# Appendix A
# Vector Calculus

Vector calculus is an indispensable mathematical tool for classical mechanics. It provides a geometric (i.e., coordinate-independent) framework for formulating the laws of mechanics, and for solving for the motion of a particle, or a system of particles, subject to external forces. In this appendix, we summarize several key results of differential and integral vector calculus, which are used repeatedly throughout the text. For a more detailed introduction to these topics, including proofs, we recommend, e.g., Schey (1996), Boas (2006), and Griffiths (1999), on which we've based our discussion. Since the majority of the calculations in classical mechanics involve working with ordinary three-dimensional spatial vectors, e.g., position, velocity, acceleration, etc., we focus attention on such vectors in this appendix. Extensions to four-dimensional vectors, which arise in the context of relativistic mechanics, are discussed in Chap. 11, and the calculus of *differential forms* is discussed in Appendix B.

## A.1 Vector Algebra

In addition to adding two vectors, $\mathbf{A}+\mathbf{B}$, and multiplying a vector by a scalar, $a\mathbf{A}$, we can form various products of vectors: (i) The **dot product** (also called the **scalar product** or **inner product**) of two vectors $\mathbf{A}$ and $\mathbf{B}$ is defined by

$$\mathbf{A}\cdot\mathbf{B}\equiv AB\,\cos\theta\,, \tag{A.1}$$

where $A\equiv|\mathbf{A}|$, $B\equiv|\mathbf{B}|$ are the magnitudes (or norms) of $\mathbf{A}$, $\mathbf{B}$, and $\theta$ is the angle between the two vectors. (ii) The **cross product** (also called the **vector product** or **exterior product**) of $\mathbf{A}$ and $\mathbf{B}$ is defined by

$$\mathbf{A}\times\mathbf{B}\equiv AB\,\sin\theta\,\hat{\mathbf{n}}\,, \tag{A.2}$$


M.J. Benacquista and J.D. Romano, *Classical Mechanics*, Undergraduate Lecture Notes in Physics, https://doi.org/10.1007/978-3-319-68780-3

where $\theta$ is as before (assumed to be between 0° and 180°), and $\hat{\mathbf{n}}$ is a unit vector perpendicular to the plane spanned by $\mathbf{A}$ and $\mathbf{B}$, whose direction is given by the right-hand rule.[1] Note that if $\mathbf{A}$ and $\mathbf{B}$ are parallel, then $\mathbf{A} \times \mathbf{B} = \mathbf{0}$, while if $\mathbf{A}$ and $\mathbf{B}$ are perpendicular, $\mathbf{A} \cdot \mathbf{B} = 0$.

The dot product and cross product of $\mathbf{A}$ and $\mathbf{B}$ can also be written rather simply in terms of the components $A_i$, $B_i$ $(i = 1, 2, 3)$ of $\mathbf{A}$ and $\mathbf{B}$ with respect to an orthonormal basis $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$. (See Appendix D.2.1 for a general review about decomposing a vector into its components with respect to a basis.) Expanding $\mathbf{A}$ and $\mathbf{B}$ as

$$\mathbf{A} = \sum_i A_i \,\hat{\mathbf{e}}_i\,, \qquad \mathbf{B} = \sum_i B_i \,\hat{\mathbf{e}}_i\,, \tag{A.3}$$

it follows that

$$\mathbf{A} \cdot \mathbf{B} = \sum_{i,j} \delta_{ij} A_i B_j = \sum_i A_i B_i \tag{A.4}$$

and

$$(\mathbf{A} \times \mathbf{B})_i = \sum_{j,k} \varepsilon_{ijk} A_j B_k\,, \tag{A.5}$$

where

$$\delta_{ij} \equiv \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{A.6}$$

is the **Kronecker delta**, and[2]

$$\varepsilon_{ijk} \equiv \begin{cases} 1 & \text{if } ijk \text{ is an even permutation of 123} \\ -1 & \text{if } ijk \text{ is an odd permutation of 123} \\ 0 & \text{otherwise} \end{cases} \tag{A.7}$$

is the **Levi-Civita symbol**. We note that the above component expressions for dot product and cross product are valid with respect to *any* orthonormal basis, and not just for Cartesian coordinates.

