c$, except in the limit as $v \to 0$. The rapidity is the additive parameter for pure Lorentz transformations ("boosts") along the same axis that corresponds to angles for rotations about the same axis. Using $1 - \tanh^2\rho = (\cosh^2\rho)^{-1}$,

$$\cosh\rho = \left(1-\beta^2\right)^{-1/2} \equiv \gamma, \qquad \sinh\rho = \beta\gamma. \tag{4.138}$$

The group of Lorentz transformations is not compact, because the limit of a sequence of rapidities going to infinity is no longer an element of the group.

The preceding special case of the velocity parallel to one space axis is easy, but it illustrates the infinitesimal velocity-exponentiation-generator technique. Now, this exact technique may be applied to derive the Lorentz transformation for the relative velocity **v not** parallel to any space axis. The matrices given by Eq. (4.136) for the case of $\mathbf{v} = \hat{\mathbf{x}}v_x$ form a subgroup. The matrices in the general case do not. The product of two Lorentz transformation matrices $\mathsf{L}(\mathbf{v}_1)$ and $\mathsf{L}(\mathbf{v}_2)$ yields a third Lorentz matrix, $\mathsf{L}(\mathbf{v}_3)$, if the two velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ are parallel. The resultant velocity, $\mathbf{v}_3$, is related to $\mathbf{v}_1$ and $\mathbf{v}_2$ by the Einstein velocity addition law, Exercise 4.5.3. If $\mathbf{v}_1$ and $\mathbf{v}_2$ are not parallel, no such simple relation exists. Specifically, consider three reference frames S, S', and S'', with S and S' related by $\mathsf{L}(\mathbf{v}_1)$ and S' and S'' related by $\mathsf{L}(\mathbf{v}_2)$. If the velocity of S'' relative to the original system S is $\mathbf{v}_3$, S'' is not obtained from S by $\mathsf{L}(\mathbf{v}_3) = \mathsf{L}(\mathbf{v}_2)\mathsf{L}(\mathbf{v}_1)$. Rather, we find that

$$\mathsf{L}(\mathbf{v}_3) = \mathsf{R}\mathsf{L}(\mathbf{v}_2)\mathsf{L}(\mathbf{v}_1), \tag{4.139}$$

where R is a 3×3 space rotation matrix embedded in our four-dimensional space–time. With $\mathbf{v}_1$ and $\mathbf{v}_2$ not parallel, the final system, S'', is **rotated** relative to S. This rotation is the origin of the Thomas precession involved in spin-orbit coupling terms in atomic and nuclear physics. Because of its presence, the pure Lorentz transformations $\mathsf{L}(\mathbf{v})$ by themselves do not form a group.

Kinematics and Dynamics in Minkowski Space–Time

We have seen that the propagation of light determines the metric

$$\mathbf{r}^2 - c^2t^2 = 0 = \mathbf{r}'^2 - c^2t'^2,$$

where $x^\mu = (ct, \mathbf{r})$ is the coordinate four-vector. For a particle moving with velocity $\mathbf{v}$, the Lorentz invariant infinitesimal version

$$c\,d\tau \equiv \sqrt{dx^\mu\,dx_\mu} = \sqrt{c^2\,dt^2 - d\mathbf{r}^2} = dt\sqrt{c^2 - \mathbf{v}^2}$$

defines the invariant proper time τ on its track. Because of time dilation in moving frames, a proper-time clock rides with the particle (in its rest frame) and runs at the slowest possible

rate compared to any other inertial frame (of an observer, for example). The four-velocity of the particle can now be defined properly as

$$\frac{dx^\mu}{d\tau} = u^\mu = \left(\frac{c}{\sqrt{c^2-\mathbf{v}^2}}, \frac{\mathbf{v}}{\sqrt{c^2-\mathbf{v}^2}}\right),$$

so $u^2 = 1$, and the four-momentum $p^\mu = cmu^\mu = (\frac{E}{c}, \mathbf{p})$ yields Einstein's famous energy relation

$$E = \frac{mc^2}{\sqrt{1-\mathbf{v}^2/c^2}} = mc^2 + \frac{m}{2}\mathbf{v}^2 \pm \cdots .$$

A consequence of $u^2 = 1$ and its physical significance is that the particle is on its mass shell $p^2 = m^2c^2$.

Now we formulate Newton's equation for a **single particle** of mass m in special relativity as $\frac{dp^\mu}{d\tau} = K^\mu$, with K^μ denoting the force four-vector, so its vector part of the equation coincides with the usual form. For $\mu = 1, 2, 3$ we use $d\tau = dt\sqrt{1-\mathbf{v}^2/c^2}$ and find

$$\frac{1}{\sqrt{1-\mathbf{v}^2/c^2}}\frac{d\mathbf{p}}{dt} = \frac{\mathbf{F}}{\sqrt{1-\mathbf{v}^2/c^2}} = \mathbf{K},$$

determining $\mathbf{K}$ in terms of the usual force $\mathbf{F}$. We need to find K^0. We proceed by analogy with the derivation of energy conservation, multiplying the force equation into the four-velocity

$$mu_\nu \frac{du^\nu}{d\tau} = \frac{m}{2}\frac{du^2}{d\tau} = 0,$$

because $u^2 = 1 = \text{const}$. The other side of Newton's equation yields

$$0 = \frac{1}{c}u \cdot K = \frac{K^0}{\sqrt{1-\mathbf{v}^2/c^2}} - \frac{\mathbf{F}\cdot\mathbf{v}/c}{\sqrt{1-\mathbf{v}^2/c^{2^2}}},$$

so $K^0 = \frac{\mathbf{F}\cdot\mathbf{v}/c}{\sqrt{1-\mathbf{v}^2/c^2}}$ is related to the rate of work done by the force on the particle.

Now we turn to two-body collisions, in which energy–momentum conservation takes the form $p_1 + p_2 = p_3 + p_4$, where p_i^μ are the particle four-momenta. Because the scalar product of any four-vector with itself is an invariant under Lorentz transformations, it is convenient to define the Lorentz invariant energy squared $s = (p_1 + p_2)^2 = P^2$, where P^μ is the total four-momentum, and to use units where the velocity of light $c = 1$. The laboratory system (lab) is defined as the rest frame of the particle with four-momentum $p_2^\mu = (m_2, \mathbf{0})$ and the center of momentum frame (cms) by the total four-momentum $P^\mu = (E_1 + E_2, \mathbf{0})$. When the incident lab energy E_1^L is given, then

$$s = p_1^2 + p_2^2 + 2p_1 \cdot p_2 = m_1^2 + m_2^2 + 2m_2E_1^L$$

is determined. Now, the cms energies of the four particles are obtained from scalar products

$$p_1 \cdot P = E_1(E_1 + E_2) = E_1\sqrt{s},$$

so

$$E_1 = \frac{p_1 \cdot (p_1 + p_2)}{\sqrt{s}} = \frac{m_1^2 + p_1 \cdot p_2}{\sqrt{s}} = \frac{m_1^2 - m_2^2 + s}{2\sqrt{s}},$$

$$E_2 = \frac{p_2 \cdot (p_1 + p_2)}{\sqrt{s}} = \frac{m_2^2 + p_1 \cdot p_2}{\sqrt{s}} = \frac{m_2^2 - m_1^2 + s}{2\sqrt{s}},$$

$$E_3 = \frac{p_3 \cdot (p_3 + p_4)}{\sqrt{s}} = \frac{m_3^2 + p_3 \cdot p_4}{\sqrt{s}} = \frac{m_3^2 - m_4^2 + s}{2\sqrt{s}},$$

$$E_4 = \frac{p_4 \cdot (p_3 + p_4)}{\sqrt{s}} = \frac{m_4^2 + p_3 \cdot p_4}{\sqrt{s}} = \frac{m_4^2 - m_3^2 + s}{2\sqrt{s}},$$

by substituting

$$2p_1 \cdot p_2 = s - m_1^2 - m_2^2, \qquad 2p_3 \cdot p_4 = s - m_3^2 - m_4^2.$$

Thus, all cms energies E_i depend only on the incident energy but not on the scattering angle. For elastic scattering, $m_3 = m_1$, $m_4 = m_2$, so $E_3 = E_1$, $E_4 = E_2$. The Lorentz invariant momentum transfer squared

$$t = (p_1 - p_3)^2 = m_1^2 + m_3^2 - 2p_1 \cdot p_3$$

depends linearly on the cosine of the scattering angle.

Example 4.5.1 KAON DECAY AND PION PHOTOPRODUCTION THRESHOLD

Find the kinetic energies of the muon of mass 106 MeV and massless neutrino into which a K meson of mass 494 MeV decays in its rest frame.

Conservation of energy and momentum gives $m_K = E_\mu + E_\nu = \sqrt{s}$. Applying the relativistic kinematics described previously yields

$$E_\mu = \frac{p_\mu \cdot (p_\mu + p_\nu)}{m_K} = \frac{m_\mu^2 + p_\mu \cdot p_\nu}{m_K},$$

$$E_\nu = \frac{p_\nu \cdot (p_\mu + p_\nu)}{m_K} = \frac{p_\mu \cdot p_\nu}{m_K}.$$

Combining both results we obtain $m_K^2 = m_\mu^2 + 2p_\mu \cdot p_\nu$, so

$$E_\mu = T_\mu + m_\mu = \frac{m_K^2 + m_\mu^2}{2m_K} = 258.4 \text{ MeV},$$

$$E_\nu = T_\nu = \frac{m_K^2 - m_\mu^2}{2m_K} = 235.6 \text{ MeV}.$$

As another example, in the production of a neutral pion by an incident photon according to $\gamma + p \to \pi^0 + p'$ at threshold, the neutral pion and proton are created at rest in the cms. Therefore,

$$s = (p_\gamma + p)^2 = m_p^2 + 2m_p E_\gamma^L = (p_\pi + p')^2 = (m_\pi + m_p)^2,$$

so $E_\gamma^L = m_\pi + \frac{m_\pi^2}{2m_p} = 144.7$ MeV. ■

Exercises

4.5.1 Two Lorentz transformations are carried out in succession: v_1 along the x-axis, then v_2 along the y-axis. Show that the resultant transformation (given by the product of these two successive transformations) **cannot** be put in the form of a single Lorentz transformation.
Note. The discrepancy corresponds to a rotation.

4.5.2 Rederive the Lorentz transformation, working entirely in the real space (x^0, x^1, x^2, x^3) with $x^0 = x_0 = ct$. Show that the Lorentz transformation may be written $\mathsf{L}(\mathbf{v}) = \exp(\rho\sigma)$, with

$$\sigma = \begin{pmatrix} 0 & -\lambda & -\mu & -\nu \\ -\lambda & 0 & 0 & 0 \\ -\mu & 0 & 0 & 0 \\ -\nu & 0 & 0 & 0 \end{pmatrix}$$

and λ, μ, ν the direction cosines of the velocity $\mathbf{v}$.

4.5.3 Using the matrix relation, Eq. (4.136), let the rapidity ρ_1 relate the Lorentz reference frames (x'^0, x'^1) and (x^0, x^1). Let ρ_2 relate (x''^0, x''^1) and (x'^0, x'^1). Finally, let ρ relate (x''^0, x''^1) and (x^0, x^1). From $\rho = \rho_1 + \rho_2$ derive the Einstein velocity addition law

$$v = \frac{v_1 + v_2}{1 + v_1 v_2/c^2}.$$

4.6 Lorentz Covariance of Maxwell's Equations

If a physical law is to hold for all orientations of our (real) coordinates (that is, to be invariant under rotations), the terms of the equation must be covariant under rotations (Sections 1.2 and 2.6). This means that we write the physical laws in the mathematical form scalar = scalar, vector = vector, second-rank tensor = second-rank tensor, and so on. Similarly, if a physical law is to hold for all inertial systems, the terms of the equation must be covariant under Lorentz transformations.

Using Minkowski space ($ct = x^0$; $x = x^1$, $y = x^2$, $z = x^3$), we have a four-dimensional space with the metric $g_{\mu\nu}$ (Eq. (4.128), Section 4.5). The Lorentz transformations are linear in space and time in this four-dimensional real space.[19]

[19] A group theoretic derivation of the Lorentz transformation in Minkowski space appears in Section 4.5. See also H. Goldstein, *Classical Mechanics*. Cambridge, MA: Addison-Wesley (1951), Chapter 6. The metric equation $x_0^2 - \mathbf{x}^2 = 0$, independent of reference frame, leads to the Lorentz transformations.

Here we consider Maxwell's equations,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{4.140a}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \rho \mathbf{v}, \tag{4.140b}$$

$$\nabla \cdot \mathbf{D} = \rho, \tag{4.140c}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{4.140d}$$

and the relations

$$\mathbf{D} = \varepsilon_0 \mathbf{E}, \qquad \mathbf{B} = \mu_0 \mathbf{H}. \tag{4.141}$$

The symbols have their usual meanings as given in Section 1.9. For simplicity we assume vacuum ($\varepsilon = \varepsilon_0$, $\mu = \mu_0$).

We assume that Maxwell's equations hold in all inertial systems; that is, Maxwell's equations are consistent with special relativity. (The covariance of Maxwell's equations under Lorentz transformations was actually shown by Lorentz and Poincaré before Einstein proposed his theory of special relativity.) Our immediate goal is to rewrite Maxwell's equations as tensor equations in Minkowski space. This will make the Lorentz covariance explicit, or manifest.

In terms of scalar, φ, and magnetic vector potentials, $\mathbf{A}$, we may solve[20] Eq. (4.140d) and then (4.140a) by

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A} \\ \mathbf{E} &= -\frac{\partial \mathbf{A}}{\partial t} - \nabla \varphi. \end{aligned} \tag{4.142}$$

Equation (4.142) specifies the curl of $\mathbf{A}$; the divergence of $\mathbf{A}$ is still undefined (compare Section 1.16). We may, and for future convenience we do, impose a further gauge restriction on the vector potential $\mathbf{A}$:

$$\nabla \cdot \mathbf{A} + \varepsilon_0 \mu_0 \frac{\partial \varphi}{\partial t} = 0. \tag{4.143}$$

This is the Lorentz gauge relation. It will serve the purpose of uncoupling the differential equations for $\mathbf{A}$ and φ that follow. The potentials $\mathbf{A}$ and φ are not yet completely fixed. The freedom remaining is the topic of Exercise 4.6.4.

Now we rewrite the Maxwell equations in terms of the potentials $\mathbf{A}$ and φ. From Eqs. (4.140c) for $\nabla \cdot \mathbf{D}$, (4.141) and (4.142),

$$\nabla^2 \varphi + \nabla \cdot \frac{\partial \mathbf{A}}{\partial t} = -\frac{\rho}{\varepsilon_0}, \tag{4.144}$$

whereas Eqs. (4.140b) for $\nabla \times \mathbf{H}$ and (4.142) and Eq. (1.86c) of Chapter 1 yield

$$\frac{\partial^2 \mathbf{A}}{\partial t^2} + \nabla \frac{\partial \varphi}{\partial t} + \frac{1}{\varepsilon_0 \mu_0} \left\{ \nabla \nabla \cdot \mathbf{A} - \nabla^2 \mathbf{A} \right\} = \frac{\rho \mathrm{v}}{\varepsilon_0}. \tag{4.145}$$

[20] Compare Section 1.13, especially Exercise 1.13.10.

Using the Lorentz relation, Eq. (4.143), and the relation $\varepsilon_0\mu_0 = 1/c^2$, we obtain

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\mathbf{A} = -\mu_0\rho\mathbf{v},$$
$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\varphi = -\frac{\rho}{\varepsilon_0}. \tag{4.146}$$

Now, the differential operator (see also Exercise 2.7.3)

$$\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \equiv -\partial^2 \equiv -\partial^\mu\partial_\mu$$

is a four-dimensional Laplacian, usually called the d'Alembertian and also sometimes denoted by $\Box$. It is a scalar by construction (see Exercise 2.7.3).

For convenience we define

$$\begin{aligned} A^1 &\equiv \frac{A_x}{\mu_0 c} = c\varepsilon_0 A_x, \qquad & A^3 &\equiv \frac{A_z}{\mu_0 c} = c\varepsilon_0 A_z, \\ A^2 &\equiv \frac{A_y}{\mu_0 c} = c\varepsilon_0 A_y, \qquad & A_0 &\equiv \varepsilon_0\varphi = A^0. \end{aligned} \tag{4.147}$$

If we further define a four-vector current density

$$\frac{\rho v_x}{c} \equiv j^1, \qquad \frac{\rho v_y}{c} \equiv j^2, \qquad \frac{\rho v_z}{c} \equiv j^3, \qquad \rho \equiv j_0 = j^0, \tag{4.148}$$

then Eq. (4.146) may be written in the form

$$\partial^2 A^\mu = j^\mu. \tag{4.149}$$

The wave equation (4.149) looks like a four-vector equation, but looks do not constitute proof. To prove that it is a four-vector equation, we start by investigating the transformation properties of the generalized current j^μ.

Since an electric charge element de is an invariant quantity, we have

$$de = \rho dx^1\,dx^2\,dx^3, \qquad \text{invariant.} \tag{4.150}$$

We saw in Section 2.9 that the four-dimensional volume element $dx^0\,dx^1\,dx^2\,dx^3$ was also invariant, a pseudoscalar. Comparing this result, Eq. (2.106), with Eq. (4.150), we see that the charge density ρ must transform the same way as dx^0, the zeroth component of a four-dimensional vector dx^λ. We put $\rho = j^0$, with j^0 now established as the zeroth component of a four-vector. The other parts of Eq. (4.148) may be expanded as

$$j^1 = \frac{\rho v_x}{c} = \frac{\rho}{c}\frac{dx^1}{dt} = j^0\frac{dx^1}{dx^0}. \tag{4.151}$$

Since we have just shown that j^0 transforms as dx^0, this means that j^1 transforms as dx^1. With similar results for j^2 and j^3, We have j^λ transforming as dx^λ, proving that j^λ is a four-vector in Minkowski space.

Equation (4.149), which follows directly from Maxwell's equations, Eqs. (4.140), is assumed to hold in all Cartesian systems (all Lorentz frames). Then, by the quotient rule, Section 2.8, A^μ is also a vector and Eq. (4.149) is a legitimate tensor equation.

Now, working backward, Eq. (4.142) may be written

$$\begin{aligned} \varepsilon_0 E_j &= -\frac{\partial A^j}{\partial x^0} - \frac{\partial A^0}{\partial x^j}, \qquad j = 1, 2, 3, \\ \frac{1}{\mu_0 c} B_i &= \frac{\partial A^k}{\partial x^j} - \frac{\partial A^j}{\partial x^k}, \qquad (i, j, k) = \text{cyclic } (1, 2, 3). \end{aligned} \tag{4.152}$$

We define a new tensor,

$$\partial^\mu A^\lambda - \partial^\lambda A^\mu = \frac{\partial A^\lambda}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\lambda} \equiv F^{\mu\lambda} = -F^{\lambda\mu} \qquad (\mu, \lambda = 0, 1, 2, 3),$$

an antisymmetric second-rank tensor, since A^λ is a vector. Written out explicitly,

$$\frac{F_{\mu\lambda}}{\varepsilon_0} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -cB_z & cB_y \\ -E_y & cB_z & 0 & -cB_x \\ -E_z & -cB_y & cB_x & 0 \end{pmatrix}, \qquad \frac{F^{\mu\lambda}}{\varepsilon_0} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -cB_z & cB_y \\ E_y & cB_z & 0 & -cB_x \\ E_z & -cB_y & cB_x & 0 \end{pmatrix}. \tag{4.153}$$

Notice that in our four-dimensional Minkowski space **E** and **B** are no longer vectors but together form a second-rank tensor. With this tensor we may write the two nonhomogeneous Maxwell equations ((4.140b) and (4.140c)) combined as a tensor equation,

$$\frac{\partial F_{\lambda\mu}}{\partial x_\mu} = j_\lambda. \tag{4.154}$$

The left-hand side of Eq. (4.154) is a four-dimensional divergence of a tensor and therefore a vector. This, of course, is equivalent to contracting a third-rank tensor $\partial F^{\lambda\mu}/\partial x_\nu$ (compare Exercises 2.7.1 and 2.7.2). The two homogeneous Maxwell equations — (4.140a) for $\nabla \times \mathbf{E}$ and (4.140d) for $\nabla \cdot \mathbf{B}$ — may be expressed in the tensor form

$$\frac{\partial F_{23}}{\partial x_1} + \frac{\partial F_{31}}{\partial x_2} + \frac{\partial F_{12}}{\partial x_3} = 0 \tag{4.155}$$

for Eq. (4.140d) and three equations of the form

$$-\frac{\partial F_{30}}{\partial x_2} - \frac{\partial F_{02}}{\partial x_3} + \frac{\partial F_{23}}{\partial x_0} = 0 \tag{4.156}$$

for Eq. (4.140a). (A second equation permutes 120, a third permutes 130.) Since

$$\partial^\lambda F^{\mu\nu} = \frac{\partial F^{\mu\nu}}{\partial x_\lambda} \equiv t^{\lambda\mu\nu}$$

is a tensor (of third rank), Eqs. (4.140a) and (4.140d) are given by the tensor equation

$$t^{\lambda\mu\nu} + t^{\nu\lambda\mu} + t^{\mu\nu\lambda} = 0. \tag{4.157}$$

From Eqs. (4.155) and (4.156) you will understand that the indices λ, μ, and ν are supposed to be different. Actually Eq. (4.157) automatically reduces to $0 = 0$ if any two indices coincide. An alternate form of Eq. (4.157) appears in Exercise 4.6.14.

Lorentz Transformation of E and B

The construction of the tensor equations ((4.154) and (4.157)) completes our initial goal of rewriting Maxwell's equations in tensor form.[21] Now we exploit the tensor properties of our four vectors and the tensor $F_{\mu\nu}$.

For the Lorentz transformation corresponding to motion along the $z(x_3)$-axis with velocity v, the "direction cosines" are given by[22]

$$\begin{aligned} x'^0 &= \gamma\left(x^0 - \beta x^3\right) \\ x'^3 &= \gamma\left(x^3 - \beta x^0\right), \end{aligned} \tag{4.158}$$

where

$$\beta = \frac{v}{c}$$

and

$$\gamma = \left(1 - \beta^2\right)^{-1/2}. \tag{4.159}$$

Using the tensor transformation properties, we may calculate the electric and magnetic fields in the moving system in terms of the values in the original reference frame. From Eqs. (2.66), (4.153), and (4.158) we obtain

$$\begin{aligned} E'_x &= \frac{1}{\sqrt{1-\beta^2}}\left(E_x - \frac{v}{c^2}B_y\right), \\ E'_y &= \frac{1}{\sqrt{1-\beta^2}}\left(E_y + \frac{v}{c^2}B_x\right), \\ E'_z &= E_z \end{aligned} \tag{4.160}$$

and

$$\begin{aligned} B'_x &= \frac{1}{\sqrt{1-\beta^2}}\left(B_x + \frac{v}{c^2}E_y\right), \\ B'_y &= \frac{1}{\sqrt{1-\beta^2}}\left(B_y - \frac{v}{c^2}E_x\right), \\ B'_z &= B_z. \end{aligned} \tag{4.161}$$

This coupling of **E** and **B** is to be expected. Consider, for instance, the case of zero electric field in the unprimed system

$$E_x = E_y = E_z = 0.$$

[21] Modern theories of quantum electrodynamics and elementary particles are often written in this "manifestly covariant" form to guarantee consistency with special relativity. Conversely, the insistence on such tensor form has been a useful guide in the construction of these theories.

[22] A group theoretic derivation of the Lorentz transformation appears in Section 4.5. See also Goldstein, *loc. cit.*, Chapter 6.

Clearly, there will be no force on a stationary charged particle. When the particle is in motion with a small velocity v along the z-axis,[23] an observer on the particle sees fields (exerting a force on his charged particle) given by

$$E'_x = -vB_y,$$
$$E'_y = vB_x,$$

where **B** is a magnetic induction field in the unprimed system. These equations may be put in vector form,

$$\mathbf{E}' = \mathbf{v} \times \mathbf{B}$$

or (4.162)

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B},$$

which is usually taken as the operational definition of the magnetic induction **B**.

Electromagnetic Invariants

Finally, the tensor (or vector) properties allow us to construct a multitude of invariant quantities. A more important one is the scalar product of the two four-dimensional vectors or four-vectors A_λ and j_λ. We have

$$\begin{aligned} A^\lambda j_\lambda &= -c\varepsilon_0 A_x \frac{\rho v_x}{c} - c\varepsilon_0 A_y \frac{\rho v_y}{c} - c\varepsilon_0 A_z \frac{\rho v_z}{c} + \varepsilon_0 \varphi \rho \\ &= \varepsilon_0(\rho\varphi - \mathbf{A}\cdot\mathbf{J}), \qquad \text{invariant,} \end{aligned} \tag{4.163}$$

with **A** the usual magnetic vector potential and **J** the ordinary current density. The first term, $\rho\varphi$, is the ordinary static electric coupling, with dimensions of energy per unit volume. Hence our newly constructed scalar invariant is an energy density. The dynamic interaction of field and current is given by the product $\mathbf{A}\cdot\mathbf{J}$. This invariant $A^\lambda j_\lambda$ appears in the electromagnetic Lagrangians of Exercises 17.3.6 and 17.5.1.

Other possible electromagnetic invariants appear in Exercises 4.6.9 and 4.6.11.

The Lorentz group is the symmetry group of electrodynamics, of the electroweak gauge theory, and of the strong interactions described by quantum chromodynamics: It governs special relativity. The metric of Minkowski space–time is Lorentz invariant and expresses the propagation of light; that is, the velocity of light is the same in all inertial frames. Newton's equations of motion are straightforward to extend to special relativity. The kinematics of two-body collisions are important applications of vector algebra in Minkowski space–time.

[23] If the velocity is not small, a relativistic transformation of force is needed.

Exercises

4.6.1 (a) Show that every four-vector in Minkowski space may be decomposed into an ordinary three-space vector and a three-space scalar. Examples: $(ct, \mathbf{r})$, $(\rho, \rho\mathbf{v}/c)$, $(\varepsilon_0\varphi, c\varepsilon_0\mathbf{A})$, $(E/c, \mathbf{p})$, $(\omega/c, \mathbf{k})$.
Hint. Consider a rotation of the three-space coordinates with time fixed.

(b) Show that the converse of (a) is **not** true — every three-vector plus scalar does **not** form a Minkowski four-vector.

4.6.2 (a) Show that

$$\partial^\mu j_\mu = \partial \cdot j = \frac{\partial j_\mu}{\partial x_\mu} = 0.$$

(b) Show how the previous tensor equation may be interpreted as a statement of continuity of charge and current in ordinary three-dimensional space and time.

(c) If this equation is known to hold in all Lorentz reference frames, why can we not conclude that j_μ is a vector?

4.6.3 Write the Lorentz gauge condition (Eq. (4.143)) as a tensor equation in Minkowski space.

4.6.4 A gauge transformation consists of varying the scalar potential φ_1 and the vector potential $\mathbf{A}_1$ according to the relation

$$\varphi_2 = \varphi_1 + \frac{\partial \chi}{\partial t},$$
$$\mathbf{A}_2 = \mathbf{A}_1 - \boldsymbol{\nabla}\chi.$$

The new function χ is required to satisfy the homogeneous wave equation

$$\boldsymbol{\nabla}^2\chi - \frac{1}{c^2}\frac{\partial^2\chi}{\partial t^2} = 0.$$

Show the following:

(a) The Lorentz gauge relation is unchanged.

(b) The new potentials satisfy the same inhomogeneous wave equations as did the original potentials.

(c) The fields **E** and **B** are unaltered.

The invariance of our electromagnetic theory under this transformation is called **gauge invariance**.

4.6.5 A charged particle, charge q, mass m, obeys the Lorentz covariant equation

$$\frac{dp^\mu}{d\tau} = \frac{q}{\varepsilon_0 mc} F^{\mu\nu} p_\nu,$$

where p^ν is the four-momentum vector $(E/c; p^1, p^2, p^3)$, τ is the proper time, $d\tau = dt\sqrt{1 - v^2/c^2}$, a Lorentz scalar. Show that the explicit space–time forms are

$$\frac{dE}{dt} = q\mathbf{v}\cdot\mathbf{E}; \qquad \frac{d\mathbf{p}}{dt} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B}).$$

4.6.6 From the Lorentz transformation matrix elements (Eq. (4.158)) derive the Einstein velocity addition law

$$u' = \frac{u - v}{1 - (uv/c^2)} \qquad \text{or} \qquad u = \frac{u' + v}{1 + (u'v/c^2)},$$

where $u = c\,dx^3/dx^0$ and $u' = c\,dx'^3/dx'^0$.
Hint. If $L_{12}(v)$ is the matrix transforming system 1 into system 2, $L_{23}(u')$ the matrix transforming system 2 into system 3, $L_{13}(u)$ the matrix transforming system 1 directly into system 3, then $L_{13}(u) = L_{23}(u')L_{12}(v)$. From this matrix relation extract the Einstein velocity addition law.

4.6.7 The dual of a four-dimensional second-rank tensor $\mathbf{B}$ may be defined by $\tilde{\mathbf{B}}$, where the elements of the dual tensor are given by

$$\tilde{B}^{ij} = \frac{1}{2!}\varepsilon^{ijkl}B_{kl}.$$

Show that $\tilde{\mathbf{B}}$ transforms as

(a) a second-rank tensor under rotations,
(b) a pseudotensor under inversions.

Note. The tilde here does **not** mean transpose.

4.6.8 Construct $\tilde{\mathbf{F}}$, the dual of $\mathbf{F}$, where $\mathbf{F}$ is the electromagnetic tensor given by Eq. (4.153).

$$\text{ANS. } \tilde{\mathbf{F}}^{\mu\nu} = \varepsilon_0 \begin{pmatrix} 0 & -cB_x & -cB_y & -cB_z \\ cB_x & 0 & E_z & -E_y \\ cB_y & -E_z & 0 & E_x \\ cB_z & E_y & -E_x & 0 \end{pmatrix}.$$

This corresponds to

$$c\mathbf{B} \to -\mathbf{E},$$

$$\mathbf{E} \to c\mathbf{B}.$$

This transformation, sometimes called a **dual transformation**, leaves Maxwell's equations in vacuum ($\rho = 0$) invariant.

4.6.9 Because the quadruple contraction of a fourth-rank pseudotensor and two second-rank tensors $\varepsilon_{\mu\lambda\nu\sigma}F^{\mu\lambda}F^{\nu\sigma}$ is clearly a pseudoscalar, evaluate it.

$$\text{ANS. } -8\varepsilon_0^2 c\mathbf{B}\cdot\mathbf{E}.$$

4.6.10 (a) If an electromagnetic field is purely electric (or purely magnetic) in one particular Lorentz frame, show that $\mathbf{E}$ and $\mathbf{B}$ will be orthogonal in other Lorentz reference systems.
(b) Conversely, if $\mathbf{E}$ and $\mathbf{B}$ are orthogonal in one particular Lorentz frame, there exists a Lorentz reference system in which $\mathbf{E}$ (or $\mathbf{B}$) vanishes. Find that reference system.

4.6.11 Show that $c^2\mathbf{B}^2 - \mathbf{E}^2$ is a Lorentz scalar.

4.6.12 Since (dx^0, dx^1, dx^2, dx^3) is a four-vector, $dx_\mu\, dx^\mu$ is a scalar. Evaluate this scalar for a moving particle in two different coordinate systems: (a) a coordinate system fixed relative to you (lab system), and (b) a coordinate system moving with a moving particle (velocity v relative to you). With the time increment labeled $d\tau$ in the particle system and dt in the lab system, show that

$$d\tau = dt\sqrt{1 - v^2/c^2}.$$

τ is the proper time of the particle, a Lorentz invariant quantity.

4.6.13 Expand the scalar expression

$$-\frac{1}{4\varepsilon_0} F_{\mu\nu} F^{\mu\nu} + \frac{1}{\varepsilon_0} j_\mu A^\mu$$

in terms of the fields and potentials. The resulting expression is the Lagrangian density used in Exercise 17.5.1.

4.6.14 Show that Eq. (4.157) may be written

$$\varepsilon_{\alpha\beta\gamma\delta} \frac{\partial F^{\alpha\beta}}{\partial x_\gamma} = 0.$$

4.7 DISCRETE GROUPS

Here we consider groups with a finite number of elements. In physics, groups usually appear as a set of operations that leave a system unchanged, invariant. This is an expression of symmetry. Indeed, a symmetry may be defined as the invariance of the Hamiltonian of a system under a group of transformations. Symmetry in this sense is important in classical mechanics, but it becomes even more important and more profound in quantum mechanics. In this section we investigate the symmetry properties of sets of objects (atoms in a molecule or crystal). This provides additional illustrations of the group concepts of Section 4.1 and leads directly to dihedral groups. The dihedral groups in turn open up the study of the 32 crystallographic point groups and 230 space groups that are of such importance in crystallography and solid-state physics. It might be noted that it was through the study of crystal symmetries that the concepts of symmetry and group theory entered physics. In physics, the abstract group conditions often take on direct physical meaning in terms of transformations of vectors, spinors, and tensors.

As a simple, but not trivial, example of a finite group, consider the set $1, a, b, c$ that combine according to the group multiplication table[24] (see Fig. 4.10). Clearly, the four conditions of the definition of "group" are satisfied. The elements a, b, c, and 1 are abstract mathematical entities, completely unrestricted except for the multiplication table of Fig. 4.10.

Now, for a specific representation of these group elements, let

$$1 \to 1, \qquad a \to i, \qquad b \to -1, \qquad c \to -i, \tag{4.164}$$

[24]The order of the factors is row–column: $ab = c$ in the indicated previous example.

	1	a	b	c
1	1	a	b	c
a	a	b	c	1
b	b	c	1	a
c	c	1	a	b

FIGURE 4.10 Group multiplication table.

combining by ordinary multiplication. Again, the four group conditions are satisfied, and these four elements form a group. We label this group C_4. Since the multiplication of the group elements is commutative, the group is labeled **commutative**, or **abelian**. Our group is also a **cyclic group**, in that the elements may be written as successive powers of one element, in this case $i^n, n = 0, 1, 2, 3$. Note that in writing out Eq. (4.164) we have selected a specific faithful representation for this group of four objects, C_4.

We recognize that the group elements $1, i, -1, -i$ may be interpreted as successive 90° rotations in the complex plane. Then, from Eq. (3.74), we create the set of four 2×2 matrices (replacing φ by $-\varphi$ in Eq. (3.74) to rotate a vector rather than rotate the coordinates):

$$\mathsf{R}(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix},$$

and for $\varphi = 0, \pi/2, \pi$, and $3\pi/2$ we have

$$\mathsf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \mathsf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$
$$\mathsf{B} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \mathsf{C} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{4.165}$$

This set of four matrices forms a group, with the law of combination being matrix multiplication. Here is a second faithful representation. By matrix multiplication one verifies that this representation is also abelian and cyclic. Clearly, there is a one-to-one correspondence of the two representations

$$1 \leftrightarrow 1 \leftrightarrow \mathsf{1} \qquad a \leftrightarrow i \leftrightarrow \mathsf{A} \qquad b \leftrightarrow -1 \leftrightarrow \mathsf{B} \qquad c \leftrightarrow -i \leftrightarrow \mathsf{C}. \tag{4.166}$$

In the group C_4 the two representations $(1, i, -1, -i)$ and $(\mathsf{1}, \mathsf{A}, \mathsf{B}, \mathsf{C})$ are isomorphic.

In contrast to this, there is no such correspondence between either of these representations of group C_4 and another group of four objects, the vierergruppe (Exercise 3.2.7). The

Table 4.3

	1	V_1	V_2	V_3
1	1	V_1	V_2	V_3
V_1	V_1	1	V_3	V_2
V_2	V_2	V_3	1	V_1
V_3	V_3	V_2	V_1	1

vierergruppe has the multiplication table shown in Table 4.3. Confirming the lack of correspondence between the group represented by $(1, i, -1, -i)$ or the matrices $(\mathsf{1}, \mathsf{A}, \mathsf{B}, \mathsf{C})$ of Eq. (4.165), note that although the vierergruppe is abelian, it is not cyclic. The cyclic group C_4 and the vierergruppe are not isomorphic.

Classes and Character

Consider a group element x transformed into a group element y by a similarity transform with respect to g_i, an element of the group

$$g_i x g_i^{-1} = y. \tag{4.167}$$

The group element y is **conjugate** to x. A **class** is a set of mutually conjugate group elements. In general, this set of elements forming a class does not satisfy the group postulates and is not a group. Indeed, the unit element 1, which is always in a class by itself, is the only class that is also a subgroup. All members of a given class are equivalent, in the sense that any one element is a similarity transform of any other element. Clearly, if a group is abelian, every element is a class by itself. We find that

1. Every element of the original group belongs to one and only one class.
2. The number of elements in a class is a factor of the order of the group.

We get a possible physical interpretation of the concept of class by noting that y is a similarity transform of x. If g_i represents a rotation of the coordinate system, then y is the same operation as x but relative to the new, related coordinates.

In Section 3.3 we saw that a real matrix transforms under rotation of the coordinates by an orthogonal similarity transformation. Depending on the choice of reference frame, essentially the same matrix may take on an infinity of different forms. Likewise, our group representations may be put in an infinity of different forms by using unitary transformations. But each such transformed representation is isomorphic with the original. From Exercise 3.3.9 the trace of each element (each matrix of our representation) is invariant under unitary transformations. Just because it is invariant, the trace (relabeled the **character**) assumes a role of some importance in group theory, particularly in applications to solid-state physics. Clearly, all members of a given class (in a given representation) have the same character. Elements of different classes may have the same character, but elements with different characters cannot be in the same class.

The concept of class is important (1) because of the trace or character and (2) because **the number of nonequivalent irreducible representations of a group is equal to the number of classes.**

Subgroups and Cosets

Frequently a subset of the group elements (including the unit element I) will by itself satisfy the four group requirements and therefore is a group. Such a subset is called a **subgroup**. Every group has two trivial subgroups: the unit element alone and the group itself. The elements 1 and b of the four-element group C_4 discussed earlier form a nontrivial

subgroup. In Section 4.1 we consider SO(3), the (continuous) group of all rotations in ordinary space. The rotations about any single axis form a subgroup of SO(3). Numerous other examples of subgroups appear in the following sections.

Consider a subgroup H with elements h_i and a group element x not in H. Then xh_i and $h_i x$ are not in subgroup H. The sets generated by

$$xh_i, \quad i = 1, 2, \ldots \qquad \text{and} \qquad h_i x, \quad i = 1, 2, \ldots$$

are called **cosets**, respectively the left and right cosets of subgroup H with respect to x. It can be shown (assume the contrary and prove a contradiction) that the coset of a subgroup has the same number of distinct elements as the subgroup. Extending this result we may express the original group G as the sum of H and cosets:

$$G = H + x_1 H + x_2 H + \cdots .$$

Then the order of any subgroup is a divisor of the order of the group. It is this result that makes the concept of coset significant. In the next section the six-element group D_3 (order 6) has subgroups of order 1, 2, and 3. D_3 cannot (and does not) have subgroups of order 4 or 5.

The similarity transform of a subgroup H by a fixed group element x **not** in H, xHx^{-1}, yields a subgroup — Exercise 4.7.8. If this new subgroup is identical with H for all x, that is,

$$xHx^{-1} = H,$$

then H is called an **invariant, normal**, or **self-conjugate subgroup**. Such subgroups are involved in the analysis of multiplets of atomic and nuclear spectra and the particles discussed in Section 4.2. All subgroups of a commutative (abelian) group are automatically invariant.

Two Objects — Twofold Symmetry Axis

Consider first the two-dimensional system of two identical atoms in the xy-plane at (1, 0) and $(-1, 0)$, Fig. 4.11. What rotations[25] can be carried out (keeping both atoms in the xy-plane) that will leave this system invariant? The first candidate is, of course, the unit operator 1. A rotation of π radians about the z-axis completes the list. So we have a rather uninteresting group of two members $(1, -1)$. The z-axis is labeled a twofold symmetry axis — corresponding to the two rotation angles, 0 and π, that leave the system invariant.

Our system becomes more interesting in three dimensions. Now imagine a molecule (or part of a crystal) with atoms of element X at $\pm a$ on the x-axis, atoms of element Y at $\pm b$ on the y-axis, and atoms of element Z at $\pm c$ on the z-axis, as show in Fig. 4.12. Clearly, each axis is now a twofold symmetry axis. Using $R_x(\pi)$ to designate a rotation of π radians about the x-axis, we may

[25] Here we deliberately exclude reflections and inversions. They must be brought in to develop the full set of 32 crystallographic point groups.

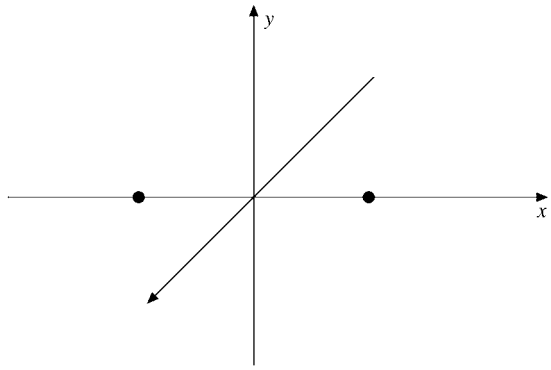

FIGURE 4.11 Diatomic molecules H_2, N_2, O_2, Cl_2.

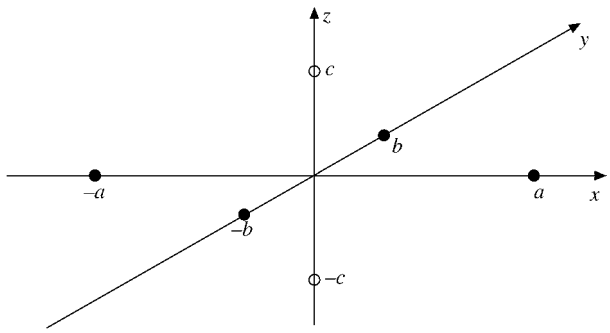

FIGURE 4.12 D_2 symmetry.

set up a matrix representation of the rotations as in Section 3.3:

$$\mathsf{R}_x(\pi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad \mathsf{R}_y(\pi) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$
$$\mathsf{R}_z(\pi) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathsf{1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.168}$$

These four elements $[\mathsf{1}, \mathsf{R}_x(\pi), \mathsf{R}_y(\pi), \mathsf{R}_z(\pi)]$ form an abelian group, with the group multiplication table shown in Table 4.4.

The products shown in Table 4.4 can be obtained in either of two distinct ways: (1) We may analyze the operations themselves—a rotation of π about the x-axis followed by a rotation of π about the y-axis is equivalent to a rotation of π about the z-axis: $\mathsf{R}_y(\pi)\mathsf{R}_x(\pi) = \mathsf{R}_z(\pi)$. (2) Alternatively, once a faithful representation is established, we

Table 4.4

	1	$R_x(\pi)$	$R_y(\pi)$	$R_z(\pi)$
1	1	R_x	R_y	R_x
$R_x(\pi)$	R_x	1	R_z	R_y
$R_y(\pi)$	R_y	R_z	1	R_x
$R_z(\pi)$	R_z	R_y	R_x	1

can obtain the products by matrix multiplication. This is where the power of mathematics is shown — when the system is too complex for a direct physical interpretation.

Comparison with Exercises 3.2.7, 4.7.2, and 4.7.3 shows that this group is the vierergruppe. The matrices of Eq. (4.168) are isomorphic with those of Exercise 3.2.7. Also, they are reducible, being diagonal. The subgroups are $(1, R_x)$, $(1, R_y)$, and $(1, R_z)$. They are invariant. It should be noted that a rotation of π about the y-axis and a rotation of π about the z-axis is equivalent to a rotation of π about the x-axis: $R_z(\pi)R_y(\pi) = R_x(\pi)$. In symmetry terms, if y and z are twofold symmetry axes, x is automatically a twofold symmetry axis.

This symmetry group,[26] the vierergruppe, is often labeled D_2, the D signifying a dihedral group and the subscript 2 signifying a twofold symmetry axis (and no higher symmetry axis).

Three Objects — Threefold Symmetry Axis

Consider now three identical atoms at the vertices of an equilateral triangle, Fig. 4.13. Rotations of the **triangle** of 0, $2\pi/3$, and $4\pi/3$ leave the triangle invariant. In matrix form, we have[27]

$$
\begin{aligned}
1 &= R_z(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
A &= R_z(2\pi/3) = \begin{pmatrix} \cos 2\pi/3 & -\sin 2\pi/3 \\ \sin 2\pi/3 & \cos 2\pi/3 \end{pmatrix} = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix} \\
B &= R_z(4\pi/3) = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}.
\end{aligned} \tag{4.169}
$$

The z-axis is a threefold symmetry axis. $(1, A, B)$ form a cyclic group, a subgroup of the complete six-element group that follows.

In the xy-plane there are three additional axes of symmetry — each atom (vertex) and the geometric center defining an axis. Each of these is a twofold symmetry axis. These rotations may most easily be described within our two-dimensional framework by introducing

[26] A **symmetry** group is a group of symmetry-preserving operations, that is, rotations, reflections, and inversions. A **symmetric** group is the group of permutations of n distinct objects — of order $n!$.

[27] Note that here we are rotating the **triangle** counterclockwise relative to fixed coordinates.

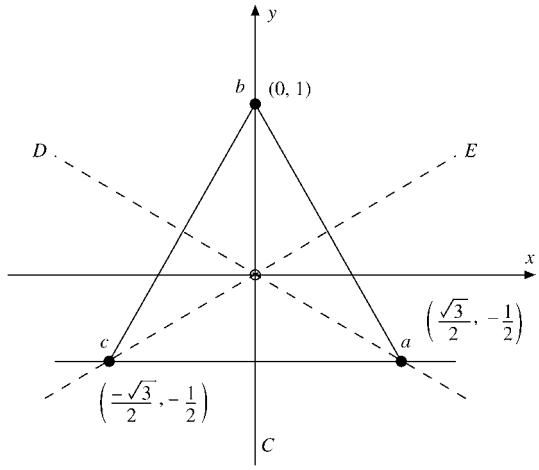

FIGURE 4.13 Symmetry operations on an equilateral triangle.

reflections. The rotation of π about the C- (or y-) axis, which means the interchanging of (structureless) atoms a and c, is just a reflection of the x-axis:

$$\mathsf{C} = \mathsf{R}_C(\pi) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.170}$$

We may replace the rotation about the D-axis by a rotation of $4\pi/3$ (about our z-axis) followed by a reflection of the x-axis ($x \to -x$) (Fig. 4.14):

$$\begin{aligned} \mathsf{D} &= \mathsf{R}_D(\pi) = \mathsf{CB} \\ &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}. \end{aligned} \tag{4.171}$$

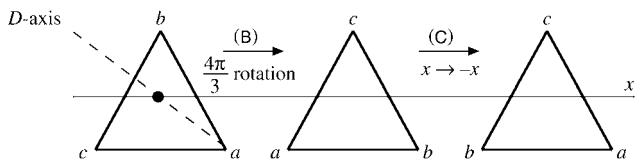

FIGURE 4.14 The triangle on the right is the triangle on the left rotated 180° about the D-axis. $\mathsf{D} = \mathsf{CB}$.

In a similar manner, the rotation of π about the E-axis, interchanging a and b, is replaced by a rotation of $2\pi/3(\mathsf{A})$ and then a reflection[28] of the x-axis:

$$\begin{aligned}\mathsf{E} &= \mathsf{R}_E(\pi) = \mathsf{CA} \\ &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}. \end{aligned} \tag{4.172}$$

The complete group multiplication table is

	1	A	B	C	D	E
1	1	A	B	C	D	E
A	A	B	1	D	E	C
B	B	1	A	E	C	D
C	C	E	D	1	B	A
D	D	C	E	A	1	B
E	E	D	C	B	A	1

Notice that each element of the group appears only once in each row and in each column, as required by the rearrangement theorem, Exercise 4.7.4. Also, from the multiplication table the group is not abelian. We have constructed a six-element group and a 2×2 irreducible matrix representation of it. The only other distinct six-element group is the cyclic group $[1, \mathsf{R}, \mathsf{R}^2, \mathsf{R}^3, \mathsf{R}^4, \mathsf{R}^5]$, with

$$R = e^{2\pi i/6} \qquad \text{or} \qquad \mathsf{R} = e^{-\pi i \sigma_2/3} = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}. \tag{4.173}$$

Our group $[1, \mathsf{A}, \mathsf{B}, \mathsf{C}, \mathsf{D}, \mathsf{E}]$ is labeled D_3 in crystallography, the dihedral group with a threefold axis of symmetry. The three axes (C, D, and E) in the xy-plane automatically become twofold symmetry axes. As a consequence, $(1, \mathsf{C})$, $(1, \mathsf{D})$, and $(1, \mathsf{E})$ all form two-element subgroups. None of these two-element subgroups of D_3 is invariant.