Geometrically, the dot product of two vectors is the projection of one vector onto the direction of the other vector, times the magnitude of the other vector. Thus, by

---

[1] Point the fingers of your right hand in the direction of $\mathbf{A}$, and then curl them toward your palm in the direction of $\mathbf{B}$. Your thumb then points in the direction of $\hat{\mathbf{n}}$.

[2] An odd (even) permutation of 123 corresponds to an odd (even) number of interchanges of two of the numbers. For example, 213 is an odd permutation of 123, while 231 is an even permutation.

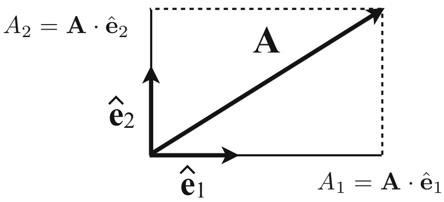



**Fig. A.1** Components of a 2-dimensional vector **A** in terms of dot products with the orthonormal basis vectors $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2$

taking the dot product of **A** with the orthonormal basis vectors $\hat{\mathbf{e}}_i$, we obtain the components of **A** with respect to this basis, i.e., $A_i = \mathbf{A} \cdot \hat{\mathbf{e}}_i$. This is shown in Fig. A.1.

**Exercise A.1** Prove that the geometric and component expressions for both the dot product, (A.1) and (A.4), and cross product, (A.2) and (A.5), are equivalent to one another, choosing a convenient coordinate system to do the calculation.

A key identity relating the Kronecker delta and Levi-Civita symbol is

$$\sum_i \varepsilon_{ijk}\varepsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl} \,. \tag{A.8}$$

Using this identity and the component forms of the dot product and cross product, one can prove the following three results:
Scalar triple product:

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}) \tag{A.9}$$

Vector triple product:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}) \tag{A.10}$$

Jacobi identity:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) + \mathbf{B} \times (\mathbf{C} \times \mathbf{A}) + \mathbf{C} \times (\mathbf{A} \times \mathbf{B}) = 0 \tag{A.11}$$

**Exercise A.2** Prove the above three identities.

## A.2 Vector Component and Coordinate Notation

Before proceeding further, we should comment on the index notation that we'll be using throughout this book.

### A.2.1 *Contravariant and Covariant Vectors*

In general, one should distinguish between vectors with components $A^i$ (so-called **contravariant vectors**) and vectors with components $A_i$ (so-called **covariant vectors** or **dual vectors**). By definition these components transform *inversely* to one another under a change of basis or coordinate system. For example, if $A^i$ and $A_i$ denote the components of a contravariant and covariant vector with respect to a coordinate basis (See Appendix A.4.1), then under a coordinate transformation $x^i \to x^{i'} = x^{i'}(x^i)$:

$$A^{i'} = \sum_i \frac{\partial x^{i'}}{\partial x^i} A^i \,, \qquad A_{i'} = \sum_i \frac{\partial x^i}{\partial x^{i'}} A_i \,. \tag{A.12}$$

But since most of the calculations that we will perform involve quantities in ordinary 3-dimensional Euclidean space with components defined with respect to an *orthonormal* basis, then

$$\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij} = \mathrm{diag}(1, 1, 1) \,, \tag{A.13}$$

and the two sets of components $A^i$ and $A_i$ can be mapped to one another using the Kronecker delta:

$$A_i = \sum_j \delta_{ij} A^j \quad \Leftrightarrow \quad A_1 = A^1 \,, \quad A_2 = A^2 \,, \quad A_3 = A^3 \,. \tag{A.14}$$

Hence, $A_i = A^i$, so it doesn't matter where we place the index. For simplicity of notation, we will typically use the subscript notation, which is the standard notation in the classical mechanics literature.