A general and most important result for finite groups of h elements is that

$$\sum_i n_i^2 = h, \tag{4.174}$$

where n_i is the dimension of the matrices of the ith irreducible representation. This equality, sometimes called the **dimensionality theorem**, is very useful in establishing the irreducible representations of a group. Here for D_3 we have $1^2 + 1^2 + 2^2 = 6$ for our three representations. No other irreducible representations of this symmetry group of three objects exist. (The other representations are the identity and ± 1, depending upon whether a reflection was involved.)

[28] Note that, as a consequence of these reflections, $\det(\mathsf{C}) = \det(\mathsf{D}) = \det(\mathsf{E}) = -1$. The rotations A and B, of course, have a determinant of $+1$.

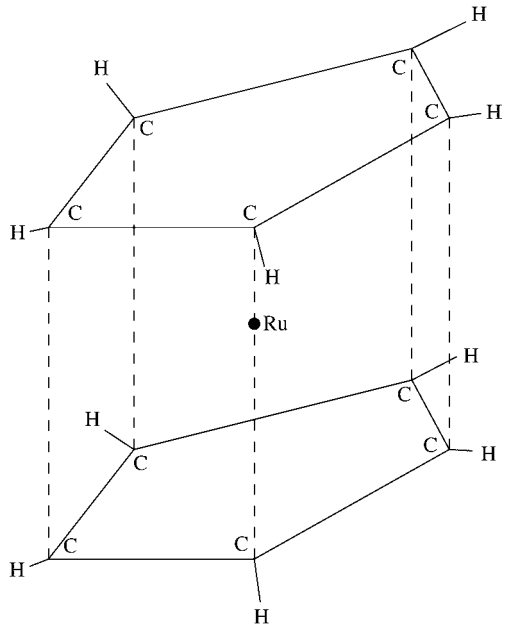

FIGURE 4.15 Ruthenocene.

Dihedral Groups, D_n

A dihedral group D_n with an n-fold symmetry axis implies n axes with angular separation of $2\pi/n$ radians, n is a positive integer, but otherwise unrestricted. If we apply the symmetry arguments to **crystal lattices**, then n is limited to 1, 2, 3, 4, and 6. The requirement of invariance of the crystal lattice under translations in the plane perpendicular to the n-fold axis excludes $n = 5, 7$, and higher values. Try to cover a plane completely with identical regular pentagons and with no overlapping.[29] For individual molecules, this constraint does not exist, although the examples with $n > 6$ are rare. $n = 5$ is a real possibility. As an example, the symmetry group for ruthenocene, $(C_5H_5)_2$Ru, illustrated in Fig. 4.15, is D_5.[30]

Crystallographic Point and Space Groups

The dihedral groups just considered are examples of the crystallographic point groups. A point group is composed of combinations of rotations and reflections (including inversions) that will leave some crystal lattice unchanged. Limiting the operations to rotations and reflections (including inversions) means that one point — the origin — remains **fixed**, hence the term **point group**. Including the cyclic groups, two cubic groups (tetrahedron and octahedron symmetries), and the improper forms (involving reflections), we come to a total of 32 crystallographic point groups.

[29] For D_6 imagine a plane covered with regular hexagons and the axis of rotation through the geometric center of one of them.
[30] Actually the full technical label is D_{5h}, with h indicating invariance under a **reflection** of the fivefold axis.

If, to the rotation and reflection operations that produced the point groups, we add the possibility of translations and still demand that some crystal lattice remain invariant, we come to the space groups. There are 230 distinct space groups, a number that is appalling except, possibly, to specialists in the field. For details (which can cover hundreds of pages) see the Additional Readings.

Exercises

4.7.1 Show that the matrices 1, A, B, and C of Eq. (4.165) are reducible. Reduce them.
Note. This means transforming A and C to diagonal form (by the same unitary transformation).
Hint. A and C are anti-Hermitian. Their eigenvectors will be orthogonal.

4.7.2 Possible operations on a crystal lattice include A_π (rotation by π), m (reflection), and i (inversion). These three operations combine as

$$A_\pi^2 = m^2 = i^2 = 1,$$
$$A_\pi \cdot m = i, \qquad m \cdot i = A_\pi, \qquad \text{and} \qquad i \cdot A_\pi = m.$$

Show that the group $(1, A_\pi, m, i)$ is isomorphic with the vierergruppe.

4.7.3 Four possible operations in the xy-plane are:

1. no change $\begin{cases} x \to x \\ y \to y \end{cases}$
2. inversion $\begin{cases} x \to -x \\ y \to -y \end{cases}$
3. reflection $\begin{cases} x \to -x \\ y \to y \end{cases}$
4. reflection $\begin{cases} x \to x \\ y \to -y. \end{cases}$

(a) Show that these four operations form a group.
(b) Show that this group is isomorphic with the vierergruppe.
(c) Set up a 2×2 matrix representation.

4.7.4 Rearrangement theorem: Given a group of n distinct elements $(I, a, b, c, \ldots, n)$, show that the set of products $(aI, a^2, ab, ac \ldots an)$ reproduces the n distinct elements in a new order.

4.7.5 Using the 2×2 matrix representation of Exercise 3.2.7 for the vierergruppe,

(a) Show that there are four classes, each with one element.

(b) Calculate the character (trace) of each class. Note that two different classes may have the same character.
(c) Show that there are three two-element subgroups. (The unit element by itself always forms a subgroup.)
(d) For any one of the two-element subgroups show that the subgroup and a single coset reproduce the original vierergruppe.

Note that subgroups, classes, and cosets are entirely different.

4.7.6 Using the 2×2 matrix representation, Eq. (4.165), of C_4,

(a) Show that there are four classes, each with one element.
(b) Calculate the character (trace) of each class.
(c) Show that there is one two-element subgroup.
(d) Show that the subgroup and a single coset reproduce the original group.

4.7.7 Prove that the number of distinct elements in a coset of a subgroup is the same as the number of elements in the subgroup.

4.7.8 A subgroup H has elements h_i. Let x be a fixed element of the original group G and **not** a member of H. The transform

$$xh_ix^{-1}, \qquad i = 1, 2, \ldots$$

generates a **conjugate subgroup** xHx^{-1}. Show that this conjugate subgroup satisfies each of the four group postulates and therefore is a group.

4.7.9 (a) A particular group is abelian. A second group is created by replacing g_i by g_i^{-1} for each element in the original group. Show that the two groups are isomorphic. *Note*. This means showing that if $a_ib_i = c_i$, then $a_i^{-1}b_i^{-1} = c_i^{-1}$.
(b) Continuing part (a), if the two groups are isomorphic, show that each must be abelian.

4.7.10 (a) Once you have a matrix representation of any group, a one-dimensional representation can be obtained by taking the determinants of the matrices. Show that the multiplicative relations are preserved in this determinant representation.
(b) Use determinants to obtain a one-dimensional representative of D_3.

4.7.11 Explain how the relation

$$\sum_i n_i^2 = h$$

applies to the vierergruppe ($h = 4$) and to the dihedral group D_3 with $h = 6$.

4.7.12 Show that the subgroup $(1, A, B)$ of D_3 is an invariant subgroup.

4.7.13 The group D_3 may be discussed as a **permutation** group of three objects. Matrix B, for instance, rotates vertex a (originally in location 1) to the position formerly occupied by c

(location 3). Vertex b moves from location 2 to location 1, and so on. As a permutation $(abc) \to (bca)$. In three dimensions

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} b \\ c \\ a \end{pmatrix}.$$

(a) Develop analogous 3×3 representations for the other elements of D_3.
(b) Reduce your 3×3 representation to the 2×2 representation of this section.

(This 3×3 representation must be reducible or Eq. (4.174) would be violated.)
Note. The actual reduction of a reducible representation may be awkward. It is often easier to develop directly a new representation of the required dimension.

4.7.14 (a) The permutation group of four objects P_4 has $4! = 24$ elements. Treating the four elements of the cyclic group C_4 as permutations, set up a 4×4 matrix representation of C_4. C_4 that becomes a subgroup of P_4.
(b) How do you know that this 4×4 matrix representation of C_4 **must** be reducible?
Note. C_4 is abelian and every abelian group of h objects has only h one-dimensional irreducible representations.

4.7.15 (a) The objects $(abcd)$ are permuted to $(dacb)$. Write out a 4×4 matrix representation of this one permutation.
(b) Is the permutation $(abdc) \to (dacb)$ odd or even?
(c) Is this permutation a possible member of the D_4 group? Why or why not?

4.7.16 The elements of the dihedral group D_n may be written in the form

$$\mathsf{S}^{\lambda}\mathsf{R}_z^{\mu}(2\pi/n), \qquad \lambda = 0, 1$$
$$\mu = 0, 1, \ldots, n-1,$$

where $\mathsf{R}_z(2\pi/n)$ represents a rotation of $2\pi/n$ about the n-fold symmetry axis, whereas S represents a rotation of π about an axis through the center of the regular polygon and one of its vertices.
For $\mathsf{S} = \mathsf{E}$ show that this form may describe the matrices A, B, C, and D of D_3.
Note. The elements R_z and S are called the generators of this finite group. Similarly, i is the generator of the group given by Eq. (4.164).

4.7.17 Show that the cyclic group of n objects, C_n, may be represented by $r^m, m = 0, 1, 2, \ldots, n-1$. Here r is a generator given by

$$r = \exp(2\pi i s/n).$$

The parameter s takes on the values $s = 1, 2, 3, \ldots, n$, each value of s yielding a different one-dimensional (irreducible) representation of C_n.

4.7.18 Develop the irreducible 2×2 matrix representation of the group of operations (rotations and reflections) that transform a square into itself. Give the group multiplication table.
Note. This is the symmetry group of a square and also the dihedral group D_4. (See Fig. 4.16.)

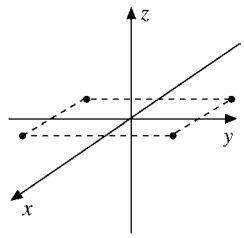

FIGURE 4.16
Square.

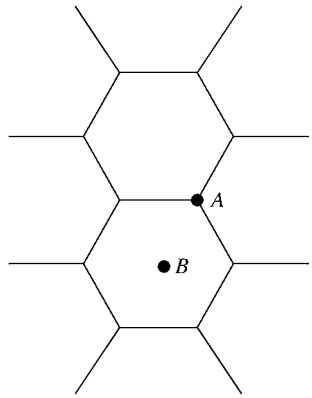

FIGURE 4.17 Hexagon.

4.7.19 The permutation group of four objects contains $4! = 24$ elements. From Exercise 4.7.18, D_4, the symmetry group for a square, has far fewer than 24 elements. Explain the relation between D_4 and the permutation group of four objects.

4.7.20 A plane is covered with regular hexagons, as shown in Fig. 4.17.

(a) Determine the dihedral symmetry of an axis perpendicular to the plane through the common vertex of three hexagons (A). That is, if the axis has n-fold symmetry, show (with careful explanation) what n is. Write out the 2×2 matrix describing the minimum (nonzero) positive rotation of the array of hexagons that is a member of your D_n group.
(b) Repeat part (a) for an axis perpendicular to the plane through the geometric center of one hexagon (B).

4.7.21 In a simple cubic crystal, we might have identical atoms at $r = (la, ma, na)$, with l, m, and n taking on all integral values.

(a) Show that each Cartesian axis is a fourfold symmetry axis.
(b) The cubic group will consist of all operations (rotations, reflections, inversion) that leave the simple cubic crystal invariant. From a consideration of the permutation

y \ x	1	A	−	−
1	1	A	−	−
A	A	1	−	−
−	−	−	1	A
−	−	−	A	1

FIGURE 4.18
Multiplication table.

of the positive and negative coordinate axes, predict how many elements this cubic group will contain.

4.7.22 (a) From the D_3 multiplication table of Fig. 4.18 construct a similarity transform table showing xyx^{-1}, where x and y each range over all six elements of D_3:

(b) Divide the elements of D_3 into classes. Using the 2×2 matrix representation of Eqs. (4.169)–(4.172) note the trace (character) of each class.

4.8 DIFFERENTIAL FORMS

In Chapters 1 and 2 we adopted the view that, in n dimensions, a vector is an n-tuple of real numbers and that its components transform properly under changes of the coordinates. In this section we start from the alternative view, in which a vector is thought of as a directed line segment, an arrow. The point of the idea is this: Although the concept of a vector as a line segment does not generalize to curved space–time (manifolds of differential geometry), except by working in the flat tangent space requiring embedding in auxiliary extra dimensions, Elie Cartan's differential forms are natural in curved space–time and a very powerful tool. Calculus can be based on differential forms, as Edwards has shown by his classic textbook (see the Additional Readings). Cartan's calculus leads to a remarkable unification of concepts and theorems of vector analysis that is worth pursuing. In differential geometry and advanced analysis (on manifolds) the use of differential forms is now widespread.

Cartan's notion of vector is based on the one-to-one correspondence between the linear spaces of displacement vectors and directional differential operators (components of the gradient form a basis). A crucial advantage of the latter is that they can be generalized to curved space–time. Moreover, describing vectors in terms of directional derivatives along curves uniquely specifies the vector at a given point without the need to invoke coordinates. Ultimately, since coordinates are needed to specify points, the Cartan formalism, though an elegant mathematical tool for the efficient derivation of theorems on tensor analysis, has in principle no advantage over the component formalism.

1-Forms

We define dx, dy, dz in three-dimensional Euclidean space as functions assigning to a directed line segment PQ from the point P to the point Q the corresponding change in x, y, z. The symbol dx represents "oriented length of the projection of a curve on the

x-axis," etc. Note that dx, dy, dz can be, but need not be, infinitesimally small, and **they must not be confused with the ordinary differentials** that we associate with integrals and differential quotients. A function of the type

$$A\,dx + B\,dy + C\,dz, \qquad A, B, C \text{ real numbers} \tag{4.175}$$

is defined as a **constant 1-form**.

Example 4.8.1 CONSTANT 1-FORM

For a constant force $\mathbf{F} = (A, B, C)$, the work done along the displacement from $P = (3, 2, 1)$ to $Q = (4, 5, 6)$ is therefore given by

$$W = A(4-3) + B(5-2) + C(6-1) = A + 3B + 5C.$$

If $\mathbf{F}$ is a force field, then its rectangular components $A(x, y, z), B(x, y, z), C(x, y, z)$ will depend on the location and the (nonconstant) 1-**form** $dW = \mathbf{F} \cdot d\mathbf{r}$ corresponds to the concept of work done against the force field $\mathbf{F}(\mathbf{r})$ along $d\mathbf{r}$ on a space curve. A finite amount of work

$$W = \int_C \big[A(x, y, z)\,dx + B(x, y, z)\,dy + C(x, y, z)\,dz\big] \tag{4.176}$$

involves the familiar line integral along an oriented curve C, where the 1-form dW describes the amount of work for small displacements (segments on the path C). In this light, the integrand $f(x)\,dx$ of an integral $\int_a^b f(x)\,dx$ consisting of the function f and of the measure dx as the oriented length is here considered to be a 1-form. The value of the integral is obtained from the ordinary line integral. ■

2-Forms

Consider a unit flow of mass in the z-direction, that is, a flow in the direction of increasing z so that a unit mass crosses a unit square of the xy-plane in unit time. The orientation symbolized by the sequence of points in Fig. 4.19,

$$(0, 0, 0) \to (1, 0, 0) \to (1, 1, 0) \to (0, 1, 0) \to (0, 0, 0),$$

will be called **counterclockwise**, as usual. A unit flow in the z-direction is defined by the function $dx\,dy$[31] assigning to oriented rectangles in space the oriented area of their projections on the xy-plane. Similarly, a unit flow in the x-direction is described by $dy\,dz$ and a unit flow in the y-direction by $dz\,dx$. The reverse order, $dz\,dx$, is dictated by the orientation convention, and $dz\,dx = -dx\,dz$ by definition. This antisymmetry is consistent with the cross product of two vectors representing oriented areas in Euclidean space. This notion generalizes to polygons and curved differentiable surfaces approximated by polygons and volumes.

[31] Many authors denote this wedge product as $dx \wedge dy$ with $dy \wedge dx = -dx \wedge dy$. Note that the product $dx\,dy = dy\,dx$ for ordinary differentials.

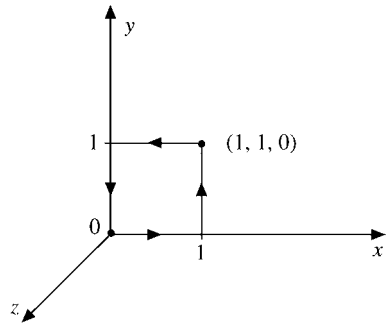

FIGURE 4.19
Counterclockwise-oriented rectangle.

Example 4.8.2 MAGNETIC FLUX ACROSS AN ORIENTED SURFACE

If $\mathbf{B} = (A, B, C)$ is a constant magnetic induction, then the constant **2-form**

$$A\,dy\,dz + B\,dz\,dx + C\,dx\,dy$$

describes the magnetic flux across an oriented rectangle. If **B** is a magnetic induction field varying across a surface S, then the flux

$$\Phi = \int_S \left[B_x(\mathbf{r})\,dy\,dz + B_y(\mathbf{r})\,dz\,dx + B_z(\mathbf{r})\,dx\,dy\right] \tag{4.177}$$

across the oriented surface S involves the familiar (Riemann) integration over approximating small oriented rectangles from which S is pieced together. ■

The definition of $\int \omega$ relies on decomposing $\omega = \sum_i \omega_i$, where the differential forms ω_i are each nonzero only in a small patch of the surface S that covers the surface. Then it can be shown that $\sum_i \int \omega_i$ converges, as the patches become smaller and more numerous, to the limit $\int \omega$, which is independent of these decompositions. For more details and proofs, we refer the reader to Edwards in the Additional Readings.

3-Forms

A 3-form $dx\,dy\,dz$ represents an oriented volume. For example, the determinant of three vectors in Euclidean space changes sign if we reverse the order of two vectors. The determinant measures the oriented volume spanned by the three vectors. In particular, $\int_V \rho(x, y, z)\,dx\,dy\,dz$ represents the total charge inside the volume V if ρ is the charge density. Higher-dimensional differential forms in higher-dimensional spaces are defined similarly and are called k-forms, with $k = 0, 1, 2, \ldots$.

If a 3-form

$$\omega = A(x_1, x_2, x_3)\,dx_1\,dx_2\,dx_3 = A'(x_1', x_2', x_3')\,dx_1'\,dx_2'\,dx_3' \tag{4.178}$$

on a 3-dimensional manifold is expressed in terms of new coordinates, then there is a one-to-one, differentiable map $x_i' = x_i'(x_1, x_2, x_3)$ between these coordinates with Jacobian

$$J = \frac{\partial(x_1', x_2', x_3')}{\partial(x_1, x_2, x_3)} = 1,$$

and $A = A'J = A'$ so that

$$\int_V \omega = \int_V A\, dx_1\, dx_2\, dx_3 = \int_{V'} A'\, dx_1'\, dx_2'\, dx_3'. \tag{4.179}$$

This statement spells out the parameter independence of integrals over differential forms, since parameterizations are essentially arbitrary. The rules governing integration of differential forms are defined on manifolds. These are continuous if we can move continuously (actually we assume them differentiable) from point to point, oriented if the orientation of curves generalizes to surfaces and volumes up to the dimension of the whole manifold. The rules on differential forms are:

- If $\omega = a\omega_1 + a'\omega_1'$, with a, a' real numbers, then $\int_S \omega = a\int_S \omega_1 + a'\int_S \omega_1'$, where S is a compact, oriented, continuous manifold with boundary.
- If the orientation is reversed, then the integral $\int_S \omega$ changes sign.

Exterior Derivative

We now introduce the **exterior derivative** d of a function f, a 0-form:

$$df \equiv \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz = \frac{\partial f}{\partial x_i}dx_i, \tag{4.180}$$

generating a 1-form $\omega_1 = df$, the differential of f (or exterior derivative), the gradient in standard vector analysis. Upon summing over the coordinates, we have used and will continue to use Einstein's summation convention. Applying the exterior derivative d to a 1-form we define

$$d(A\,dx + B\,dy + C\,dz) = dA\,dx + dB\,dy + dC\,dz \tag{4.181}$$

with functions A, B, C. This definition in conjunction with df as just given ties vectors to differential operators $\partial_i = \frac{\partial}{\partial x_i}$. Similarly, we extend d to k-forms. However, applying d twice gives zero, $ddf = 0$, because

$$\begin{aligned} d(df) &= d\frac{\partial f}{\partial x}dx + d\frac{\partial f}{\partial y}dy \\ &= \left(\frac{\partial^2 f}{\partial x^2}dx + \frac{\partial^2 f}{\partial x\,\partial y}dy\right)dx + \left(\frac{\partial^2 f}{\partial y \partial x}dx + \frac{\partial^2 f}{\partial y^2}dy\right)dy \\ &= \left(\frac{\partial^2 f}{\partial y\,\partial x} - \frac{\partial^2 f}{\partial x\,\partial y}\right)dx\,dy = 0. \end{aligned} \tag{4.182}$$

This follows from the fact that in mixed partial derivatives their order does not matter provided all functions are sufficiently differentiable. Similarly we can show $dd\omega_1 = 0$ for a 1-form ω_1, etc.

The rules governing differential forms, with ω_k denoting a k-form, that we have used so far are

- $dx\,dx = 0 = dy\,dy = dz\,dz$, $dx_i^2 = 0$;
- $dx\,dy = -dy\,dx$, $dx_i\,dx_j = -dx_j\,dx_i$, $i \neq j$,
- $dx_1\,dx_2 \cdots dx_k$ is totally antisymmetric in the dx_i, $i = 1, 2, \ldots, k$.
- $df = \frac{\partial f}{\partial x_i} dx_i$;
- $d(\omega_k + \Omega_k) = d\omega_k + d\Omega_k$, linearity;
- $dd\omega_k = 0$.

Now we apply the exterior derivative d to products of differential forms, starting with functions (0-forms). We have

$$d(fg) = \frac{\partial(fg)}{\partial x_i} dx_i = \left(f \frac{\partial g}{\partial x_i} + \frac{\partial f}{\partial x_i} g \right) dx_i = f\,dg + df g. \tag{4.183}$$

If $\omega_1 = \frac{\partial g}{\partial x_i} dx_i$ is a 1-form and f is a function, then

$$\begin{aligned} d(f\omega_1) &= d\left(f \frac{\partial g}{\partial x_i} dx_i \right) = d\left(f \frac{\partial g}{\partial x_i} \right) dx_i \\ &= \frac{\partial \left(f \frac{\partial g}{\partial x_i} \right)}{\partial x_j} dx_j\,dx_i = \left(\frac{\partial f}{\partial x_j} \frac{\partial g}{\partial x_i} + f \frac{\partial^2 g}{\partial x_i\,\partial x_j} \right) dx_j\,dx_i \\ &= df\,\omega_1 + f\,d\omega_1, \end{aligned} \tag{4.184}$$

as expected. But if $\omega_1' = \frac{\partial f}{\partial x_j} dx_j$ is another 1-form, then

$$\begin{aligned} d(\omega_1\omega_1') &= d\left(\frac{\partial g}{\partial x_i} dx_i \frac{\partial f}{\partial x_j} dx_j \right) = d\left(\frac{\partial g}{\partial x_i} \frac{\partial f}{\partial x_j} \right) dx_i\,dx_j \\ &= \frac{\partial \left(\frac{\partial g}{\partial x_i} \frac{\partial f}{\partial x_j} \right)}{\partial x_k} dx_k\,dx_i\,dx_j \\ &= \frac{\partial^2 g}{\partial x_i \partial x_k} dx_k\,dx_i \frac{\partial f}{\partial x_j} dx_j - \frac{\partial g}{\partial x_i} dx_i \frac{\partial^2 f}{\partial x_j\,\partial x_k} dx_k\,dx_j \\ &= d\omega_1\,\omega_1' - \omega_1\,d\omega_1'. \end{aligned} \tag{4.185}$$

This proof is valid for more general 1-forms $\omega = f_i\,dx_i$ with functions f_i. In general, therefore, we define for k-forms:

$$d(\omega_k\omega_k') = (d\omega_k)\omega_k' + (-1)^k \omega_k (d\omega_k'). \tag{4.186}$$

In general, the exterior derivative of a k-form is a $(k+1)$-form.

Example 4.8.3 POTENTIAL ENERGY

As an application in two dimensions (for simplicity), consider the potential $V(\mathbf{r})$, a 0-form, and dV, its exterior derivative. Integrating V along an oriented path C from $\mathbf{r}_1$ to $\mathbf{r}_2$ gives

$$V(\mathbf{r}_2) - V(\mathbf{r}_1) = \int_C dV = \int_C \left(\frac{\partial V}{\partial x} dx + \frac{\partial V}{\partial y} dy \right) = \int_C \nabla V \cdot d\mathbf{r}, \tag{4.187}$$

where the last integral is the standard formula for the potential energy difference that forms part of the energy conservation theorem. The path and parameterization independence are manifest in this special case. ■

Pullbacks

If a linear map $\mathcal{L}_2$ from the uv-plane to the xy-plane has the form

$$x = au + bv + c, \qquad y = eu + fv + g, \tag{4.188}$$

oriented polygons in the uv-plane are mapped onto similar polygons in the xy-plane, provided the determinant $af - be$ of the map $\mathcal{L}_2$ is nonzero. The 2-form

$$dx\,dy = (a\,du + b\,dv)(e\,du + f\,dv) = (af - be)du\,dv \tag{4.189}$$

can be pulled back from the xy- to the uv-plane. That is to say, an integral over a simply connected surface S becomes

$$\int_{\mathcal{L}_2(S)} dx\,dy = (af - be) \int_S du\,dv, \tag{4.190}$$

and $(af - be)\,du\,dv$ is the pullback of $dx\,dy$, opposite to the direction of the map $\mathcal{L}_2$ from the uv-plane to the xy-plane. Of course, the determinant $af - be$ of the map $\mathcal{L}_2$ is simply the Jacobian, generated without effort by the differential forms in Eq. (4.189).

Similarly, a linear map $\mathcal{L}_3$ from the $u_1u_2u_3$-space to the $x_1x_2x_3$-space

$$x_i = a_{ij}u_j + b_i, \qquad i = 1, 2, 3, \tag{4.191}$$

automatically generates its Jacobian from the 3-form

$$\begin{aligned} dx_1\,dx_2\,dx_3 &= \left(\sum_{j=1}^{3} a_{1j}\,du_j \right) \left(\sum_{j=1}^{3} a_{2j}\,du_j \right) \left(\sum_{j=1}^{3} a_{3j}\,du_j \right) \\ &= (a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} \pm \cdots) du_1\,du_2\,du_3 \\ &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} du_1\,du_2\,du_3. \end{aligned} \tag{4.192}$$

Thus, differential forms generate the rules governing determinants.

Given two linear maps in a row, it is straightforward to prove that the pullback under a composed map is the pullback of the pullback. This theorem is the differential-forms analog of matrix multiplication.

Let us now consider a curve C defined by a parameter t in contrast to a curve defined by an equation. For example, the circle $\{(\cos t, \sin t); 0 \le t \le 2\pi\}$ is a parameterization by t, whereas the circle $\{(x, y); x^2 + y^2 = 1\}$ is a definition by an equation. Then the line integral

$$\int_C \big[A(x, y)dx + B(x, y)\, dy\big] = \int_{t_i}^{t_f} \left[A\frac{dx}{dt} + B\frac{dy}{dt}\right] dt \tag{4.193}$$

for continuous functions $A, B, dx/dt, dy/dt$ becomes a one-dimensional integral over the oriented interval $t_i \le t \le t_f$. Clearly, the 1-form $[A\frac{dx}{dt} + B\frac{dy}{dt}]\,dt$ on the t-line is obtained from the 1-form $A\,dx + B\,dy$ on the xy-plane via the map $x = x(t), y = y(t)$ from the t-line to the curve C in the xy-plane. The 1-form $[A\frac{dx}{dt} + B\frac{dy}{dt}]\,dt$ is called the pullback of the 1-form $A\,dx + B\,dy$ under the map $x = x(t)$, $y = y(t)$. Using pullbacks we can show that integrals over 1-forms are independent of the parameterization of the path.

In this sense, the differential quotient $\frac{dy}{dx}$ can be considered as the coefficient of dx in the pullback of dy under the function $y = f(x)$, or $dy = f'(x)\,dx$. This concept of pullback readily generalizes to maps in three or more dimensions and to k-forms with $k > 1$. In particular, the chain rule can be seen to be a pullback: If

$$\begin{aligned} y_i &= f_i(x_1, x_2, \ldots, x_n), \qquad i = 1, 2, \ldots, l \quad \text{and} \\ z_j &= g_j(y_1, y_2, \ldots, y_l), \qquad j = 1, 2, \ldots, m \end{aligned} \tag{4.194}$$

are differentiable maps from $\mathbb{R}^n \to \mathbb{R}^l$ and $\mathbb{R}^l \to \mathbb{R}^m$, then the composed map $\mathbb{R}^n \to \mathbb{R}^m$ is differentiable and the pullback of any k-form under the composed map is equal to the pullback of the pullback. This theorem is useful for establishing that integrals of k-forms are parameter independent.

Similarly, we define the differential df as the pullback of the 1-form dz under the function $z = f(x, y)$:

$$dz = df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy. \tag{4.195}$$

Example 4.8.4 STOKES' THEOREM

As another application let us first sketch the standard derivation of the simplest version of Stokes' theorem for a rectangle $S = [a \le x \le b, c \le y \le d]$ oriented counterclockwise, with ∂S its boundary

$$\begin{aligned} \int_{\partial S} (A\,dx + B\,dy) &= \int_a^b A(x, c)\,dx + \int_c^d B(b, y)\,dy + \int_b^a A(x, d)\,dx + \int_d^c B(a, y)\,dy \\ &= \int_c^d \big[B(b, y) - B(a, y)\big]\,dy - \int_a^b \big[A(x, d) - A(x, c)\big]\,dx \\ &= \int_c^d \int_a^b \frac{\partial B}{\partial x}dx\,dy - \int_a^b \int_c^d \frac{\partial A}{\partial y}dy\,dx \\ &= \int_S \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right) dx\,dy, \end{aligned} \tag{4.196}$$

which holds for any simply connected surface S that can be pieced together by rectangles.

Now we demonstrate the use of differential forms to obtain the same theorem (again in two dimensions for simplicity):

$$\begin{aligned} d(A\,dx + B\,dy) &= dA\,dx + dB\,dy \\ &= \left(\frac{\partial A}{\partial x}dx + \frac{\partial A}{\partial y}dy\right)dx + \left(\frac{\partial B}{\partial x}dx + \frac{\partial B}{\partial y}dy\right)dy = \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right)dx\,dy, \end{aligned} \tag{4.197}$$

using the rules highlighted earlier. Integrating over a surface S and its boundary ∂S, respectively, we obtain

$$\int_{\partial S}(A\,dx + B\,dy) = \int_S d(A\,dx + B\,dy) = \int_S \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right)dx\,dy. \tag{4.198}$$

Here contributions to the left-hand integral from inner boundaries cancel as usual because they are oriented in opposite directions on adjacent rectangles. For each oriented inner rectangle that makes up the simply connected surface S we have used,

$$\int_R dd x = \int_{\partial R} dx = 0. \tag{4.199}$$

Note that the exterior derivative automatically generates the z component of the curl.

In three dimensions, Stokes' theorem derives from the differential-form identity involving the vector potential $\mathbf{A}$ and magnetic induction $\mathbf{B} = \nabla \times \mathbf{A}$,

$$\begin{aligned} d(A_x\,dx + A_y\,dy + A_z\,dz) &= dA_x\,dx + dA_y\,dy + dA_z\,dz \\ &= \left(\frac{\partial A_x}{\partial x}dx + \frac{\partial A_x}{\partial y}dy + \frac{\partial A_x}{\partial z}dz\right)dx + \cdots \\ &= \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)dy\,dz + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right)dz\,dx + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)dx\,dy, \end{aligned} \tag{4.200}$$

generating all components of the curl in three-dimensional space. This identity is integrated over each oriented rectangle that makes up the simply connected surface S (which has no holes, that is, where every curve contracts to a point of the surface) and then is summed over all adjacent rectangles to yield the magnetic flux across S,

$$\begin{aligned} \Phi &= \int_S [B_x\,dy\,dz + B_y\,dz\,dx + B_z\,dx\,dy] \\ &= \int_{\partial S} [A_x\,dx + A_y\,dy + A_z\,dz], \end{aligned} \tag{4.201}$$

or, in the standard notation of vector analysis (Stokes' theorem, Chapter 1),

$$\int_S \mathbf{B}\cdot d\mathbf{a} = \int_S (\nabla \times \mathbf{A})\cdot d\mathbf{a} = \int_{\partial S} \mathbf{A}\cdot d\mathbf{r}. \tag{4.202}$$

■

Example 4.8.5 GAUSS' THEOREM

Consider Gauss' law, Section 1.14. We integrate the electric density $\rho = \frac{1}{\varepsilon_0}\nabla \cdot \mathbf{E}$ over the volume of a single parallelepiped $V = [a \le x \le b, c \le y \le d, e \le z \le f]$ oriented by $dx\,dy\,dz$ (right-handed), the side $x = b$ of V is oriented by $dy\,dz$ (counterclockwise, as seen from $x > b$), and so on. Using

$$E_x(b, y, z) - E_x(a, y, z) = \int_a^b \frac{\partial E_x}{\partial x} dx, \tag{4.203}$$

we have, in the notation of differential forms, summing over all adjacent parallelepipeds that make up the volume $\mathcal{V}$,

$$\int_{\partial \mathcal{V}} E_x\, dy\, dz = \int_{\mathcal{V}} \frac{\partial E_x}{\partial x} dx\, dy\, dz. \tag{4.204}$$

Integrating the electric flux (2-form) identity

$$\begin{aligned} d(E_x\, dy\, dz + E_y\, dz\, dx + E_z\, dx\, dy) &= dE_x\, dy\, dz + dE_y\, dz\, dx + dE_z\, dx\, dy \\ &= \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right) dx\, dy\, dz \end{aligned} \tag{4.205}$$

across the simply connected surface $\partial \mathcal{V}$ we have Gauss' theorem,

$$\int_{\partial \mathcal{V}} (E_x\, dy\, dz + E_y\, dz\, dx + E_z\, dx\, dy) = \int_{\mathcal{V}} \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right) dx\, dy\, dz, \tag{4.206}$$

or, in standard notation of vector analysis,

$$\int_{\partial \mathcal{V}} \mathbf{E} \cdot d\mathbf{a} = \int_{\mathcal{V}} \nabla \cdot \mathbf{E}\, d^3r = \frac{q}{\varepsilon_0}. \tag{4.207}$$

■

These examples are different cases of a single theorem on differential forms. To explain why, let us begin with some terminology, a preliminary **definition of a differentiable manifold M**: It is a collection of points (m-tuples of real numbers) that are smoothly (that is, differentiably) connected with each other so that the neighborhood of each point looks like a simply connected piece of an m-dimensional Cartesian space "close enough" around the point and containing it. Here, m, which stays constant from point to point, is called the dimension of the manifold. Examples are the m-dimensional Euclidean space $\mathbb{R}^m$ and the m-dimensional sphere

$$S_m = \left[\left(x^1, \ldots, x^{m+1}\right); \qquad \sum_{i=1}^{m+1} \left(x^i\right)^2 = 1 \right].$$

Any surface with sharp edges, corners, or kinks is not a manifold in our sense, that is, is not differentiable. In differential geometry, all movements, such as translation and parallel displacement, are local, that is, are defined infinitesimally. If we apply the exterior derivative d to a function $f(x^1, \ldots, x^m)$ on **M**, we generate basic 1-forms:

$$df = \frac{\partial f}{\partial x_i} dx^i, \tag{4.208}$$

where $x^i(P)$ are coordinate functions. As before we have $d(df) = 0$ because

$$d(df) = d\left(\frac{\partial f}{\partial x^i}\right) dx^i = \frac{\partial^2 f}{\partial x^j\, \partial x^i} dx^j\, dx^i$$
$$= \sum_{j<i} \left(\frac{\partial^2 f}{\partial x^j\, \partial x^i} - \frac{\partial^2 f}{\partial x^i\, \partial x^j}\right) dx^j\, dx^i = 0 \tag{4.209}$$

because the order of derivatives does not matter. Any 1-form is a linear combination $\omega = \sum_i \omega_i\, dx^i$ with functions ω_i.

Generalized Stokes' Theorem on Differential Forms

Let ω be a continuous $(k-1)$-form in $x_1x_2\cdots x_n$-space defined everywhere on a compact, oriented, differentiable k-dimensional manifold S with boundary ∂S in $x_1x_2\cdots x_n$-space. Then

$$\int_{\partial S} \omega = \int_S d\omega. \tag{4.210}$$

Here

$$d\omega = d(A\, dx_1\, dx_2 \cdots dx_{k-1} + \cdots) = dA\, dx_1\, dx_2 \cdots dx_{k-1} + \cdots. \tag{4.211}$$

The potential energy in Example 4.8.3 given this theorem for the potential $\omega = V$, a 0-form; Stokes' theorem in Example 4.8.4 is this theorem for the vector potential 1-form $\sum_i A_i\, dx_i$ (for Euclidean spaces $dx^i = dx_i$); and Gauss' theorem in Example 4.8.5 is Stokes' theorem for the electric flux 2-form in three-dimensional Euclidean space.

The method of integration by parts can be generalized to differential forms using Eq. (4.186):

$$\int_S d\omega_1\, \omega_2 = \int_{\partial S} \omega_1\omega_2 - (-1)^{k_1} \int_S \omega_1\, d\omega_2. \tag{4.212}$$

This is proved by integrating the identity

$$d(\omega_1\omega_2) = d\omega_1\, \omega_2 + (-1)^{k_1}\omega_1\, d\omega_2, \tag{4.213}$$

with the integrated term $\int_S d(\omega_1\omega_2) = \int_{\partial S} \omega_1\omega_2$.

Our next goal is to cast Sections 2.10 and 2.11 in the language of differential forms. So far we have worked in two- or three-dimensional Euclidean space.

Example 4.8.6 RIEMANN MANIFOLD

Let us look at the curved Riemann space–time of Sections 2.10–2.11 and reformulate some of this tensor analysis in curved spaces in the language of differential forms. Recall that dishinguishing between upper and lower indices is important here. The metric g_{ij} in Eq. (2.123) can be written in terms of tangent vectors, Eq. (2.114), as follows:

$$g_{ij} = \frac{\partial x^l}{\partial q^i}\frac{\partial x_l}{\partial q^j}, \tag{4.214}$$

where the sum over the index l denotes the inner product of the tangent vectors. (Here we continue to use Einstein's summation convention over repeated indices. As before, the metric tensor is used to raise and lower indices.) The key concept of connection involves the Christoffel symbols, which we address first. The exterior derivative of a tangent vector can be expanded in terms of the basis of tangent vectors (compare Eq. (2.131a)),

$$d\left(\frac{\partial x^l}{\partial q^i}\right) = \Gamma^k{}_{ij}\frac{\partial x^l}{\partial q^k}dq^j, \tag{4.215}$$

thus introducing the Christoffel symbols of the second kind. Applying d to Eq. (4.214) we obtain

$$\begin{aligned} dg_{ij} &= \frac{\partial g_{ij}}{\partial q^m}dq^m = d\left(\frac{\partial x^l}{\partial q^i}\right)\frac{\partial x_l}{\partial q^j} + \frac{\partial x^l}{\partial q^i}d\left(\frac{\partial x_l}{\partial q^j}\right) \\ &= \left(\Gamma^k{}_{im}\frac{\partial x_l}{\partial q^k}\frac{\partial x^l}{\partial q^j} + \Gamma^k{}_{jm}\frac{\partial x^l}{\partial q^i}\frac{\partial x_l}{\partial q^k}\right)dq^m = \left(\Gamma^k{}_{im}g_{kj} + \Gamma^k{}_{jm}g_{ik}\right)dq^m. \end{aligned} \tag{4.216}$$

Comparing the coefficients of dq^m yields

$$\frac{\partial g_{ij}}{\partial q^m} = \Gamma^k{}_{im}g_{kj} + \Gamma^k{}_{jm}g_{ik}. \tag{4.217}$$

Using the Christoffel symbol of the first kind,

$$[ij, m] = g_{km}\Gamma^k{}_{ij}, \tag{4.218}$$

we can rewrite Eq. (4.217) as

$$\frac{\partial g_{ij}}{\partial q^m} = [im, j] + [jm, i], \tag{4.219}$$

which corresponds to Eq. (2.136) and implies Eq. (2.137). We check that

$$[ij, m] = \frac{1}{2}\left(\frac{\partial g_{im}}{\partial q^j} + \frac{\partial g_{jm}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^m}\right) \tag{4.220}$$

is the unique solution of Eq. (4.219) and that

$$\Gamma^k{}_{ij} = g^{mk}[ij, m] = \frac{1}{2}g^{mk}\left(\frac{\partial g_{im}}{\partial q^j} + \frac{\partial g_{jm}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^m}\right) \tag{4.221}$$

follows. ■

Hodge * Operator

The differentials dx^i, $i = 1, 2, \ldots, m$, form a basis of a vector space that is (seen to be) dual to the derivatives $\partial_i = \frac{\partial}{\partial x^i}$; they are basic 1-forms. For example, the vector space $\mathcal{V} = \{(a_1, a_2, a_3)\}$ is dual to the vector space of planes (linear functions f) in three-dimensional Euclidean space $\mathcal{V}^* = \{f \equiv a_1x_1 + a_2x_2 + a_3x_3 - d = 0\}$. The gradient

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}\right) = (a_1, a_2, a_3) \tag{4.222}$$

provides a one-to-one, differentiable map from $\mathcal{V}^*$ to $\mathcal{V}$. Such dual relationships are generalized by the Hodge * operator, based on the Levi-Civita symbol of Section 2.9.

Let the unit vectors $\hat{\mathbf{x}}_i$ be an oriented orthonormal basis of three-dimensional Euclidean space. Then the Hodge * of scalars is defined by the basis element

$$*1 \equiv \frac{1}{3!}\varepsilon^{ijk}\hat{\mathbf{x}}_i\hat{\mathbf{x}}_j\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_1\hat{\mathbf{x}}_2\hat{\mathbf{x}}_3, \tag{4.223}$$

which corresponds to $(\hat{\mathbf{x}}_1 \times \hat{\mathbf{x}}_2) \cdot \hat{\mathbf{x}}_3$ in standard vector notation. Here $\hat{\mathbf{x}}_i\hat{\mathbf{x}}_j\hat{\mathbf{x}}_k$ is the totally antisymmetric exterior product of the unit vectors that corresponds to $(\hat{\mathbf{x}}_i \times \hat{\mathbf{x}}_j) \cdot \hat{\mathbf{x}}_k$ in standard vector notation. For vectors, * is defined for the basis of unit vectors as

$$*\hat{\mathbf{x}}_i \equiv \frac{1}{2!}\varepsilon_i{}^{jk}\hat{\mathbf{x}}_j\hat{\mathbf{x}}_k. \tag{4.224}$$

In particular,

$$*\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_2\hat{\mathbf{x}}_3, \qquad *\hat{\mathbf{x}}_2 = \hat{\mathbf{x}}_3\hat{\mathbf{x}}_1, \qquad *\hat{\mathbf{x}}_3 = \hat{\mathbf{x}}_1\hat{\mathbf{x}}_2. \tag{4.225}$$

For oriented areas, * is defined on basis area elements as

$$*(\hat{\mathbf{x}}_i\hat{\mathbf{x}}_j) \equiv \varepsilon^k{}_{ij}\hat{\mathbf{x}}_k, \tag{4.226}$$

so

$$\begin{aligned} &*(\hat{\mathbf{x}}_1\hat{\mathbf{x}}_2) = \varepsilon^3{}_{12}\hat{\mathbf{x}}_3 = \hat{\mathbf{x}}_3, \qquad *(\hat{\mathbf{x}}_1\hat{\mathbf{x}}_3) = \varepsilon^2{}_{13}\hat{\mathbf{x}}_2 = -\hat{\mathbf{x}}_2, \\ &*(\hat{\mathbf{x}}_2\hat{\mathbf{x}}_3) = \varepsilon^1{}_{23}\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_1. \end{aligned} \tag{4.227}$$

For volumes, * is defined as

$$*(\hat{\mathbf{x}}_1\hat{\mathbf{x}}_2\hat{\mathbf{x}}_3) \equiv \varepsilon_{123} = 1. \tag{4.228}$$

Example 4.8.7 CROSS PRODUCT OF VECTORS

The exterior product of two vectors

$$\mathbf{a} = \sum_{i=1}^{3} a^i\hat{\mathbf{x}}_i, \qquad \mathbf{b} = \sum_{i=1}^{3} b^i\hat{\mathbf{x}}_i \tag{4.229}$$

is given by

$$\mathbf{ab} = \left(\sum_{i=1}^{3} a^i\hat{\mathbf{x}}_i\right)\left(\sum_{j=1}^{3} b^i\hat{\mathbf{x}}_j\right) = \sum_{i<j}(a^ib^j - a^jb^i)\hat{\mathbf{x}}_i\hat{\mathbf{x}}_j, \tag{4.230}$$

whereas Eq. (4.224) implies that

$$*(\mathbf{ab}) = \mathbf{a} \times \mathbf{b}. \tag{4.231}$$

■

Next, let us analyze Sections 2.1–2.2 on curvilinear coordinates in the language of differential forms.

Example 4.8.8 LAPLACIAN IN ORTHOGONAL COORDINATES

Consider orthogonal coordinates where the metric (Eq. (2.5)) leads to length elements

$$ds_i = h_i\, dq_i, \qquad \text{not summed.} \tag{4.232}$$

Here the dq_i are ordinary differentials. The 1-forms associated with the directions $\hat{\mathbf{q}}_i$ are

$$\varepsilon^i = h_i dq_i, \qquad \text{not summed.} \tag{4.233}$$

Then the gradient is defined by the 1-form

$$df = \frac{\partial f}{\partial q_i} dq_i = \left(\frac{1}{h_i}\frac{\partial f}{\partial q_i}\right)\varepsilon^i. \tag{4.234}$$

We apply the hodge star operator to df, generating the 2-form

$$\begin{aligned} *df &= \left(\frac{1}{h_i}\frac{\partial f}{\partial q_i}\right) * \varepsilon^i = \left(\frac{1}{h_1}\frac{\partial f}{\partial q_1}\right)\varepsilon^2\varepsilon^3 + \left(\frac{1}{h_2}\frac{\partial f}{\partial q_2}\right)\varepsilon^3\varepsilon^1 + \left(\frac{1}{h_3}\frac{\partial f}{\partial q_3}\right)\varepsilon^1\varepsilon^2 \\ &= \left(\frac{h_2 h_3}{h_1}\frac{\partial f}{\partial q_1}\right) dq_2\, dq_3 + \left(\frac{h_1 h_3}{h_2}\frac{\partial f}{\partial q_2}\right) dq_3\, dq_1 + \left(\frac{h_1 h_2}{h_3}\frac{\partial f}{\partial q_3}\right) dq_1\, dq_2. \end{aligned} \tag{4.235}$$

Applying another exterior derivative d, we get the Laplacian

$$\begin{aligned} d(*df) &= \frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial f}{\partial q_1}\right) dq_1\, dq_2\, dq_3 + \frac{\partial}{\partial q_2}\left(\frac{h_1 h_3}{h_2}\frac{\partial f}{\partial q_2}\right) dq_2\, dq_1\, dq_2\, dq_3 \\ &\quad + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial f}{\partial q_3}\right) dq_3\, dq_1\, dq_2\, dq_3 \\ &= \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial f}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_1 h_3}{h_2}\frac{\partial f}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial f}{\partial q_3}\right)\right] \\ &\quad \cdot \varepsilon^1\varepsilon^2\varepsilon^3 = \nabla^2 f\, dq_1\, dq_2\, dq_3. \end{aligned} \tag{4.236}$$

Dividing by the volume element gives Eq. (2.22). Recall that the volume elements $dx\, dy\, dz$ and $\varepsilon^1\varepsilon^2\varepsilon^3$ must be equal because ε^i and dx, dy, dz are orthonormal 1-forms and the map from the xyz to the q_i coordinates is one-to-one. ■

Example 4.8.9 MAXWELL'S EQUATIONS

We now work in four-dimensional Minkowski space, the homogeneous, flat space–time of special relativity, to discuss classical electrodynamics in terms of differential forms. We start by introducing the electromagnetic field 2-form (field tensor in standard relativistic notation):

$$\begin{aligned} F &= -E_x\, dt\, dx - E_y\, dt\, dy - E_z\, dt\, dz + B_x\, dy\, dz + B_y\, dz\, dx + B_z\, dx\, dy \\ &= \frac{1}{2} F_{\mu\nu}\, dx^\mu\, dx^\nu, \end{aligned} \tag{4.237}$$

which contains the electric 1-form $E = E_x\,dx + E_y\,dy + E_z\,dz$ and the magnetic flux 2-form. Here, terms with 1-forms in opposite order have been combined. (For Eq. (4.237) to be valid, the magnetic induction is in units of c; that is, $B_i \to cB_i$, with c the velocity of light; or we work in units where $c = 1$. Also, F is in units of $1/\varepsilon_0$, the dielectric constant of the vacuum. Moreover, the vector potential is defined as $A_0 = \varepsilon_0\phi$, with the nonstatic electric potential ϕ and $A^1 = \frac{A_x}{\mu_0 c}, \ldots$; see Section 4.6 for more details.) The field 2-form F encompasses Faraday's induction law that a moving charge is acted on by magnetic forces.