This equality between covariant and contravariant components will not hold, however, when we discuss spacetime 4-vectors in the context of special relativity (Chap. 11). This is because a set of orthonormal spacetime basis vectors satisfies (See Sect. 11.5.2.1)

$$\mathbf{e}_\alpha \cdot \mathbf{e}_\beta = \eta_{\alpha\beta} = \text{diag}(-1, 1, 1, 1)\,, \tag{A.15}$$

where $\alpha = 0, 1, 2, 3$ labels the spacetime coordinates $x^\alpha \equiv (ct, x, y, z)$ of an inertial reference frame. Thus, the two sets of components $A^\alpha$ and $A_\alpha$ are related by

$$A_\alpha = \sum_\beta \eta_{\alpha\beta} A^\beta \quad \Leftrightarrow \quad A_0 = -A^0\,, \quad A_1 = A^1\,, \quad A_2 = A^2\,, \quad A_3 = A^3\,. \tag{A.16}$$

So for relativistic mechanics, $A^0$ and $A_0$ differ by a minus sign, and hence it will be important to distinguish between the components of contravariant and covariant vectors. Whether an index is a superscript or a subscript *does* make a difference in special relativity.

**Exercise A.3** Show that under a coordinate transformation $x^i \to x^{i'}(x^i)$, the components $A^i$ of a contravariant vector transform like the coordinate differentials $\mathrm{d}x^i$, while the components $A_i$ of a covariant vector transform like the partial derivative operators $\partial/\partial x^i$.

### *A.2.2 Coordinate Notation*

Regarding coordinates, we will generally use superscripts (as we have above) to denote the collection of coordinates as a whole, e.g.,

(i) $x^i \equiv (x^1, x^2, x^3) = (x, y, z)$, $(r, \theta, \phi)$, or $(\rho, \phi, z)$ for ordinary 3-dimensional Euclidean space,
(ii) $x^\alpha \equiv (ct, x, y, z)$ for Minkowski spacetime,
(iii) $q^a \equiv (q^1, q^2, \ldots, q^n)$ for generalized coordinates defining the configuration of a system of particles having $n$ degrees of freedom.

Note that the superscripts just label the different coordinates, e.g., $x^2$ and $x^3$, and do *not* correspond to the square or cube of a single coordinate $x$. We choose this notation because tangents to curves in these spaces naturally define contravariant vectors, e.g.,

$$v^i \equiv \frac{\mathrm{d}x^i(\lambda)}{\mathrm{d}\lambda}\,, \tag{A.17}$$

and partial derivatives of scalars with respect to the coordinates naturally define covariant vectors, e.g.,

$$\omega_i \equiv \frac{\partial \varphi}{\partial x^i}\,, \tag{A.18}$$

with the placement of the superscript or subscript indices matching on both sides of the equation.[3] This is valid even for spaces that are not Euclidean and for coordinates that are not Cartesian, e.g., the angular coordinates describing the configuration of a planar double pendulum (See e.g., Problem 1.4).

### *A.2.3 Other Indices*

All other types of indices that we might need to use, e.g., to label different functions, basis vectors, or particles in a system, etc., will be placed as either superscripts or subscripts in whichever way is most notationally convenient for the discussion at hand. There is no "transformation law" associated with changes in these types of indices, so there is no standard convention for their placement.

## A.3 Differential Vector Calculus

To do calculus with vectors, we need *fields*—both **scalar fields**, which assign a real number to each position in space, and **vector fields**, which assign a three-dimensional vector to each position. An example of a scalar field is the gravitational potential $\Phi(\mathbf{r})$ for a stationary mass distribution, written as a function of the spatial location $\mathbf{r}$. An example of a vector field is the velocity $\mathbf{v}(\mathbf{r}, t)$ of a fluid at a fixed time $t$, which is a function of position $\mathbf{r}$ within the fluid.

Given a scalar field $U(\mathbf{r})$ and vector field $\mathbf{A}(\mathbf{r})$, we can define the following derivatives:

**(i) Gradient**:

$$(\nabla U) \cdot \hat{\mathbf{t}} \equiv \lim_{\Delta s \to 0} \left[ \frac{U(\mathbf{r}_2) - U(\mathbf{r}_1)}{\Delta s} \right], \tag{A.19}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the endpoints (i.e., the 'boundary') of the vector displacement $\Delta \mathbf{s} \equiv \Delta s\, \hat{\mathbf{t}}$. Thus, $(\nabla U) \cdot \hat{\mathbf{t}}$ measures the change in $U$ in the direction of $\hat{\mathbf{t}}$. This is the **directional derivative** of the scalar field $U$. The direction of $\nabla U$ is *perpendicular* to the contour lines $U(\mathbf{r}) = \text{const}$, since the right-hand side is zero for points that

[3] If we had swapped the notation and denoted the components of contravariant vectors with subscripts and the components of covariant vectors with superscripts, then to match indices would require denoting the collection of coordinates with subscripts, like $x_i$. In retrospect, this might have been a less confusing notation for coordinates (e.g., no chance of confusing the second coordinate $x_2$ with $x$-squared, etc.). But we will stick with the coordinate index notation that we have adopted above since it is the standard notation in the literature.

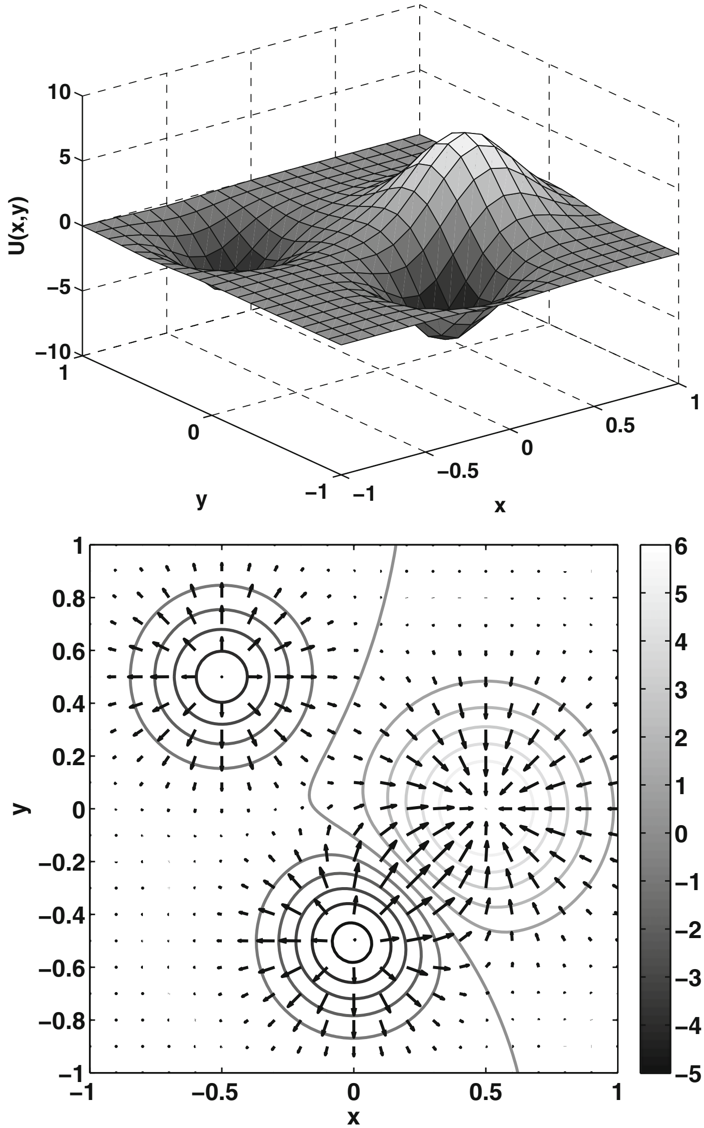



**Fig. A.2** Top panel: Function $U(x, y)$ displayed as a 2-dimensional surface. Bottom panel: Contour plot (lines of constant $U$, lighter lines corresponding to larger values) with gradient vector field $\nabla U$ superimposed. Note that the direction of $\nabla U$ is perpendicular to the $U(x, y) = \text{const}$ lines and is largest in magnitude where the change in $U$ is greatest

lie along a contour. Hence the gradient $\nabla U$ points in the direction of *steepest ascent* of the function $U$. This is illustrated graphically in Fig. A.2 for a function of two variables $U(x, y)$.