Applying the exterior derivative d generates Maxwell's homogeneous equations automatically from F:

$$\begin{aligned} dF &= -\left(\frac{\partial E_x}{\partial y}dy + \frac{\partial E_x}{\partial z}dz\right)dt\,dx - \left(\frac{\partial E_y}{\partial x}dx + \frac{\partial E_y}{\partial z}dz\right)dt\,dy \\ &\quad - \left(\frac{\partial E_z}{\partial x}dx + \frac{\partial E_z}{\partial y}dy\right)dt\,dz + \left(\frac{\partial B_x}{\partial x}dx + \frac{\partial B_x}{\partial t}dt\right)dy\,dz \\ &\quad + \left(\frac{\partial B_y}{\partial t}dt + \frac{\partial B_y}{\partial y}dy\right)dz\,dx + \left(\frac{\partial B_z}{\partial t}dt + \frac{\partial B_z}{\partial z}dz\right)dx\,dy \\ &= \left(-\frac{\partial E_x}{\partial y} + \frac{\partial E_y}{\partial x} + \frac{\partial B_z}{\partial t}\right)dt\,dx\,dy + \left(-\frac{\partial E_x}{\partial z} + \frac{\partial E_z}{\partial x} - \frac{\partial B_y}{\partial t}\right)dt\,dx\,dz \\ &\quad + \left(-\frac{\partial E_y}{\partial z} + \frac{\partial E_z}{\partial y} + \frac{\partial B_x}{\partial t}\right)dt\,dy\,dz = 0 \end{aligned} \tag{4.238}$$

which, in standard notation of vector analysis, takes the familiar vector form of Maxwell's homogeneous equations,

$$\boldsymbol{\nabla} \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0. \tag{4.239}$$

Since $dF = 0$, that is, there is no driving term so that F is closed, there must be a 1-form $\omega = A_\mu\,dx^\mu$ so that $F = d\omega$. Now,

$$d\omega = \partial_\nu A_\mu\,dx^\nu\,dx^\mu, \tag{4.240}$$

which, in standard notation, leads to the conventional relativistic form of the electromagnetic field tensor,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{4.241}$$

Maxwell's homogeneous equations, $dF = 0$, are thus equivalent to $\partial^\nu F_{\mu\nu} = 0$.

In order to derive similarly the inhomogeneous Maxwell's equations, we introduce the dual electromagnetic field tensor

$$\tilde{F}^{\mu\nu} = \frac{1}{2}\varepsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}, \tag{4.242}$$

and, in terms of differential forms,

$$*F = *\left(F_{\mu\nu}\,dx^\mu\,dx^\nu\right) = F_{\mu\nu} * \left(dx^\mu\,dx^\nu\right) = \frac{1}{2}F_{\mu\nu}\varepsilon^{\mu\nu}{}_{\alpha\beta}\,dx^\alpha\,dx^\beta. \tag{4.243}$$

Applying the exterior derivative yields

$$d(*F) = \frac{1}{2}\varepsilon^{\mu\nu}{}_{\alpha\beta}(\partial_\gamma F_{\mu\nu})\,dx^\gamma\,dx^\alpha\,dx^\beta, \tag{4.244}$$

the left-hand side of Maxwell's inhomogeneous equations, a 3-form. Its driving term is the dual of the electric current density, a 3-form:

$$\begin{aligned} *J &= J_\alpha\big(*dx^\alpha\big) = J_\alpha\varepsilon^\alpha{}_{\mu\nu\lambda}\,dx^\mu\,dx^\nu\,dx^\lambda \\ &= \rho\,dx\,dy\,dz - J_x\,dt\,dy\,dz - J_y\,dt\,dz\,dx - J_z\,dt\,dx\,dy. \end{aligned} \tag{4.245}$$

Altogether Maxwell's inhomogeneous equations take the elegant form

$$d(*F) = *J. \tag{4.246}$$

■

The differential-form framework has brought considerable unification to vector algebra and to tensor analysis on manifolds more generally, such as uniting Stokes' and Gauss' theorems and providing an elegant reformulation of Maxwell's equations and an efficient derivation of the Laplacian in curved orthogonal coordinates, among others.

Exercises

4.8.1 Evaluate the 1-form $a\,dx + 2b\,dy + 4c\,dz$ on the line segment PQ, with $P = (3, 5, 7)$, $Q = (7, 5, 3)$.

4.8.2 If the force field is constant and moving a particle from the origin to $(3, 0, 0)$ requires a units of work, from $(-1, -1, 0)$ to $(-1, 1, 0)$ takes b units of work, and from $(0, 0, 4)$ to $(0, 0, 5)$ c units of work, find the 1-form of the work.

4.8.3 Evaluate the flow described by the 2-form $dx\,dy + 2\,dy\,dz + 3\,dz\,dx$ across the oriented triangle PQR with corners at

$$P = (3, 1, 4), \qquad Q = (-2, 1, 4), \qquad R = (1, 4, 1).$$

4.8.4 Are the points, in this order,

$$(0, 1, 1), \qquad (3, -1, -2), \qquad (4, 2, -2), \qquad (-1, 0, 1)$$

coplanar, or do they form an oriented volume (right-handed or left-handed)?

4.8.5 Write Oersted's law,

$$\int_{\partial S} \mathbf{H}\cdot d\mathbf{r} = \int_S \nabla\times\mathbf{H}\cdot d\mathbf{a} \sim I,$$

in differential form notation.

4.8.6 Describe the electric field by the 1-form $E_1\,dx + E_2\,dy + E_3\,dz$ and the magnetic induction by the 2-form $B_1\,dy\,dz + B_2\,dz\,dx + B_3\,dx\,dy$. Then formulate Faraday's induction law in terms of these forms.

4.8.7 Evaluate the 1-form

$$\frac{x\,dy}{x^2+y^2}-\frac{y\,dx}{x^2+y^2}$$

on the unit circle about the origin oriented counterclockwise.

4.8.8 Find the pullback of $dx\,dz$ under $x=u\cos v$, $y=u-v$, $z=u\sin v$.

4.8.9 Find the pullback of the 2-form $dy\,dz+dz\,dx+dx\,dy$ under the map $x=\sin\theta\cos\varphi$, $y=\sin\theta\sin\varphi$, $z=\cos\theta$.

4.8.10 Parameterize the surface obtained by rotating the circle $(x-2)^2+z^2=1$, $y=0$, about the z-axis in a counterclockwise orientation, as seen from outside.

4.8.11 A 1-form $A\,dx+B\,dy$ is defined as **closed** if $\frac{\partial A}{\partial y}=\frac{\partial B}{\partial x}$. It is called **exact** if there is a function f so that $\frac{\partial f}{\partial x}=A$ and $\frac{\partial f}{\partial y}=B$. Determine which of the following 1-forms are closed, or exact, and find the corresponding functions f for those that are exact:

$$y\,dx+x\,dy,\qquad \frac{y\,dx+x\,dy}{x^2+y^2},\qquad \big[\ln(xy)+1\big]dx+\frac{x}{y}dy,$$

$$-\frac{y\,dx}{x^2+y^2}+\frac{x\,dy}{x^2+y^2},\qquad f(z)\,dz \text{ with } z=x+iy.$$

4.8.12 Show that $\sum_{i=1}^{n}x_i^2=a^2$ defines a differentiable manifold of dimension $D=n-1$ if $a\neq 0$ and $D=0$ if $a=0$.

4.8.13 Show that the set of orthogonal 2×2 matrices form a differentiable manifold, and determine its dimension.

4.8.14 Determine the value of the 2-form $A\,dy\,dz+B\,dz\,dx+C\,dx\,dy$ on a parallelogram with sides $\mathbf{a}$, $\mathbf{b}$.

4.8.15 Prove Lorentz invariance of Maxwell's equations in the language of differential forms.

Additional Readings

Buerger, M. J., *Elementary Crystallography*. New York: Wiley (1956). A comprehensive discussion of crystal symmetries. Buerger develops all 32 point groups and all 230 space groups. Related books by this author include *Contemporary Crystallography*. New York: McGraw-Hill (1970); *Crystal Structure Analysis*. New York: Krieger (1979) (reprint, 1960); and *Introduction to Crystal Geometry*. New York: Krieger (1977) (reprint, 1971).

Burns, G., and A. M. Glazer, *Space Groups for Solid-State Scientists*. New York: Academic Press (1978). A well-organized, readable treatment of groups and their application to the solid state.

de-Shalit, A., and I. Talmi, *Nuclear Shell Model*. New York: Academic Press (1963). We adopt the Condon–Shortley phase conventions of this text.

Edmonds, A. R., *Angular Momentum in Quantum Mechanics*. Princeton, NJ: Princeton University Press (1957).

Edwards, H. M., *Advanced Calculus: A Differential Forms Approach*. Boston: Birkhäuser (1994).

Falicov, L. M., *Group Theory and Its Physical Applications*. Notes compiled by A. Luehrmann. Chicago: University of Chicago Press (1966). Group theory, with an emphasis on applications to crystal symmetries and solid-state physics.

Gell-Mann, M., and Y. Ne'eman, *The Eightfold Way*. New York: Benjamin (1965). A collection of reprints of significant papers on SU(3) and the particles of high-energy physics. Several introductory sections by Gell-Mann and Ne'eman are especially helpful.

Greiner, W., and B. Müller, *Quantum Mechanics Symmetries*. Berlin: Springer (1989). We refer to this textbook for more details and numerous exercises that are worked out in detail.

Hamermesh, M., *Group Theory and Its Application to Physical Problems*. Reading, MA: Addison-Wesley (1962). A detailed, rigorous account of both finite and continuous groups. The 32 point groups are developed. The continuous groups are treated, with Lie algebra included. A wealth of applications to atomic and nuclear physics.

Hassani, S., *Foundations of Mathematical Physics*. Boston: Allyn and Bacon (1991).

Heitler, W., *The Quantum Theory of Radiation*, 2nd ed. Oxford: Oxford University Press (1947). Reprinted, New York: Dover (1983).

Higman, B., *Applied Group-Theoretic and Matrix Methods*. Oxford: Clarendon Press (1955). A rather complete and unusually intelligible development of matrix analysis and group theory.

Jackson, J. D., *Classical Electrodynamics*, 3rd ed. New York: Wiley (1998).

Messiah, A., *Quantum Mechanics*, Vol. II. Amsterdam: North-Holland (1961).

Panofsky, W. K. H., and M. Phillips, *Classical Electricity and Magnetism*, 2nd ed. Reading, MA: Addison-Wesley (1962). The Lorentz covariance of Maxwell's equations is developed for both vacuum and material media. Panofsky and Phillips use contravariant and covariant tensors.

Park, D., Resource letter SP-1 on symmetry in physics. *Am. J. Phys.* **36**: 577–584 (1968). Includes a large selection of basic references on group theory and its applications to physics: atoms, molecules, nuclei, solids, and elementary particles.

Ram, B., Physics of the SU(3) symmetry model. *Am. J. Phys.* **35**: 16 (1967). An excellent discussion of the applications of SU(3) to the strongly interacting particles (baryons). For a sequel to this see R. D. Young, Physics of the quark model. *Am. J. Phys.* **41**: 472 (1973).

Rose, M. E., *Elementary Theory of Angular Momentum*. New York: Wiley (1957). Reprinted. New York: Dover (1995). As part of the development of the quantum theory of angular momentum, Rose includes a detailed and readable account of the rotation group.

Wigner, E. P., *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (translated by J. J. Griffin). New York: Academic Press (1959). This is the classic reference on group theory for the physicist. The rotation group is treated in considerable detail. There is a wealth of applications to atomic physics.

CHAPTER 5

INFINITE SERIES

5.1 FUNDAMENTAL CONCEPTS

Infinite series, literally summations of an infinite number of terms, occur frequently in both pure and applied mathematics. They may be used by the pure mathematician to define functions as a fundamental approach to the theory of functions, as well as for calculating accurate values of transcendental constants and transcendental functions. In the mathematics of science and engineering infinite series are ubiquitous, for they appear in the evaluation of integrals (Sections 5.6 and 5.7), in the solution of differential equations (Sections 9.5 and 9.6), and as Fourier series (Chapter 14) and compete with integral representations for the description of a host of special functions (Chapters 11, 12, and 13). In Section 16.3 the Neumann series solution for integral equations provides one more example of the occurrence and use of infinite series.

Right at the start we face the problem of attaching meaning to the sum of an infinite number of terms. The usual approach is by partial sums. If we have an infinite sequence of terms $u_1, u_2, u_3, u_4, u_5, \ldots$, we define the ith partial sum as

$$s_i = \sum_{n=1}^{i} u_n. \tag{5.1}$$

This is a finite summation and offers no difficulties. If the partial sums s_i converge to a (finite) limit as $i \to \infty$,

$$\lim_{i\to\infty} s_i = S, \tag{5.2}$$

the infinite series $\sum_{n=1}^{\infty} u_n$ is said to be **convergent** and to have the value S. Note that we reasonably, plausibly, but still arbitrarily **define** the infinite series as equal to S and that a necessary condition for this convergence to a limit is that $\lim_{n\to\infty} u_n = 0$. This condition, however, is not sufficient to guarantee convergence. Equation (5.2) is usually written in formal mathematical notation:

The condition for the existence of a limit S is that for each $\varepsilon > 0$, there is a fixed $N = N(\varepsilon)$ such that

$$|S - s_i| < \varepsilon, \qquad \text{for } i > N.$$

This condition is often derived from the Cauchy criterion applied to the partial sums s_i. The **Cauchy criterion** is:

A necessary and sufficient condition that a sequence (s_i) converge is that for each $\varepsilon > 0$ there is a fixed number N such that

$$|s_j - s_i| < \varepsilon, \qquad \text{for all } i, j > N.$$

This means that the individual partial sums must cluster together as we move far out in the sequence.

The Cauchy criterion may easily be extended to sequences of functions. We see it in this form in Section 5.5 in the definition of uniform convergence and in Section 10.4 in the development of Hilbert space. Our partial sums s_i may not converge to a single limit but may oscillate, as in the case

$$\sum_{n=1}^{\infty} u_n = 1 - 1 + 1 - 1 + 1 + \cdots - (-1)^n + \cdots.$$

Clearly, $s_i = 1$ for i odd but $s_i = 0$ for i even. There is no convergence to a limit, and series such as this one are labeled **oscillatory**. Whenever the sequence of partial sums diverges (approaches $\pm\infty$), the infinite series is said to **diverge**. Often the term *divergent* is extended to include oscillatory series as well. Because we evaluate the partial sums by ordinary arithmetic, the convergent series, defined in terms of a limit of the partial sums, assumes a position of supreme importance. Two examples may clarify the nature of convergence or divergence of a series and will also serve as a basis for a further detailed investigation in the next section.

Example 5.1.1 THE GEOMETRIC SERIES

The geometrical sequence, starting with a and with a ratio r $(= a_{n+1}/a_n$ independent of $n)$, is given by

$$a + ar + ar^2 + ar^3 + \cdots + ar^{n-1} + \cdots.$$

The nth partial sum is given by[1]

$$s_n = a\frac{1 - r^n}{1 - r}. \tag{5.3}$$

Taking the limit as $n \to \infty$,

$$\lim_{n\to\infty} s_n = \frac{a}{1 - r}, \quad \text{for } |r| < 1. \tag{5.4}$$

[1]Multiply and divide $s_n = \sum_{m=0}^{n-1} ar^m$ by $1 - r$.

Hence, by definition, the infinite geometric series converges for $|r| < 1$ and is given by

$$\boxed{\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r}.} \tag{5.5}$$

On the other hand, if $|r| \geq 1$, the necessary condition $u_n \to 0$ is not satisfied and the infinite series diverges. ■

Example 5.1.2 THE HARMONIC SERIES

As a second and more involved example, we consider the harmonic series

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + \cdots. \tag{5.6}$$

We have the $\lim_{n\to\infty} u_n = \lim_{n\to\infty} 1/n = 0$, but this is not sufficient to guarantee convergence. If we group the terms (no change in order) as

$$1 + \tfrac{1}{2} + \left(\tfrac{1}{3} + \tfrac{1}{4}\right) + \left(\tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8}\right) + \left(\tfrac{1}{9} + \cdots + \tfrac{1}{16}\right) + \cdots, \tag{5.7}$$

each pair of parentheses encloses p terms of the form

$$\frac{1}{p+1} + \frac{1}{p+2} + \cdots + \frac{1}{p+p} > \frac{p}{2p} = \frac{1}{2}. \tag{5.8}$$

Forming partial sums by adding the parenthetical groups one by one, we obtain

$$\begin{aligned} s_1 &= 1, & s_4 &> \frac{5}{2}, \\ s_2 &= \frac{3}{2}, & s_5 &> \frac{6}{2}, \cdots \\ s_3 &> \frac{4}{2}, & s_n &> \frac{n+1}{2}. \end{aligned} \tag{5.9}$$

The harmonic series considered in this way is certainly divergent.[2] An alternate and independent demonstration of its divergence appears in Section 5.2. ■

If the $u_n > 0$ are monotonically decreasing to zero, that is, $u_n > u_{n+1}$ for all n, then $\sum_n u_n$ is converging to S if, and only if, $s_n - nu_n$ converges to S. As the partial sums s_n converge to S, this theorem implies that $nu_n \to 0$, for $n \to \infty$.

To prove this **theorem**, we start by concluding from $0 < u_{n+1} < u_n$ and

$$s_{n+1} - (n+1)u_{n+1} = s_n - nu_{n+1} = s_n - nu_n + n(u_n - u_{n+1}) > s_n - nu_n$$

that $s_n - nu_n$ increases as $n \to \infty$. As a consequence of $s_n - nu_n < s_n \leq S$, $s_n - nu_n$ converges to a value $s \leq S$. Deleting the tail of positive terms $u_i - u_n$ from $i = \nu + 1$ to n,

[2] The (finite) harmonic series appears in an interesting note on the maximum stable displacement of a stack of coins. P. R. Johnson, The Leaning Tower of Lire. *Am. J. Phys.* **23**: 240 (1955).

we infer from $s_n - nu_n > u_0 + (u_1 - u_n) + \cdots + (u_\nu - u_n) = s_\nu - \nu u_n$ that $s_n - nu_n \geq s_\nu$ for $n \to \infty$. Hence also $s \geq S$, so $s = S$ and $nu_n \to 0$.

When this theorem is applied to the **harmonic series** $\sum_n \frac{1}{n}$ with $n\frac{1}{n} = 1$ it implies that it does not converge; it diverges to $+\infty$.

Addition, Subtraction of Series

If we have two convergent series $\sum_n u_n \to s$ and $\sum_n v_n \to S$, their sum and difference will also converge to $s \pm S$ because their partial sums satisfy

$$\left|s_j \pm S_j - (s_i \pm S_i)\right| = \left|s_j - s_i \pm (S_j - S_i)\right| \leq |s_j - s_i| + |S_j - S_i| < 2\epsilon$$

using the triangle inequality

$$|a| - |b| \leq |a + b| \leq |a| + |b|$$

for $a = s_j - s_i, b = S_j - S_i$.

A convergent series $\sum_n u_n \to S$ may be multiplied termwise by a real number a. The new series will converge to aS because

$$|as_j - as_i| = \left|a(s_j - s_i)\right| = |a||s_j - s_i| < |a|\epsilon.$$

This multiplication by a constant can be generalized to a multiplication by terms c_n of a bounded sequence of numbers.

If $\sum_n u_n$ converges to S and $0 < c_n \leq M$ are bounded, then $\sum_n u_n c_n$ is convergent. If $\sum_n u_n$ is divergent and $c_n > M > 0$, then $\sum_n u_n c_n$ diverges.

To prove this **theorem** we take i, j sufficiently large so that $|s_j - s_i| < \epsilon$. Then

$$\sum_{i+1}^{j} u_n c_n \leq M \sum_{i+1}^{j} u_n = M|s_j - s_i| < M\epsilon.$$

The divergent case follows from

$$\sum_n u_n c_n > M \sum_n u_n \to \infty.$$

Using the binomial theorem[3] (Section 5.6), we may expand the function $(1 + x)^{-1}$:

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots + (-x)^{n-1} + \cdots . \tag{5.10}$$

If we let $x \to 1$, this series becomes

$$1 - 1 + 1 - 1 + 1 - 1 + \cdots , \tag{5.11}$$

a series that we labeled oscillatory earlier in this section. Although it does not converge in the usual sense, meaning can be attached to this series. Euler, for example, assigned a value of $1/2$ to this oscillatory sequence on the basis of the correspondence between this series and the well-defined function $(1 + x)^{-1}$. Unfortunately, such correspondence between series and function is not unique, and this approach must be refined. Other methods

[3] Actually Eq. (5.10) may be verified by multiplying both sides by $1 + x$.

of assigning a meaning to a divergent or oscillatory series, methods of defining a sum, have been developed. See G. H. Hardy, *Divergent Series*, Chelsea Publishing Co. 2nd ed. (1992). In general, however, this aspect of infinite series is of relatively little interest to the scientist or the engineer. An exception to this statement, the very important asymptotic or semiconvergent series, is considered in Section 5.10.

Exercises

5.1.1 Show that

$$\sum_{n=1}^{\infty} \frac{1}{(2n-1)(2n+1)} = \frac{1}{2}.$$

Hint. Show (by mathematical induction) that $s_m = m/(2m+1)$.

5.1.2 Show that

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1.$$

Find the partial sum s_m and verify its correctness by mathematical induction.
Note. The method of expansion in partial fractions, Section 15.8, offers an alternative way of solving Exercises 5.1.1 and 5.1.2.

5.2 Convergence Tests

Although nonconvergent series may be useful in certain special cases (compare Section 5.10), we usually insist, as a matter of convenience if not necessity, that our series be convergent. It therefore becomes a matter of extreme importance to be able to tell whether a given series is convergent. We shall develop a number of possible tests, starting with the simple and relatively insensitive tests and working up to the more complicated but quite sensitive tests. For the present let us consider a **series of positive terms** $a_n \geq 0$, postponing negative terms until the next section.

Comparison Test

If term by term a series of terms $0 \leq u_n \leq a_n$, in which the a_n form a convergent series, the series $\sum_n u_n$ is also convergent. If $u_n \leq a_n$ for all n, then $\sum_n u_n \leq \sum_n a_n$ and $\sum_n u_n$ therefore is **convergent**. If term by term a series of terms $v_n \geq b_n$, in which the b_n, form a divergent series, the series $\sum_n v_n$ is also **divergent**. Note that comparisons of u_n with b_n or v_n with a_n yield no information. If $v_n \geq b_n$ for all n, then $\sum_n v_n \geq \sum_n b_n$ and $\sum_n v_n$ therefore is divergent.

For the convergent series a_n we already have the geometric series, whereas the harmonic series will serve as the divergent comparison series b_n. As other series are identified as either convergent or divergent, they may be used for the known series in this comparison test. All tests developed in this section are essentially comparison tests. Figure 5.1 exhibits these tests and the interrelationships.

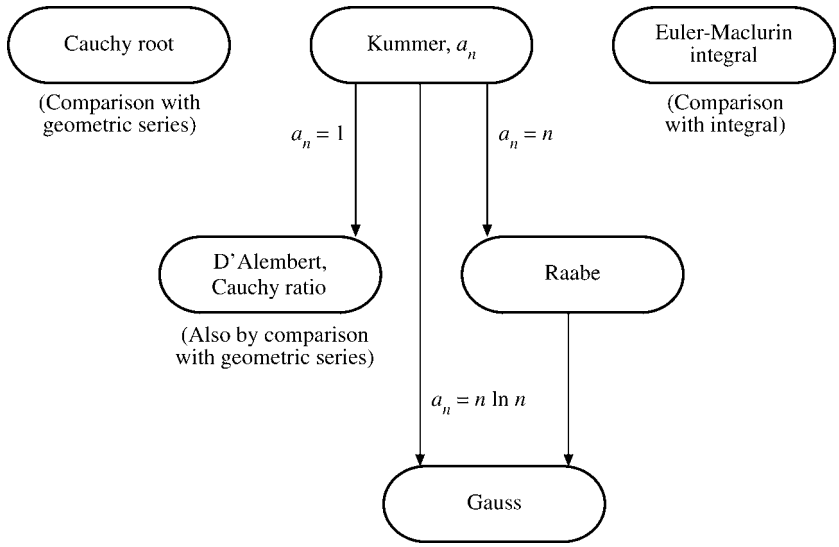

FIGURE 5.1 Comparison tests.

Example 5.2.1 A DIRICHLET SERIES

Test $\sum_{n=1}^{\infty} n^{-p}$, $p = 0.999$, for convergence. Since $n^{-0.999} > n^{-1}$ and $b_n = n^{-1}$ forms the divergent harmonic series, the comparison test shows that $\sum_n n^{-0.999}$ is divergent. Generalizing, $\sum_n n^{-p}$ is seen to be divergent for all $p \leq 1$ but convergent for $p > 1$ (see Example 5.2.3). ■

Cauchy Root Test

If $(a_n)^{1/n} \leq r < 1$ for all sufficiently large n, with r independent of n, then $\sum_n a_n$ is convergent. If $(a_n)^{1/n} \geq 1$ for all sufficiently large n, then $\sum_n a_n$ is divergent.

The first part of this test is verified easily by raising $(a_n)^{1/n} \leq r$ to the nth power. We get

$$a_n \leq r^n < 1.$$

Since r^n is just the nth term in a convergent geometric series, $\sum_n a_n$ is convergent by the comparison test. Conversely, if $(a_n)^{1/n} \geq 1$, then $a_n \geq 1$ and the series must diverge. This root test is particularly useful in establishing the properties of power series (Section 5.7).

D'Alembert (or Cauchy) Ratio Test

If $a_{n+1}/a_n \leq r < 1$ for all sufficiently large n and r is independent of n, then $\sum_n a_n$ is convergent. If $a_{n+1}/a_n \geq 1$ for all sufficiently large n, then $\sum_n a_n$ is divergent.

Convergence is proved by direct comparison with the geometric series $(1 + r + r^2 + \cdots)$. In the second part, $a_{n+1} \geq a_n$ and divergence should be reasonably obvious. Although not

quite so sensitive as the Cauchy root test, this D'Alembert ratio test is one of the easiest to apply and is widely used. An alternate statement of the ratio test is in the form of a limit: If

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n}\begin{cases}<1, & \text{convergence,}\\ >1, & \text{divergence,}\\ =1, & \text{indeterminate.}\end{cases} \tag{5.12}$$

Because of this final indeterminate possibility, the ratio test is likely to fail at crucial points, and more delicate, sensitive tests are necessary. The alert reader may wonder how this indeterminacy arose. Actually it was concealed in the first statement, $a_{n+1}/a_n \le r < 1$. We might encounter $a_{n+1}/a_n < 1$ for all **finite** n but be unable to choose an $r < 1$ **and independent of n** such that $a_{n+1}/a_n \le r$ for all sufficiently large n. An example is provided by the harmonic series

$$\frac{a_{n+1}}{a_n} = \frac{n}{n+1} < 1. \tag{5.13}$$

Since

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n} = 1, \tag{5.14}$$

no fixed ratio $r < 1$ exists and the ratio test fails.

Example 5.2.2 D'ALEMBERT RATIO TEST

Test $\sum_n n/2^n$ for convergence.

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)/2^{n+1}}{n/2^n} = \frac{1}{2}\cdot\frac{n+1}{n}. \tag{5.15}$$

Since

$$\frac{a_{n+1}}{a_n} \le \frac{3}{4} \qquad \text{for } n \ge 2, \tag{5.16}$$

we have convergence. Alternatively,

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n} = \frac{1}{2} \tag{5.17}$$

and again — convergence. ■

Cauchy (or Maclaurin) Integral Test

This is another sort of comparison test, in which we compare a series with an integral. Geometrically, we compare the area of a series of unit-width rectangles with the area under a curve.

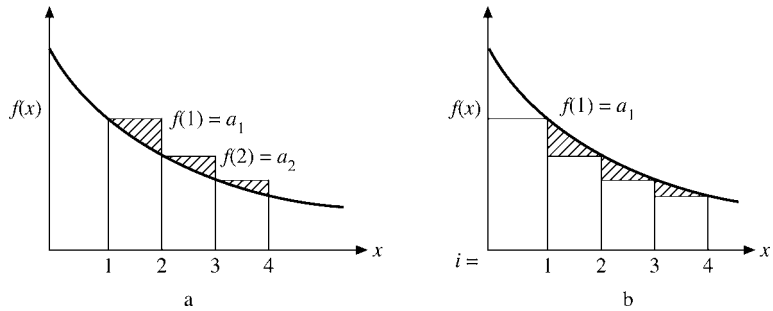

FIGURE 5.2 (a) Comparison of integral and sum-blocks leading. (b) Comparison of integral and sum-blocks lagging.

Let $f(x)$ be a continuous, **monotonic decreasing function** in which $f(n) = a_n$. Then $\sum_n a_n$ converges if $\int_1^\infty f(x)\,dx$ is finite and diverges if the integral is infinite. For the ith partial sum,

$$s_i = \sum_{n=1}^{i} a_n = \sum_{n=1}^{i} f(n). \tag{5.18}$$

But

$$s_i > \int_1^{i+1} f(x)\,dx \tag{5.19}$$

from Fig. 5.2a, $f(x)$ being monotonic decreasing. On the other hand, from Fig. 5.2b,

$$s_i - a_1 < \int_1^{i} f(x)\,dx, \tag{5.20}$$

in which the series is represented by the inscribed rectangles. Taking the limit as $i \to \infty$, we have

$$\boxed{\int_1^\infty f(x)\,dx \le \sum_{n=1}^{\infty} a_n \le \int_1^\infty f(x)\,dx + a_1.} \tag{5.21}$$

Hence the infinite series converges or diverges as the corresponding integral converges or diverges. This integral test is particularly useful in setting upper and lower bounds on the remainder of a series after some number of initial terms have been summed. That is,

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{N} a_n + \sum_{n=N+1}^{\infty} a_n,$$

where

$$\boxed{\int_{N+1}^\infty f(x)\,dx \le \sum_{n=N+1}^{\infty} a_n \le \int_{N+1}^\infty f(x)\,dx + a_{N+1}.}$$

To free the integral test from the quite restrictive requirement that the interpolating function $f(x)$ be positive and monotonic, we show for any function $f(x)$ with a continuous derivative that

$$\sum_{n=N_i+1}^{N_f} f(n) = \int_{N_i}^{N_f} f(x)\,dx + \int_{N_i}^{N_f} \big(x-[x]\big) f'(x)\,dx \tag{5.22}$$

holds. Here $[x]$ denotes the largest integer below x, so $x-[x]$ varies sawtoothlike between 0 and 1. To derive Eq. (5.22) we observe that

$$\int_{N_i}^{N_f} x f'(x)\,dx = N_f f(N_f) - N_i f(N_i) - \int_{N_i}^{N_f} f(x)\,dx, \tag{5.23}$$

using integration by parts. Next we evaluate the integral

$$\begin{aligned}\int_{N_i}^{N_f} [x] f'(x)\,dx &= \sum_{n=N_i}^{N_f-1} n \int_n^{n+1} f'(x)\,dx = \sum_{n=N_i}^{N_f-1} n\big\{f(n+1)-f(n)\big\} \\ &= -\sum_{n=N_i+1}^{N_f} f(n) - N_i f(N_i) + N_f f(N_f). \end{aligned}\tag{5.24}$$

Subtracting Eq. (5.24) from (5.23) we arrive at Eq. (5.22). Note that $f(x)$ may go up or down and even change sign, so Eq. (5.22) applies to alternating series (see Section 5.3) as well. Usually $f'(x)$ falls faster than $f(x)$ for $x\to\infty$, so the remainder term in Eq. (5.22) converges better. It is easy to improve Eq. (5.22) by replacing $x-[x]$ by $x-[x]-\frac{1}{2}$, which varies between $-\frac{1}{2}$ and $\frac{1}{2}$:

$$\begin{aligned}\sum_{N_i<n\le N_f} f(n) = \int_{N_i}^{N_f} f(x)\,dx + \int_{N_i}^{N_f} \big(x-[x]-\tfrac{1}{2}\big) f'(x)\,dx \\ + \tfrac{1}{2}\big\{f(N_f)-f(N_i)\big\}.\end{aligned}\tag{5.25}$$

Then the $f'(x)$-integral becomes even smaller, if $f'(x)$ does not change sign too often. For an application of this integral test to an alternating series see Example 5.3.1.

Example 5.2.3 RIEMANN ZETA FUNCTION

The Riemann zeta function is defined by

$$\boxed{\zeta(p) = \sum_{n=1}^{\infty} n^{-p},} \tag{5.26}$$

provided the series converges. We may take $f(x)=x^{-p}$, and then

$$\begin{aligned}\int_1^{\infty} x^{-p}\,dx &= \frac{x^{-p+1}}{-p+1}\bigg|_1^{\infty}, \qquad p\neq 1 \\ &= \ln x\,|_{x=1}^{\infty}, \qquad p=1.\end{aligned}\tag{5.27}$$

The integral and therefore the series are divergent for $p \le 1$, convergent for $p > 1$. Hence Eq. (5.26) should carry the condition $p > 1$. This, incidentally, is an independent proof that the harmonic series ($p = 1$) diverges logarithmically. The sum of the first million terms $\sum^{1,000,000} n^{-1}$ is only 14.392 726.... ■

This integral comparison may also be used to set an upper limit to the Euler–Mascheroni constant,[4] defined by

$$\boxed{\gamma = \lim_{n\to\infty}\left(\sum_{m=1}^{n} m^{-1} - \ln n\right).} \tag{5.28}$$

Returning to partial sums, Eq. (5.20) yields

$$s_n = \sum_{m=1}^{n} m^{-1} - \ln n \le \int_1^n \frac{dx}{x} - \ln n + 1. \tag{5.29}$$

Evaluating the integral on the right, $s_n < 1$ for all n and therefore $\gamma \le 1$. Exercise 5.2.12 leads to more restrictive bounds. Actually the Euler–Mascheroni constant is 0.57721566....

Kummer's Test

This is the first of three tests that are somewhat more difficult to apply than the preceding tests. Their importance lies in their power and sensitivity. Frequently, at least one of the three will work when the simpler, easier tests are indecisive. It must be remembered, however, that these tests, like those previously discussed, are ultimately based on comparisons. It can be shown that there is no most slowly converging series and no most slowly diverging series. This means that all convergence tests given here, including Kummer's, may fail sometime.

We consider a series of positive terms u_i and a sequence of finite positive constants a_i. If

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} \ge C > 0 \tag{5.30}$$

for all $n \ge N$, where N is some fixed number,[5] then $\sum_{i=1}^{\infty} u_i$ **converges**. If

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} \le 0 \tag{5.31}$$

and $\sum_{i=1}^{\infty} a_i^{-1}$ diverges, then $\sum_{i=1}^{\infty} u_i$ **diverges**.

[4]This is the notation of National Bureau of Standards, *Handbook of Mathematical Functions,* Applied Mathematics Series-55 (AMS-55). New York: Dover (1972).

[5]With u_m finite, the partial sum s_N will always be finite for N finite. The convergence or divergence of a series depends on the behavior of the last infinity of terms, not on the first N terms.

The proof of this powerful test is remarkably simple. From Eq. (5.30), with C some positive constant,

$$\begin{aligned} Cu_{N+1} &\le a_N u_N - a_{N+1}u_{N+1} \\ Cu_{N+2} &\le a_{N+1}u_{N+1} - a_{N+2}u_{N+2} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ Cu_n &\le a_{n-1}u_{n-1} - a_n u_n. \end{aligned} \tag{5.32}$$

Adding and dividing by C (and recalling that $C \neq 0$), we obtain

$$\sum_{i=N+1}^{n} u_i \le \frac{a_N u_N}{C} - \frac{a_n u_n}{C}. \tag{5.33}$$

Hence for the partial sum s_n,

$$\begin{aligned} s_n &\le \sum_{i=1}^{N} u_i + \frac{a_N u_N}{C} - \frac{a_n u_n}{C} \\ &< \sum_{i=1}^{N} u_i + \frac{a_N u_N}{C}, \qquad \text{a constant, independent of } n. \end{aligned} \tag{5.34}$$

The partial sums therefore have an upper bound. With zero as an obvious lower bound, the series $\sum u_i$ must converge.

Divergence is shown as follows. From Eq. (5.31) for $u_{n+1} > 0$,

$$a_n u_n \ge a_{n-1}u_{n-1} \ge \cdots \ge a_N u_N, \qquad n > N. \tag{5.35}$$

Thus, for $a_n > 0$,

$$u_n \ge \frac{a_N u_N}{a_n} \tag{5.36}$$

and

$$\sum_{i=N+1}^{\infty} u_i \ge a_N u_N \sum_{i=N+1}^{\infty} a_i^{-1}. \tag{5.37}$$

If $\sum_{i=1}^{\infty} a_i^{-1}$ diverges, then by the comparison test $\sum_i u_i$ diverges. Equations (5.30) and (5.31) are often given in a limit form:

$$\lim_{n\to\infty}\left(a_n \frac{u_n}{u_{n+1}} - a_{n+1}\right) = C. \tag{5.38}$$

Thus for $C > 0$ we have convergence, whereas for $C < 0$ (and $\sum_i a_i^{-1}$ divergent) we have divergence. It is perhaps useful to show the close relation of Eq. (5.38) and Eqs. (5.30) and (5.31) and to show why indeterminacy creeps in when the limit $C = 0$. From the definition of limit,

$$\left| a_n \frac{u_n}{u_{n+1}} - a_{n+1} - C \right| < \varepsilon \tag{5.39}$$

for all $n \geq N$ and all $\varepsilon > 0$, no matter how small ε may be. When the absolute value signs are removed,

$$C - \varepsilon < a_n \frac{u_n}{u_{n+1}} - a_{n+1} < C + \varepsilon. \tag{5.40}$$

Now, if $C > 0$, Eq. (5.30) follows from ε sufficiently small. On the other hand, if $C < 0$, Eq. (5.31) follows. However, if $C = 0$, the center term, $a_n(u_n/u_{n+1}) - a_{n+1}$, may be either positive or negative and the proof fails. The primary use of Kummer's test is to prove other tests, such as Raabe's (compare also Exercise 5.2.3).

If the positive constants a_n of Kummer's test are chosen $a_n = n$, we have Raabe's test.

Raabe's Test

If $u_n > 0$ and if

$$n\left(\frac{u_n}{u_{n+1}} - 1\right) \geq P > 1 \tag{5.41}$$

for all $n \geq N$, where N is a positive integer independent of n, **then** $\sum_i u_i$ **converges**. Here, $P = C + 1$ of Kummer's test. If

$$n\left(\frac{u_n}{u_{n+1}} - 1\right) \leq 1, \tag{5.42}$$

then $\sum_i u_i$ diverges (as $\sum_n n^{-1}$ diverges). The limit form of Raabe's test is

$$\lim_{n\to\infty} n\left(\frac{u_n}{u_{n+1}} - 1\right) = P. \tag{5.43}$$

We have convergence for $P > 1$, divergence for $P < 1$, and no conclusion for $P = 1$, exactly as with the Kummer test. This indeterminacy is pointed up by Exercise 5.2.4, which presents a convergent series and a divergent series, with both series yielding $P = 1$ in Eq. (5.43).

Raabe's test is more sensitive than the d'Alembert ratio test (Exercise 5.2.3) because $\sum_{n=1}^{\infty} n^{-1}$ diverges more slowly than $\sum_{n=1}^{\infty} 1$. We obtain a more sensitive test (and one that is still fairly easy to apply) by choosing $a_n = n \ln n$. This is Gauss' test.

Gauss' Test

If $u_n > 0$ for all finite n and

$$\frac{u_n}{u_{n+1}} = 1 + \frac{h}{n} + \frac{B(n)}{n^2}, \tag{5.44}$$

in which $B(n)$ is a bounded function of n for $n \to \infty$, then $\sum_i u_i$ converges for $h > 1$ and diverges for $h \leq 1$: There is no indeterminate case here.

The Gauss test is an extremely sensitive test of series convergence. It will work for all series the physicist is likely to encounter. For $h > 1$ or $h < 1$ the proof follows directly from Raabe's test

$$\lim_{n\to\infty} n\left[1+\frac{h}{n}+\frac{B(n)}{n^2}-1\right]=\lim_{n\to\infty}\left[h+\frac{B(n)}{n}\right]=h. \tag{5.45}$$

If $h = 1$, Raabe's test fails. However, if we return to Kummer's test and use $a_n = n\ln n$, Eq. (5.38) leads to

$$\begin{aligned}\lim_{n\to\infty}&\left\{n\ln n\left[1+\frac{1}{n}+\frac{B(n)}{n^2}\right]-(n+1)\ln(n+1)\right\}\\ &=\lim_{n\to\infty}\left[n\ln n\cdot\frac{n+1}{n}-(n+1)\ln(n+1)\right]\\ &=\lim_{n\to\infty}(n+1)\left[\ln n-\ln n-\ln\left(1+\frac{1}{n}\right)\right].\end{aligned} \tag{5.46}$$

Borrowing a result from Section 5.6 (which is not dependent on Gauss' test), we have

$$\begin{aligned}\lim_{n\to\infty}-(n+1)\ln\left(1+\frac{1}{n}\right)&=\lim_{n\to\infty}-(n+1)\left(\frac{1}{n}-\frac{1}{2n^2}+\frac{1}{3n^3}\cdots\right)\\ &=-1<0.\end{aligned} \tag{5.47}$$

Hence we have divergence for $h = 1$. This is an example of a successful application of Kummer's test when Raabe's test had failed.

Example 5.2.4 LEGENDRE SERIES

The recurrence relation for the series solution of Legendre's equation (Exercise 9.5.5) may be put in the form

$$\frac{a_{2j+2}}{a_{2j}}=\frac{2j(2j+1)-l(l+1)}{(2j+1)(2j+2)}. \tag{5.48}$$

For $u_j = a_{2j}$ and $B(j) = O(1/j^2) \to 0$ (that is, $|B(j)j^2| \le C$, $C > 0$, a constant) as $j \to \infty$ in Gauss' test we apply Eq. (5.45). Then, for $j \gg l$,[6]

$$\frac{u_j}{u_{j+1}}\to\frac{(2j+1)(2j+2)}{2j(2j+1)}=\frac{2j+2}{2j}=1+\frac{1}{j}. \tag{5.49}$$

By Eq. (5.44) the series is divergent. ■

[6]The l dependence enters $B(j)$ but does not affect h in Eq. (5.45).

Improvement of Convergence

This section so far has been concerned with establishing convergence as an abstract mathematical property. In practice, the **rate** of convergence may be of considerable importance. Here we present one method of improving the rate of convergence of a convergent series. Other techniques are given in Sections 5.4 and 5.9.

The basic principle of this method, due to Kummer, is to form a linear combination of our slowly converging series and one or more series whose sum is known. For the known series the collection

$$\alpha_1 = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$$

$$\alpha_2 = \sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)} = \frac{1}{4}$$

$$\alpha_3 = \sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)(n+3)} = \frac{1}{18}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\alpha_p = \sum_{n=1}^{\infty} \frac{1}{n(n+1)\cdots(n+p)} = \frac{1}{p \cdot p!}$$

is particularly useful.[7] The series are combined term by term and the coefficients in the linear combination chosen to cancel the most slowly converging terms.

Example 5.2.5 RIEMANN ZETA FUNCTION, $\zeta(3)$

Let the series to be summed be $\sum_{n=1}^{\infty} n^{-3}$. In Section 5.9 this is identified as the Riemann zeta function, $\zeta(3)$. We form a linear combination

$$\sum_{n=1}^{\infty} n^{-3} + a_2\alpha_2 = \sum_{n=1}^{\infty} n^{-3} + \frac{a_2}{4}.$$

α_1 is not included since it converges more slowly than $\zeta(3)$. Combining terms, we obtain on the left-hand side

$$\sum_{n=1}^{\infty} \left\{ \frac{1}{n^3} + \frac{a_2}{n(n+1)(n+2)} \right\} = \sum_{n=1}^{\infty} \frac{n^2(1+a_2) + 3n + 2}{n^3(n+1)(n+2)}.$$

If we choose $a_2 = -1$, the preceding equations yield

$$\zeta(3) = \sum_{n=1}^{\infty} n^{-3} = \frac{1}{4} + \sum_{n=1}^{\infty} \frac{3n+2}{n^3(n+1)(n+2)}. \tag{5.50}$$

[7] These series sums may be verified by expanding the forms by partial fractions, writing out the initial terms, and inspecting the pattern of cancellation of positive and negative terms.

The resulting series may not be beautiful but it does converge as n^{-4}, faster than n^{-3}. A more convenient form comes from Exercise 5.2.21. There, the symmetry leads to convergence as n^{-5}. ■

The method can be extended, including $a_3\alpha_3$ to get convergence as n^{-5}, $a_4\alpha_4$ to get convergence as n^{-6}, and so on. Eventually, you have to reach a compromise between how much algebra you do and how much arithmetic the computer does. As computers get faster, the balance is steadily shifting to less algebra for you and more arithmetic for them.

Exercises

5.2.1 (a) Prove that if

$$\lim_{n\to\infty} n^p u_n = A < \infty, \qquad p > 1,$$

the series $\sum_{n=1}^{\infty} u_n$ converges.

(b) Prove that if

$$\lim_{n\to\infty} n u_n = A > 0,$$

the series diverges. (The test fails for $A = 0$.)

These two tests, known as **limit tests**, are often convenient for establishing the convergence of a series. They may be treated as comparison tests, comparing with

$$\sum_n n^{-q}, \qquad 1 \le q < p.$$

5.2.2 If

$$\lim_{n\to\infty} \frac{b_n}{a_n} = K,$$

a constant with $0 < K < \infty$, show that $\sum_n b_n$ converges or diverges with $\sum a_n$.
Hint. If $\sum a_n$ converges, use $b'_n = \frac{1}{2K} b_n$. If $\sum_n a_n$ diverges, use $b''_n = \frac{2}{K} b_n$.

5.2.3 Show that the complete d'Alembert ratio test follows directly from Kummer's test with $a_i = 1$.

5.2.4 Show that Raabe's test is indecisive for $P = 1$ by establishing that $P = 1$ for the series

(a) $u_n = \dfrac{1}{n \ln n}$ and that this series diverges.

(b) $u_n = \dfrac{1}{n(\ln n)^2}$ and that this series converges.

Note. By direct addition $\sum_2^{100,000} [n(\ln n)^2]^{-1} = 2.02288$. The remainder of the series $n > 10^5$ yields 0.08686 by the integral comparison test. The total, then, 2 to ∞, is 2.1097.

5.2.5 Gauss' test is often given in the form of a test of the ratio

$$\frac{u_n}{u_{n+1}} = \frac{n^2 + a_1 n + a_0}{n^2 + b_1 n + b_0}.$$

For what values of the parameters a_1 and b_1 is there convergence? divergence?

ANS. Convergent for $a_1 - b_1 > 1$,
divergent for $a_1 - b_1 \leq 1$.

5.2.6 Test for convergence

(a) $\displaystyle\sum_{n=2}^{\infty} (\ln n)^{-1}$ (d) $\displaystyle\sum_{n=1}^{\infty} \big[n(n+1)\big]^{-1/2}$

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{n!}{10^n}$ (e) $\displaystyle\sum_{n=0}^{\infty} \frac{1}{2n+1}$.

(c) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{2n(2n+1)}$

5.2.7 Test for convergence

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ (d) $\displaystyle\sum_{n=1}^{\infty} \ln\left(1 + \frac{1}{n}\right)$

(b) $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n \ln n}$ (e) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n \cdot n^{1/n}}$.

(c) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n 2^n}$

5.2.8 For what values of p and q will the following series converge? $\sum_{n=2}^{\infty} \frac{1}{n^p (\ln n)^q}$.

$$\text{ANS. Convergent for } \begin{cases} p > 1, & \text{all } q, \\ p = 1, & q > 1, \end{cases} \qquad \text{divergent for } \begin{cases} p < 1, & \text{all } q, \\ p = 1, & q \leq 1. \end{cases}$$

5.2.9 Determine the range of convergence for Gauss's **hypergeometric series**

$$F(\alpha, \beta, \gamma; x) = 1 + \frac{\alpha\beta}{1!\gamma} x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)} x^2 + \cdots.$$

Hint. Gauss developed his test for the specific purpose of establishing the convergence of this series.

ANS. Convergent for $-1 < x < 1$ and $x = \pm 1$ if $\gamma > \alpha + \beta$.

5.2.10 A pocket calculator yields

$$\sum_{n=1}^{100} n^{-3} = 1.202\ 007.$$

Show that

$$1.202\,056 \le \sum_{n=1}^{\infty} n^{-3} \le 1.202\,057.$$

Hint. Use integrals to set upper and lower bounds on $\sum_{n=101}^{\infty} n^{-3}$.
Note. A more exact value for summation of $\zeta(3) = \sum_{n=1}^{\infty} n^{-3}$ is 1.202 056 903...; $\zeta(3)$ is known to be an irrational number, but it is not linked to known constants such as $e, \pi, \gamma, \ln 2$.

5.2.11 Set upper and lower bounds on $\sum_{n=1}^{1,000,000} n^{-1}$, assuming that

(a) the Euler–Mascheroni constant is known.

$$\text{ANS. } 14.392\,726 < \sum_{n=1}^{1,000,000} n^{-1} < 14.392\,727.$$

(b) The Euler–Mascheroni constant is unknown.

5.2.12 Given $\sum_{n=1}^{1,000} n^{-1} = 7.485\,470\ldots$ set upper and lower bounds on the Euler–Mascheroni constant.

$$\text{ANS. } 0.5767 < \gamma < 0.5778.$$

5.2.13 (From **Olbers' paradox**.) Assume a static universe in which the stars are uniformly distributed. Divide all space into shells of constant thickness; the stars in any one shell by themselves subtend a solid angle of ω_0. **Allowing for the blocking out of distant stars by nearer stars**, show that the total net solid angle subtended by all stars, shells extending to infinity, is **exactly** 4π. [Therefore the night sky should be ablaze with light. For more details, see E. Harrison, *Darkness at Night: A Riddle of the Universe*. Cambridge, MA: Harvard University Press (1987).]

5.2.14 Test for convergence

$$\sum_{n=1}^{\infty} \left[\frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)} \right]^2 = \frac{1}{4} + \frac{9}{64} + \frac{25}{256} + \cdots.$$

5.2.15 The Legendre series $\sum_{j \text{ even}} u_j(x)$ satisfies the recurrence relations

$$u_{j+2}(x) = \frac{(j+1)(j+2) - l(l+1)}{(j+2)(j+3)} x^2 u_j(x),$$

in which the index j is even and l is some constant (but, in this problem, **not** a non-negative odd integer). Find the range of values of x for which this Legendre series is convergent. Test the endpoints.

$$\text{ANS. } -1 < x < 1.$$

5.2.16 A series solution (Section 9.5) of the Chebyshev equation leads to successive terms having the ratio

$$\frac{u_{j+2}(x)}{u_j(x)} = \frac{(k+j)^2 - n^2}{(k+j+1)(k+j+2)} x^2,$$

with $k = 0$ and $k = 1$. Test for convergence at $x = \pm 1$.

ANS. Convergent.

5.2.17 A series solution for the ultraspherical (Gegenbauer) function $C_n^{\alpha}(x)$ leads to the recurrence

$$a_{j+2} = a_j \frac{(k+j)(k+j+2\alpha) - n(n+2\alpha)}{(k+j+1)(k+j+2)}.$$

Investigate the convergence of each of these series at $x = \pm 1$ as a function of the parameter α.

ANS. Convergent for $\alpha < 1$,
divergent for $\alpha \geq 1$.

5.2.18 A series expansion of the incomplete beta function (Section 8.4) yields

$$B_x(p,q) = x^p \left\{ \frac{1}{p} + \frac{1-q}{p+1} x + \frac{(1-q)(2-q)}{2!(p+2)} x^2 + \cdots \right.$$
$$\left. + \frac{(1-q)(2-q)\cdots(n-q)}{n!(p+n)} x^n + \cdots \right\}.$$

Given that $0 \leq x \leq 1$, $p > 0$, and $q > 0$, test this series for convergence. What happens at $x = 1$?

5.2.19 Show that the following series is convergent.

$$\sum_{s=0}^{\infty} \frac{(2s-1)!!}{(2s)!!(2s+1)}.$$

Note. $(2s-1)!! = (2s-1)(2s-3)\cdots 3\cdot 1$ with $(-1)!! = 1$; $(2s)!! = (2s)(2s-2)\cdots 4\cdot 2$ with $0!! = 1$. The series appears as a series expansion of $\sin^{-1}(1)$ and equals $\pi/2$, and $\sin^{-1} x \equiv \arcsin x \neq (\sin x)^{-1}$.

5.2.20 Show how to combine $\zeta(2) = \sum_{n=1}^{\infty} n^{-2}$ with α_1 and α_2 to obtain a series converging as n^{-4}.
Note. $\zeta(2)$ is known: $\zeta(2) = \pi^2/6$ (see Section 5.9).

5.2.21 The convergence improvement of Example 5.2.5 may be carried out more expediently (in this special case) by putting α_2 into a more symmetric form: Replacing n by $n-1$, we have

$$\alpha_2' = \sum_{n=2}^{\infty} \frac{1}{(n-1)n(n+1)} = \frac{1}{4}.$$

(a) Combine $\zeta(3)$ and α_2' to obtain convergence as n^{-5}.
(b) Let α_4' be α_4 with $n \to n-2$. Combine $\zeta(3)$, α_2', and α_4' to obtain convergence as n^{-7}.
(c) If $\zeta(3)$ is to be calculated to six = decimal = place accuracy (error 5×10^{-7}), how many terms are required for $\zeta(3)$ alone? combined as in part (a)? combined as in part (b)?

Note. The error may be estimated using the corresponding integral.

ANS. (a) $\zeta(3) = \dfrac{5}{4} - \displaystyle\sum_{n=2}^{\infty} \frac{1}{n^3(n^2-1)}$.

5.2.22 Catalan's constant ($\beta(2)$ of M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (AMS-55), Wash, D. C. National Bureau of Standards (1972); reprinted Dover (1974), Chapter 23) is defined by

$$\beta(2) = \sum_{k=0}^{\infty} (-1)^k (2k+1)^{-2} = \frac{1}{1^2} - \frac{1}{3^2} + \frac{1}{5^2} \cdots .$$

Calculate $\beta(2)$ to six-digit accuracy.
Hint. The rate of convergence is enhanced by pairing the terms:

$$(4k-1)^{-2} - (4k+1)^{-2} = \frac{16k}{(16k^2-1)^2}.$$

If you have carried enough digits in your series summation, $\sum_{1 \le k \le N} 16k/(16k^2-1)^2$, additional significant figures may be obtained by setting upper and lower bounds on the tail of the series, $\sum_{k=N+1}^{\infty}$. These bounds may be set by comparison with integrals, as in the Maclaurin integral test.

ANS. $\beta(2) = 0.9159\ 6559\ 4177 \ldots$.

5.3 ALTERNATING SERIES

In Section 5.2 we limited ourselves to series of positive terms. Now, in contrast, we consider infinite series in which the signs alternate. The partial cancellation due to alternating signs makes convergence more rapid and much easier to identify. We shall prove the Leibniz criterion, a general condition for the convergence of an alternating series. For series with more irregular sign changes, like Fourier series of Chapter 14 (see Example 5.3.1), the integral test of Eq. (5.25) is often helpful.

Leibniz Criterion

Consider the series $\sum_{n=1}^{\infty} (-1)^{n+1} a_n$ with $a_n > 0$. If a_n, is **monotonically decreasing** (for sufficiently large n) and $\lim_{n\to\infty} a_n = 0$, then the series converges. To prove this theorem,

we examine the even partial sums

$$\begin{aligned} s_{2n} &= a_1 - a_2 + a_3 - \cdots - a_{2n}, \\ s_{2n+2} &= s_{2n} + (a_{2n+1} - a_{2n+2}). \end{aligned} \tag{5.51}$$

Since $a_{2n+1} > a_{2n+2}$, we have

$$s_{2n+2} > s_{2n}. \tag{5.52}$$

On the other hand,

$$s_{2n+2} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \cdots - a_{2n+2}. \tag{5.53}$$

Hence, with each pair of terms $a_{2p} - a_{2p+1} > 0$,

$$s_{2n+2} < a_1. \tag{5.54}$$

With the even partial sums bounded $s_{2n} < s_{2n+2} < a_1$ and the terms a_n decreasing monotonically and approaching zero, this alternating series converges.

One further important result can be extracted from the partial sums of the same alternating series. From the difference between the series limit S and the partial sum s_n,

$$\begin{aligned} S - s_n &= a_{n+1} - a_{n+2} + a_{n+3} - a_{n+4} + \cdots \\ &= a_{n+1} - (a_{n+2} - a_{n+3}) - (a_{n+4} - a_{n+5}) - \cdots, \end{aligned} \tag{5.55}$$

or

$$S - s_n < a_{n+1}. \tag{5.56}$$

Equation (5.56) says that the error in cutting off an alternating series whose terms are monotonically decreasing after n terms is less than a_{n+1}, the first term dropped. A knowledge of the error obtained this way may be of great practical importance.

Absolute Convergence

Given a series of terms u_n in which u_n may vary in sign, if $\sum |u_n|$ converges, then $\sum u_n$ is said to be absolutely convergent. If $\sum u_n$ converges but $\sum |u_n|$ diverges, the convergence is called **conditional**.

The alternating harmonic series is a simple example of this conditional convergence. We have

$$\sum_{n=1}^{\infty} (-1)^{n-1} n^{-1} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{(-1)^{n-1}}{n} + \cdots, \tag{5.57}$$

convergent by the Leibniz criterion; but

$$\sum_{n=1}^{\infty} n^{-1} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + \cdots$$

has been shown to be divergent in Sections 5.1 and 5.2.

Note that most tests developed in Section 5.2 assume a series of positive terms. Therefore these tests in that section guarantee absolute convergence.

Example 5.3.1 SERIES WITH IRREGULAR SIGN CHANGES

For $0 < x < 2\pi$ the Fourier series (see Chapter 14.1)

$$\sum_{n=1}^{\infty} \frac{\cos(nx)}{n} = -\ln\left(2\sin\frac{x}{2}\right) \tag{5.58}$$

converges, having coefficients that change sign often, but not so that the Leibniz convergence criterion applies easily. Let us apply the integral test of Eq. (5.22). Using integration by parts we see immediately that

$$\int_1^{\infty} \frac{\cos(nx)}{n}\,dn = \left[\frac{\sin(nx)}{nx}\right]_1^{\infty} + \frac{1}{x}\int_{n=1}^{\infty} \frac{\sin(nx)}{n^2}\,dn$$

converges, and the integral on the right-hand side even converges absolutely. The derivative term in Eq. (5.22) has the form

$$\int_1^{\infty} \big(n - [n]\big)\left\{-\frac{x}{n}\sin(nx) - \frac{\cos(nx)}{n^2}\right\} dn,$$

where the second term converges absolutely and need not be considered further. Next we observe that $g(N) = \int_1^N (n - [n])\sin(nx)\,dn$ is bounded for $N \to \infty$, just as $\int^N \sin(nx)\,dn$ is bounded because of the periodic nature of $\sin(nx)$ and its regular sign changes. Using integration by parts again,

$$\int_1^{\infty} \frac{g'(n)}{n}\,dn = \left[\frac{g(n)}{n}\right]_{n=1}^{\infty} + \int_1^{\infty} \frac{g(n)}{n^2}\,dn,$$

we see that the second term is absolutely convergent and that the first goes to zero at the upper limit. Hence the series in Eq. (5.58) converges, which is hard to see from other convergence tests.

Alternatively, we may apply the $q = 1$ case of the Euler–Maclaurin integration formula in Eq. (5.168b),

$$\sum_{\nu=1}^{n} f(\nu) = \int_1^n f(x)\,dx + \frac{1}{2}\big\{f(n) + f(1)\big\} + \frac{1}{12}\big\{f'(n) - f'(1)\big\}$$
$$- \frac{1}{2}\int_0^1 \left(x^2 - x + \frac{1}{6}\right)\sum_{\nu=1}^{n-1} f''(x+\nu)\,dx,$$

which is straightforward but more tedious because of the second derivative. ■

Exercises

5.3.1 (a) From the electrostatic two-hemisphere problem (Exercise 12.3.20) we obtain the series

$$\sum_{s=0}^{\infty}(-1)^s(4s+3)\frac{(2s-1)!!}{(2s+2)!!}.$$

Test it for convergence.

(b) The corresponding series for the surface charge density is

$$\sum_{s=0}^{\infty}(-1)^s(4s+3)\frac{(2s-1)!!}{(2s)!!}.$$

Test it for convergence.

The !! notation is explained in Section 8.1 and Exercise 5.2.19.

5.3.2 Show by direct numerical computation that the sum of the first 10 terms of

$$\lim_{x\to 1}\ln(1+x)=\ln 2=\sum_{n=1}^{\infty}(-1)^{n-1}n^{-1}$$

differs from $\ln 2$ by less than the eleventh term: $\ln 2 = 0.69314\,71806\ldots$.

5.3.3 In Exercise 5.2.9 the hypergeometric series is shown convergent for $x=\pm 1$, if $\gamma > \alpha+\beta$. Show that there is conditional convergence for $x=-1$ for γ down to $\gamma > \alpha+\beta-1$.

Hint. The asymptotic behavior of the factorial function is given by Stirling's series, Section 8.3.

5.4 Algebra of Series

The establishment of absolute convergence is important because it can be proved that absolutely convergent series may be reordered according to the familiar rules of algebra or arithmetic.

- If an infinite series is absolutely convergent, the series sum is independent of the order in which the terms are added.
- The series may be multiplied with another absolutely convergent series. The limit of the product will be the product of the individual series limits. The product series, a double series, will also converge absolutely.

No such guarantees can be given for conditionally convergent series. Again consider the alternating harmonic series. If we write

$$1-\tfrac{1}{2}+\tfrac{1}{3}-\tfrac{1}{4}+\cdots=1-\left(\tfrac{1}{2}-\tfrac{1}{3}\right)-\left(\tfrac{1}{4}-\tfrac{1}{5}\right)-\cdots, \tag{5.59}$$

it is clear that the sum

$$\sum_{n=1}^{\infty}(-1)^{n-1}n^{-1} < 1. \tag{5.60}$$

However, if we rearrange the terms slightly, we may make the alternating harmonic series converge to $\frac{3}{2}$. We regroup the terms of Eq. (5.59), taking

$$\begin{aligned}&\left(1+\tfrac{1}{3}+\tfrac{1}{5}\right)-\left(\tfrac{1}{2}\right)+\left(\tfrac{1}{7}+\tfrac{1}{9}+\tfrac{1}{11}+\tfrac{1}{13}+\tfrac{1}{15}\right)-\left(\tfrac{1}{4}\right)\\&\quad+\left(\tfrac{1}{17}+\cdots+\tfrac{1}{25}\right)-\left(\tfrac{1}{6}\right)+\left(\tfrac{1}{27}+\cdots+\tfrac{1}{35}\right)-\left(\tfrac{1}{8}\right)+\cdots.\end{aligned} \tag{5.61}$$

Treating the terms grouped in parentheses as single terms for convenience, we obtain the partial sums

$$\begin{aligned}s_1 &= 1.5333 & s_2 &= 1.0333\\ s_3 &= 1.5218 & s_4 &= 1.2718\\ s_5 &= 1.5143 & s_6 &= 1.3476\\ s_7 &= 1.5103 & s_8 &= 1.3853\\ s_9 &= 1.5078 & s_{10} &= 1.4078.\end{aligned}$$

From this tabulation of s_n and the plot of s_n versus n in Fig. 5.3, the convergence to $\frac{3}{2}$ is fairly clear. We have rearranged the terms, taking positive terms until the partial sum was equal to or greater than $\frac{3}{2}$ and then adding in negative terms until the partial sum just fell below $\frac{3}{2}$ and so on. As the series extends to infinity, all original terms will eventually appear, but the partial sums of this rearranged alternating harmonic series converge to $\frac{3}{2}$.

By a suitable rearrangement of terms, a conditionally convergent series may be made to converge to any desired value or even to diverge. This statement is sometimes given

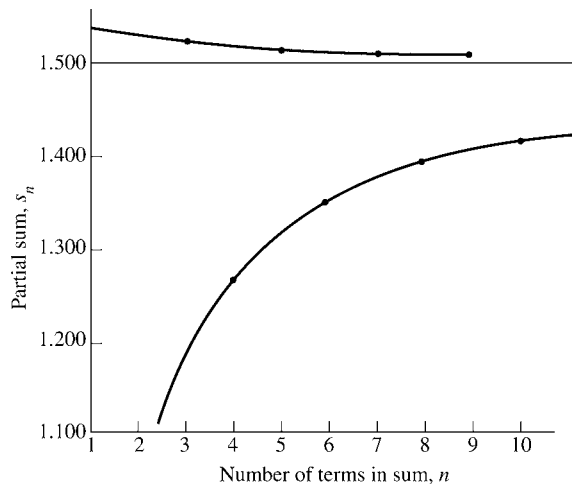

FIGURE 5.3 Alternating harmonic series — terms rearranged to give convergence to 1.5.

as **Riemann's theorem**. Obviously, conditionally convergent series must be treated with caution.

Absolutely convergent series can be multiplied without problems. This follows as a special case from the rearrangement of double series. However, conditionally convergent series cannot always be multiplied to yield convergent series, as the following example shows.

Example 5.4.1 SQUARE OF A CONDITIONALLY CONVERGENT SERIES MAY DIVERGE

The series $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{\sqrt{n}}$ converges, by the Leibniz criterion. Its square,

$$\left[\sum_n \frac{(-1)^{n-1}}{\sqrt{n}}\right]^2 = \sum_n (-1)^n \left[\frac{1}{\sqrt{1}}\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{2}}\frac{1}{\sqrt{n-2}} + \cdots + \frac{1}{\sqrt{n-1}}\frac{1}{\sqrt{1}}\right],$$

has the general term in brackets consisting of $n-1$ additive terms, each of which is greater than $\frac{1}{\sqrt{n-1}\sqrt{n-1}}$, so the product term in brackets is greater than $\frac{n-1}{n-1}$ and does not go to zero. Hence this product oscillates and therefore diverges. ■

Hence for a product of two series to converge, we have to demand as a sufficient condition that at least one of them converge absolutely. To prove this **product convergence theorem** that if $\sum_n u_n$ converges absolutely to U, $\sum_n v_n$ converges to V, then

$$\sum_n c_n, \qquad c_n = \sum_{m=0}^{n} u_m v_{n-m}$$

converges to UV, it is sufficient to show that the difference terms $D_n \equiv c_0 + c_1 + \cdots + c_{2n} - U_n V_n \to 0$ for $n \to \infty$, where U_n, V_n are the partial sums of our series. As a result, the partial sum differences

$$\begin{aligned} D_n &= u_0 v_0 + (u_0 v_1 + u_1 v_0) + \cdots + (u_0 v_{2n} + u_1 v_{2n-1} + \cdots + u_{2n} v_0) \\ &\quad - (u_0 + u_1 + \cdots + u_n)(v_0 + v_1 + \cdots + v_n) \\ &= u_0(v_{n+1} + \cdots + v_{2n}) + u_1(v_{n+1} + \cdots + v_{2n-1}) + \cdots + u_{n+1} v_{n+1} \\ &\quad + v_{n+1}(v_0 + \cdots + v_{n-1}) + \cdots + u_{2n} v_0, \end{aligned}$$

so for all sufficiently large n,

$$|D_n| < \epsilon\big(|u_0| + \cdots + |u_{n-1}|\big) + M\big(|u_{n+1}| + \cdots + |u_{2n}|\big) < \epsilon(a + M),$$

because $|v_{n+1} + v_{n+2} + \cdots + v_{n+m}| < \epsilon$ for sufficiently large n and all positive integers m as $\sum v_n$ converges, and the partial sums $V_n < B$ of $\sum_n v_n$ are bounded by M, because the sum converges. Finally we call $\sum_n |u_n| = a$, as $\sum u_n$ converges absolutely.

Two series can be multiplied, provided one of them converges absolutely. Addition and subtraction of series is also valid termwise if one series converges absolutely.

Improvement of Convergence, Rational Approximations

The series

$$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}, \qquad -1 < x \le 1, \tag{5.61a}$$

converges very slowly as x approaches $+1$. The **rate** of convergence may be improved substantially by multiplying both sides of Eq. (5.61a) by a polynomial and adjusting the polynomial coefficients to cancel the more slowly converging portions of the series. Consider the simplest possibility: Multiply $\ln(1+x)$ by $1 + a_1 x$:

$$(1 + a_1 x)\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} + a_1 \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^{n+1}}{n}.$$

Combining the two series on the right, term by term, we obtain

$$\begin{aligned}(1 + a_1 x)\ln(1+x) &= x + \sum_{n=2}^{\infty} (-1)^{n-1} \left(\frac{1}{n} - \frac{a_1}{n-1} \right) x^n \\ &= x + \sum_{n=2}^{\infty} (-1)^{n-1} \frac{n(1-a_1) - 1}{n(n-1)} x^n.\end{aligned}$$

Clearly, if we take $a_1 = 1$, the n in the numerator disappears and our combined series converges as n^{-2}.

Continuing this process, we find that $(1 + 2x + x^2)\ln(1+x)$ vanishes as n^{-3} and that $(1 + 3x + 3x^2 + x^3)\ln(1+x)$ vanishes as n^{-4}. In effect we are shifting from a simple series expansion of Eq. (5.61a) to a rational fraction representation in which the function $\ln(1+x)$ is represented by the ratio of a series and a polynomial:

$$\ln(1+x) = \frac{x + \sum_{n=2}^{\infty} (-1)^n x^n / [n(n-1)]}{1+x}.$$

Such rational approximations may be both compact and accurate.

Rearrangement of Double Series

Another aspect of the rearrangement of series appears in the treatment of double series (Fig. 5.4):

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{n,m}.$$

Let us substitute

$$n = q \ge 0, \qquad m = p - q \ge 0 \qquad (q \le p).$$

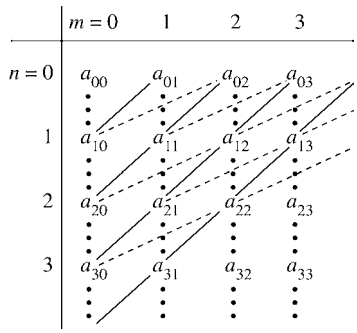

FIGURE 5.4 Double series — summation over n indicated by vertical dashed lines.

This results in the identity

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{n,m} = \sum_{p=0}^{\infty}\sum_{q=0}^{p} a_{q,p-q}. \tag{5.62}$$

The summation over p and q of Eq. (5.62) is illustrated in Fig. 5.5. The substitution

$$n = s \geq 0, \qquad m = r - 2s \geq 0 \qquad \left(s \leq \frac{r}{2}\right)$$

leads to

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{n,m} = \sum_{r=0}^{\infty}\sum_{s=0}^{[r/2]} a_{s,r-2s}, \tag{5.63}$$

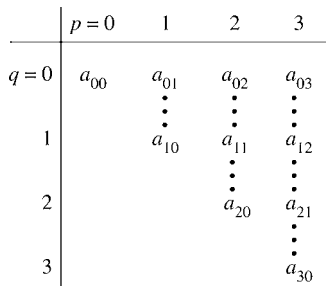

FIGURE 5.5 Double series — again, the first summation is represented by vertical dashed lines, but these vertical lines correspond to diagonals in Fig. 5.4.

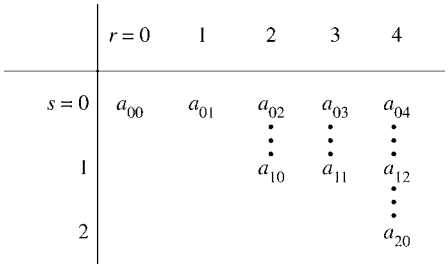

FIGURE 5.6 Double series. The summation over s corresponds to a summation along the almost-horizontal dashed lines in Fig. 5.4.

with $[r/2] = r/2$ for r even and $(r-1)/2$ for r odd. The summation over r and s of Eq. (5.63) is shown in Fig. 5.6. Equations (5.62) and (5.63) are clearly rearrangements of the array of coefficients a_{nm}, rearrangements that are valid as long as we have absolute convergence.

The combination of Eqs. (5.62) and (5.63),

$$\sum_{p=0}^{\infty}\sum_{q=0}^{p} a_{q,p-q} = \sum_{r=0}^{\infty}\sum_{s=0}^{[r/2]} a_{s,r-2s}, \tag{5.64}$$

is used in Section 12.1 in the determination of the series form of the Legendre polynomials.

Exercises

5.4.1 Given the series (derived in Section 5.6)

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4}\cdots, \qquad -1 < x \le 1,$$

show that

$$\ln\left(\frac{1+x}{1-x}\right) = 2\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots\right), \qquad -1 < x < 1.$$

The original series, $\ln(1+x)$, appears in an analysis of binding energy in crystals. It is $\frac{1}{2}$ the Madelung constant $(2\ln 2)$ for a chain of atoms. The second series is useful in normalizing the Legendre polynomials (Section 12.3) and in developing a second solution for Legendre's differential equation (Section 12.10).

5.4.2 Determine the values of the coefficients a_1, a_2, and a_3 that will make $(1 + a_1x + a_2x^2 + a_3x^3)\ln(1+x)$ converge as n^{-4}. Find the resulting series.

5.4.3 Show that

$$\text{(a)} \sum_{n=2}^{\infty}\big[\zeta(n) - 1\big] = 1, \qquad \text{(b)} \sum_{n=2}^{\infty}(-1)^n\big[\zeta(n) - 1\big] = \tfrac{1}{2},$$

where $\zeta(n)$ is the Riemann zeta function.

5.4.4 Write a program that will rearrange the terms of the alternating harmonic series to make the series converge to 1.5. Group your terms as indicated in Eq. (5.61). List the first 100 successive partial sums that just climb above 1.5 or just drop below 1.5, and list the new terms included in each such partial sum.

ANS.

n	1	2	3	4	5
s_n	1.5333	1.0333	1.5218	1.2718	1.5143

5.5 SERIES OF FUNCTIONS

We extend our concept of infinite series to include the possibility that each term u_n may be a function of some variable, $u_n = u_n(x)$. Numerous illustrations of such series of functions appear in Chapters 11–14. The partial sums become functions of the variable x,

$$s_n(x) = u_1(x) + u_2(x) + \cdots + u_n(x), \tag{5.65}$$

as does the series sum, defined as the limit of the partial sums:

$$\sum_{n=1}^{\infty} u_n(x) = S(x) = \lim_{n\to\infty} s_n(x). \tag{5.66}$$

So far we have concerned ourselves with the behavior of the partial sums as a function of n. Now we consider how the foregoing quantities depend on x. The key concept here is that of uniform convergence.

Uniform Convergence

If for any small $\varepsilon > 0$ there exists a number N, **independent of** x in the interval $[a, b]$ (that is, $a \le x \le b$) such that

$$\left|S(x) - s_n(x)\right| < \varepsilon, \qquad \text{for all} \quad n \ge N, \tag{5.67}$$

then the series is said to be uniformly convergent in the interval $[a, b]$. This says that for our series to be uniformly convergent, it must be possible to find a finite N so that the tail of the infinite series, $|\sum_{i=N+1}^{\infty} u_i(x)|$, will be less than an arbitrarily small ε for all x in the given interval.

This condition, Eq. (5.67), which defines uniform convergence, is illustrated in Fig. 5.7. The point is that no matter how small ε is taken to be, we can always choose n large enough so that the absolute magnitude of the difference between $S(x)$ and $s_n(x)$ is less than ε for all x, $a \le x \le b$. If this cannot be done, then $\sum u_n(x)$ is not uniformly convergent in $[a, b]$.

Example 5.5.1 NONUNIFORM CONVERGENCE

$$\sum_{n=1}^{\infty} u_n(x) = \sum_{n=1}^{\infty} \frac{x}{[(n-1)x + 1][nx + 1]}. \tag{5.68}$$

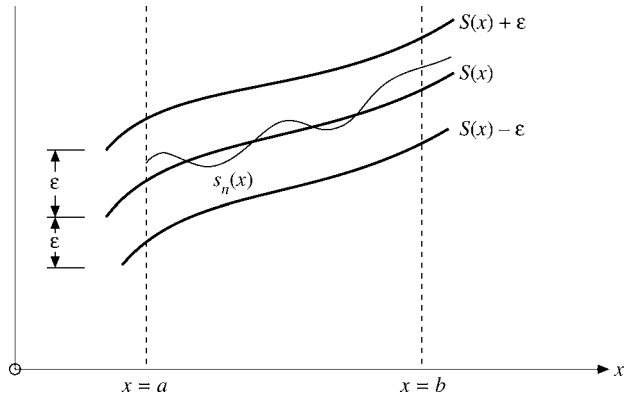

FIGURE 5.7 Uniform convergence.

The partial sum $s_n(x) = nx(nx+1)^{-1}$, as may be verified by **mathematical induction**. By inspection this expression for $s_n(x)$ holds for $n = 1, 2$. We assume it holds for n terms and then prove it holds for $n+1$ terms:

$$\begin{aligned} s_{n+1}(x) &= s_n(x) + \frac{x}{[nx+1][(n+1)x+1]} \\ &= \frac{nx}{[nx+1]} + \frac{x}{[nx+1][(n+1)x+1]} \\ &= \frac{(n+1)x}{(n+1)x+1}, \end{aligned}$$

completing the proof.

Letting n approach infinity, we obtain

$$\begin{aligned} S(0) &= \lim_{n\to\infty} s_n(0) = 0, \\ S(x \neq 0) &= \lim_{n\to\infty} s_n(x \neq 0) = 1. \end{aligned}$$

We have a discontinuity in our series limit at $x = 0$. However, $s_n(x)$ is a continuous function of $x, 0 \leq x \leq 1$, for all finite n. No matter how small ε may be, Eq. (5.67) will be violated for all sufficiently small x. Our series does not converge uniformly. ■

Weierstrass *M* (Majorant) Test

The most commonly encountered test for uniform convergence is the Weierstrass M test. If we can construct a series of numbers $\sum_1^\infty M_i$, in which $M_i \geq |u_i(x)|$ for all x in the interval $[a, b]$ and $\sum_1^\infty M_i$ is convergent, our series $u_i(x)$ will be **uniformly** convergent in $[a, b]$.

The proof of this Weierstrass M test is direct and simple. Since $\sum_i M_i$ converges, some number N exists such that for $n+1 \geq N$,

$$\sum_{i=n+1}^{\infty} M_i < \varepsilon. \tag{5.69}$$

This follows from our definition of convergence. Then, with $|u_i(x)| \leq M_i$ for all x in the interval $a \leq x \leq b$,

$$\sum_{i=n+1}^{\infty} \left|u_i(x)\right| < \varepsilon. \tag{5.70}$$

Hence

$$\left|S(x) - s_n(x)\right| = \left|\sum_{i=n+1}^{\infty} u_i(x)\right| < \varepsilon, \tag{5.71}$$

and by definition $\sum_{i=1}^{\infty} u_i(x)$ is uniformly convergent in $[a, b]$. Since we have specified absolute values in the statement of the Weierstrass M test, the series $\sum_{i=1}^{\infty} u_i(x)$ is also seen to be **absolutely** convergent.

Note that uniform convergence and absolute convergence are independent properties. Neither implies the other. For specific examples,

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n + x^2}, \qquad -\infty < x < \infty, \tag{5.72}$$

and

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} = \ln(1 + x), \qquad 0 \leq x \leq 1, \tag{5.73}$$

converge uniformly in the indicated intervals but do not converge absolutely. On the other hand,

$$\begin{aligned} \sum_{n=0}^{\infty} (1 - x)x^n &= 1, \qquad 0 \leq x < 1 \\ &= 0, \qquad x = 1, \end{aligned} \tag{5.74}$$

converges absolutely but does not converge uniformly in $[0, 1]$.

From the definition of uniform convergence we may show that any series

$$f(x) = \sum_{n=1}^{\infty} u_n(x) \tag{5.75}$$

cannot converge uniformly in any interval that includes a discontinuity of $f(x)$ if all $u_n(x)$ are continuous.

Since the Weierstrass M test establishes both uniform and absolute convergence, it will necessarily fail for series that are uniformly but conditionally convergent.

Abel's Test

A somewhat more delicate test for uniform convergence has been given by Abel. If

$$u_n(x) = a_n f_n(x),$$

$$\sum a_n = A, \quad \text{convergent}$$

and the functions $f_n(x)$ are monotonic $[f_{n+1}(x) \le f_n(x)]$ and bounded, $0 \le f_n(x) \le M$, for all x in $[a, b]$, then $\sum_n u_n(x)$ **converges uniformly** in $[a, b]$.

This test is especially useful in analyzing power series (compare Section 5.7). Details of the proof of Abel's test and other tests for uniform convergence are given in the Additional Readings listed at the end of this chapter.

Uniformly convergent series have three particularly useful properties.

1. If the individual terms $u_n(x)$ are continuous, the series sum

$$f(x) = \sum_{n=1}^{\infty} u_n(x) \tag{5.76}$$

is also continuous.

2. If the individual terms $u_n(x)$ are continuous, the series may be integrated term by term. The sum of the integrals is equal to the integral of the sum.

$$\int_a^b f(x)\,dx = \sum_{n=1}^{\infty} \int_a^b u_n(x)\,dx. \tag{5.77}$$

3. The derivative of the series sum $f(x)$ equals the sum of the individual term derivatives:

$$\frac{d}{dx} f(x) = \sum_{n=1}^{\infty} \frac{d}{dx} u_n(x), \tag{5.78}$$

provided the following conditions are satisfied:

$$u_n(x) \text{ and } \frac{du_n(x)}{dx} \text{ are continuous in } [a, b].$$

$$\sum_{n=1}^{\infty} \frac{du_n(x)}{dx} \text{ is uniformly convergent in } [a, b].$$

Term-by-term integration of a uniformly convergent series[8] requires only continuity of the individual terms. This condition is almost always satisfied in physical applications. Term-by-term differentiation of a series is often not valid because more restrictive conditions must be satisfied. Indeed, we shall encounter Fourier series in Chapter 14 in which term-by-term differentiation of a uniformly convergent series leads to a divergent series.

[8] Term-by-term integration may also be valid in the absence of uniform convergence.

Exercises

5.5.1 Find the range of **uniform** convergence of the Dirichlet series

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^x}$, (b) $\zeta(x) = \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^x}$.

ANS. (a) $0 < s \le x < \infty$.
(b) $1 < s \le x < \infty$.

5.5.2 For what range of x is the geometric series $\sum_{n=0}^{\infty} x^n$ uniformly convergent?

ANS. $-1 < -s \le x \le s < 1$.

5.5.3 For what range of positive values of x is $\sum_{n=0}^{\infty} 1/(1+x^n)$

(a) convergent? (b) uniformly convergent?

5.5.4 If the series of the coefficients $\sum a_n$ and $\sum b_n$ are absolutely convergent, show that the Fourier series

$$\sum (a_n \cos nx + b_n \sin nx)$$

is **uniformly** convergent for $-\infty < x < \infty$.

5.6 TAYLOR'S EXPANSION

This is an expansion of a function into an infinite series of powers of a variable x or into a finite series plus a remainder term. The coefficients of the successive terms of the series involve the successive derivatives of the function. We have already used Taylor's expansion in the establishment of a physical interpretation of divergence (Section 1.7) and in other sections of Chapters 1 and 2. Now we derive the Taylor expansion.

We assume that our function $f(x)$ has a continuous nth derivative[9] in the interval $a \le x \le b$. Then, integrating this nth derivative n times,

$$\int_a^x f^{(n)}(x_1)\,dx_1 = f^{(n-1)}(x_1)\Big|_a^x = f^{(n-1)}(x) - f^{(n-1)}(a),$$

$$\int_a^x dx_2 \int_a^{x_2} dx_1 f^{(n)}(x_1) = \int_a^x dx_2 \left[f^{(n-1)}(x_2) - f^{(n-1)}(a)\right] \tag{5.79}$$

$$= f^{(n-2)}(x) - f^{(n-2)}(a) - (x-a) f^{(n-1)}(a).$$

Continuing, we obtain

$$\int_a^x dx_3 \int_a^{x_3} dx_2 \int_a^{x_2} dx_1 f^{(n)}(x_1) = f^{(n-3)}(x) - f^{(n-3)}(a) - (x-a) f^{(n-2)}(a)$$

$$- \frac{(x-a)^2}{2!} f^{(n-1)}(a). \tag{5.80}$$

[9] Taylor's expansion may be derived under slightly less restrictive conditions; compare H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, 3rd ed. Cambridge: Cambridge University Press (1956), Section 1.133.

Finally, on integrating for the nth time,

$$\int_a^x dx_n \cdots \int_a^{x_2} dx_1 f^{(n)}(x_1) = f(x) - f(a) - (x-a)f'(a) - \frac{(x-a)^2}{2!}f''(a) - \cdots - \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a). \tag{5.81}$$

Note that this expression is exact. No terms have been dropped, no approximations made. Now, solving for $f(x)$, we have

$$\boxed{f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + R_n.} \tag{5.82}$$

The remainder, R_n, is given by the n-fold integral

$$R_n = \int_a^x dx_n \cdots \int_a^{x_2} dx_1 f^{(n)}(x_1). \tag{5.83}$$

This remainder, Eq. (5.83), may be put into a perhaps more practical form by using the **mean value theorem** of integral calculus:

$$\int_a^x g(x)\,dx = (x-a)g(\xi), \tag{5.84}$$

with $a \le \xi \le x$. By integrating n times we get the Lagrangian form[10] of the remainder:

$$R_n = \frac{(x-a)^n}{n!}f^{(n)}(\xi). \tag{5.85}$$

With Taylor's expansion in this form we are not concerned with any questions of infinite series convergence. This series is finite, and the only questions concern the magnitude of the remainder.

When the function $f(x)$ is such that

$$\lim_{n\to\infty} R_n = 0, \tag{5.86}$$

Eq. (5.82) becomes Taylor's series:

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \cdots = \sum_{n=0}^{\infty} \frac{(x-a)^n}{n!}f^{(n)}(a).^{11} \tag{5.87}$$

[10] An alternate form derived by Cauchy is

$$R_n = \frac{(x-\zeta)^{n-1}(x-a)}{(n-1)!}f^{(n)}(\zeta),$$

with $a \le \zeta \le x$.

[11] Note that $0! = 1$ (compare Section 8.1).

Our Taylor series specifies the value of a function at one point, x, in terms of the value of the function and its derivatives at a reference point a. It is an expansion in powers of the **change** in the variable, $\Delta x = x - a$ in this case. The notation may be varied at the user's convenience. With the substitution $x \to x + h$ and $a \to x$ we have an alternate form,

$$f(x+h) = \sum_{n=0}^{\infty} \frac{h^n}{n!} f^{(n)}(x).$$

When we use the **operator** $D = d/dx$, the Taylor expansion becomes

$$f(x+h) = \sum_{n=0}^{\infty} \frac{h^n D^n}{n!} f(x) = e^{hD} f(x).$$

(The transition to the exponential form anticipates Eq. (5.90), which follows.) An equivalent operator form of this Taylor expansion appears in Exercise 4.2.4. A derivation of the Taylor expansion in the context of complex variable theory appears in Section 6.5.

Maclaurin Theorem

If we expand about the origin ($a = 0$), Eq. (5.87) is known as Maclaurin's series:

$$\boxed{f(x) = f(0) + xf'(0) + \frac{x^2}{2!} f''(0) + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!} f^{(n)}(0).} \tag{5.88}$$

An immediate application of the Maclaurin series (or the Taylor series) is in the expansion of various transcendental functions into infinite (power) series.

Example 5.6.1 EXPONENTIAL FUNCTION

Let $f(x) = e^x$. Differentiating, we have

$$f^{(n)}(0) = 1 \tag{5.89}$$

for all $n, n = 1, 2, 3, \ldots$. Then, with Eq. (5.88), we have

$$\boxed{e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.} \tag{5.90}$$

This is the series expansion of the exponential function. Some authors use this series to define the exponential function.

Although this series is clearly convergent for all x, we should check the remainder term, R_n. By Eq. (5.85) we have

$$R_n = \frac{x^n}{n!} f^{(n)}(\xi) = \frac{x^n}{n!} e^{\xi}, \qquad 0 \le |\xi| \le x. \tag{5.91}$$

Therefore

$$|R_n| \le \frac{x^n e^x}{n!} \tag{5.92}$$

and

$$\lim_{n\to\infty} R_n = 0 \tag{5.93}$$

for all **finite** values of x, which indicates that this Maclaurin expansion of e^x converges absolutely over the range $-\infty < x < \infty$. ■

Example 5.6.2 LOGARITHM

Let $f(x) = \ln(1+x)$. By differentiating, we obtain

$$\begin{aligned} f'(x) &= (1+x)^{-1}, \\ f^{(n)}(x) &= (-1)^{n-1}(n-1)!(1+x)^{-n}. \end{aligned} \tag{5.94}$$

The Maclaurin expansion (Eq. (5.88)) yields

$$\begin{aligned} \ln(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + R_n \\ &= \sum_{p=1}^{n} (-1)^{p-1} \frac{x^p}{p} + R_n. \end{aligned} \tag{5.95}$$

In this case our remainder is given by

$$\begin{aligned} R_n &= \frac{x^n}{n!} f^{(n)}(\xi), \qquad 0 \le \xi \le x \\ &\le \frac{x^n}{n}, \qquad 0 \le \xi \le x \le 1. \end{aligned} \tag{5.96}$$

Now, the remainder approaches zero as n is increased indefinitely, provided $0 \le x \le 1$.[12] As an infinite series,

$$\boxed{\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}} \tag{5.97}$$

converges for $-1 < x \le 1$. The range $-1 < x < 1$ is easily established by the d'Alembert ratio test (Section 5.2). Convergence at $x = 1$ follows by the Leibniz criterion (Section 5.3). In particular, at $x = 1$ we have

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots = \sum_{n=1}^{\infty} (-1)^{n-1} n^{-1}, \tag{5.98}$$

the conditionally convergent alternating harmonic series. ■

[12] This range can easily be extended to $-1 < x \le 1$ but not to $x = -1$.

Binomial Theorem

A second, extremely important application of the Taylor and Maclaurin expansions is the derivation of the binomial theorem for negative and/or nonintegral powers.

Let $f(x) = (1+x)^m$, in which m may be negative and is not limited to integral values. Direct application of Eq. (5.88) gives

$$(1+x)^m = 1 + mx + \frac{m(m-1)}{2!}x^2 + \cdots + R_n. \tag{5.99}$$

For this function the remainder is

$$R_n = \frac{x^n}{n!}(1+\xi)^{m-n}m(m-1)\cdots(m-n+1) \tag{5.100}$$

and ξ lies between 0 and x, $0 \le \xi \le x$. Now, for $n > m$, $(1+\xi)^{m-n}$ is a maximum for $\xi = 0$. Therefore

$$R_n \le \frac{x^n}{n!}m(m-1)\cdots(m-n+1). \tag{5.101}$$

Note that the m dependent factors do not yield a zero unless m is a nonnegative integer; R_n tends to zero as $n \to \infty$ if x is restricted to the range $0 \le x < 1$. The binomial expansion therefore is shown to be

$$(1+x)^m = 1 + mx + \frac{m(m-1)}{2!}x^2 + \frac{m(m-1)(m-2)}{3!}x^3 + \cdots. \tag{5.102}$$

In other, equivalent notation,

$$\boxed{(1+x)^m = \sum_{n=0}^{\infty}\frac{m!}{n!(m-n)!}x^n = \sum_{n=0}^{\infty}\binom{m}{n}x^n.} \tag{5.103}$$

The quantity $\binom{m}{n}$, which equals $m!/[n!(m-n)!]$, is called a **binomial coefficient**. Although we have only shown that the remainder vanishes,

$$\lim_{n\to\infty} R_n = 0,$$

for $0 \le x < 1$, the series in Eq. (5.102) actually may be shown to be convergent for the extended range $-1 < x < 1$. For m an integer, $(m-n)! = \pm\infty$ if $n > m$ (Section 8.1) and the series automatically terminates at $n = m$.

Example 5.6.3 RELATIVISTIC ENERGY

The total relativistic energy of a particle of mass m and velocity v is

$$E = mc^2\left(1 - \frac{v^2}{c^2}\right)^{-1/2}. \tag{5.104}$$

Compare this expression with the classical kinetic energy, $mv^2/2$.

By Eq. (5.102) with $x = -v^2/c^2$ and $m = -1/2$ we have

$$E = mc^2\left[1 - \frac{1}{2}\left(-\frac{v^2}{c^2}\right) + \frac{(-1/2)(-3/2)}{2!}\left(-\frac{v^2}{c^2}\right)^2 + \frac{(-1/2)(-3/2)(-5/2)}{3!}\left(-\frac{v^2}{c^2}\right)^3 + \cdots\right],$$

or

$$E = mc^2 + \frac{1}{2}mv^2 + \frac{3}{8}mv^2 \cdot \frac{v^2}{c^2} + \frac{5}{16}mv^2 \cdot \left(\frac{v^2}{c^2}\right)^2 + \cdots. \tag{5.105}$$

The first term, mc^2, is identified as the rest mass energy. Then

$$E_{\text{kinetic}} = \frac{1}{2}mv^2\left[1 + \frac{3}{4}\frac{v^2}{c^2} + \frac{5}{8}\left(\frac{v^2}{c^2}\right)^2 + \cdots\right]. \tag{5.106}$$

For particle velocity $v \ll c$, the velocity of light, the expression in the brackets reduces to unity and we see that the kinetic portion of the total relativistic energy agrees with the classical result. ■

For polynomials we can generalize the binomial expansion to

$$(a_1 + a_2 + \cdots + a_m)^n = \sum \frac{n!}{n_1!n_2!\cdots n_m!} a_1^{n_1} a_2^{n_2} \cdots a_m^{n_m},$$

where the summation includes all different combinations of $n_1, n_2, \ldots, n_m$ with $\sum_{i=1}^{m} n_i = n$. Here n_i and n are all integral. This generalization finds considerable use in statistical mechanics.

Maclaurin series may sometimes appear indirectly rather than by direct use of Eq. (5.88). For instance, the most convenient way to obtain the series expansion

$$\sin^{-1} x = \sum_{n=0}^{\infty} \frac{(2n-1)!!}{(2n)!!} \cdot \frac{x^{2n+1}}{(2n+1)} = x + \frac{x^3}{6} + \frac{3x^5}{40} + \cdots, \tag{5.106a}$$

is to make use of the relation (from $\sin y = x$, get $dy/dx = 1/\sqrt{1-x^2}$)

$$\sin^{-1} x = \int_0^x \frac{dt}{(1-t^2)^{1/2}}.$$

We expand $(1-t^2)^{-1/2}$ (binomial theorem) and then integrate term by term. This term-by-term integration is discussed in Section 5.7. The result is Eq. (5.106a). Finally, we may take the limit as $x \to 1$. The series converges by Gauss' test, Exercise 5.2.5.

Taylor Expansion — More Than One Variable

If the function f has more than one independent variable, say, $f = f(x, y)$, the Taylor expansion becomes

$$\begin{aligned} f(x,y) = f(a,b) &+ (x-a)\frac{\partial f}{\partial x} + (y-b)\frac{\partial f}{\partial y} \\ &+ \frac{1}{2!}\left[(x-a)^2\frac{\partial^2 f}{\partial x^2} + 2(x-a)(y-b)\frac{\partial^2 f}{\partial x\,\partial y} + (y-b)^2\frac{\partial^2 f}{\partial y^2}\right] \\ &+ \frac{1}{3!}\left[(x-a)^3\frac{\partial^3 f}{\partial x^3} + 3(x-a)^2(y-b)\frac{\partial^3 f}{\partial x^2\,\partial y}\right. \\ &\left. + 3(x-a)(y-b)^2\frac{\partial^3 f}{\partial x\,\partial y^2} + (y-b)^3\frac{\partial^3 f}{\partial y^3}\right] + \cdots, \end{aligned} \tag{5.107}$$

with all derivatives evaluated at the point (a, b). Using $\alpha_j t = x_j - x_{j0}$, we may write the Taylor expansion for m independent variables in the symbolic form

$$f(x_1, \ldots, x_m) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\left(\sum_{i=1}^{m} \alpha_i \frac{\partial}{\partial x_i}\right)^n f(x_1, \ldots, x_m)\Big|_{(x_k = x_{k0}, k=1,\ldots,m)}. \tag{5.108}$$

A convenient vector form for $m = 3$ is

$$\psi(\mathbf{r} + \mathbf{a}) = \sum_{n=0}^{\infty} \frac{1}{n!}(\mathbf{a} \cdot \nabla)^n \psi(\mathbf{r}). \tag{5.109}$$

Exercises

5.6.1 Show that

(a) $\sin x = \displaystyle\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$,

(b) $\cos x = \displaystyle\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$.

In Section 6.1, e^{ix} is defined by a series expansion such that

$$e^{ix} = \cos x + i \sin x.$$

This is the basis for the polar representation of complex quantities. As a special case we find, with $x = \pi$, the intriguing relation

$$e^{i\pi} = -1.$$