**(ii) Curl**:

$$(\nabla \times \mathbf{A}) \cdot \hat{\mathbf{n}} \equiv \lim_{\Delta a \to 0} \left[ \frac{1}{\Delta a} \oint_C \mathbf{A} \cdot d\mathbf{s} \right], \tag{A.20}$$

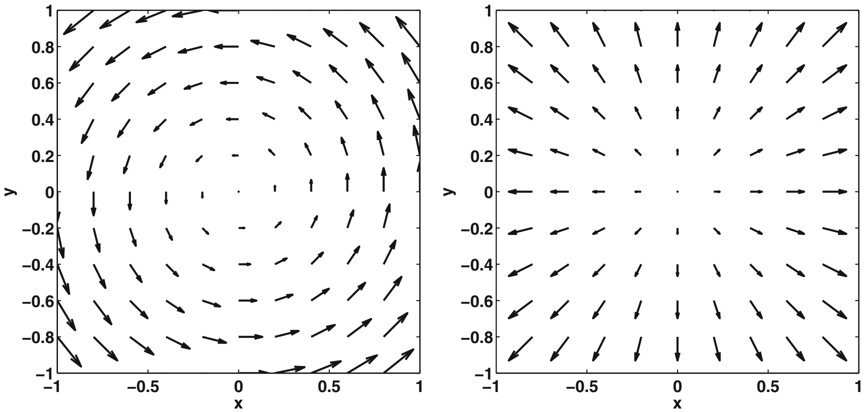
where $C$ is the boundary of the area element $\Delta \mathbf{a} \equiv \hat{\mathbf{n}}\, \Delta a$, and $d\mathbf{s}$ is the infinitesimal displacement vector tangent to $C$. The curl measures the circulation of $\mathbf{A}(\mathbf{r})$ around an infinitesimal closed curve. An example of a vector field with a non-zero curl is shown in panel (a) of Fig. A.3.

**(iii) Divergence**:

$$\nabla \cdot \mathbf{A} \equiv \lim_{\Delta V \to 0} \left[ \frac{1}{\Delta V} \oint_S \mathbf{A} \cdot \hat{\mathbf{n}}\, da \right], \tag{A.21}$$

where $S$ is the boundary of the volume $\Delta V$. The divergence measures the flux of $\mathbf{A}(\mathbf{r})$ through the surface bounding an infinitesimal volume element. An example of a vector field with a non-zero divergence is shown in panel (b) of Fig. A.3.

The beauty of the above definitions is that they are *geometric* and do not refer to a particular coordinate system. In Appendix A.5, we will write down expressions for the gradient, curl, and divergence in *arbitrary* orthogonal curvilinear coordinates $(u, v, w)$, which can be derived from the above definitions. In Cartesian coordinates $(x, y, z)$, the expressions for the three different derivatives turn out to be particularly simple:




(a) (b)

**Fig. A.3** Panel (a) Example of a vector field, $\mathbf{A}(\mathbf{r}) = -y\,\hat{\mathbf{x}} + x\,\hat{\mathbf{y}}$, with a non-zero curl, $\nabla \times \mathbf{A} = 2\hat{\mathbf{z}}$. Panel (b) Example of a vector field $\mathbf{A}(\mathbf{r}) = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}$, with a non-zero divergence, $\nabla \cdot \mathbf{A} = 3$. In both cases, just the $z = 0$ values of the vector fields are shown in these figures

$$\nabla U = \frac{\partial U}{\partial x}\,\hat{\mathbf{x}} + \frac{\partial U}{\partial y}\,\hat{\mathbf{y}} + \frac{\partial U}{\partial z}\,\hat{\mathbf{z}}\,,$$
$$\nabla \times \mathbf{A} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)\hat{\mathbf{x}} + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right)\hat{\mathbf{y}} + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)\hat{\mathbf{z}}\,,$$
$$\nabla \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\,. \tag{A.22}$$

In more compact form,

$$(\nabla U)_i = \partial_i U\,, \quad (\nabla \times \mathbf{A})_i = \sum_{j,k} \varepsilon_{ijk}\partial_j A_k\,, \quad \nabla \cdot \mathbf{A} = \sum_i \partial_i A_i\,, \tag{A.23}$$

where $\partial_i$ is shorthand for the partial derivative $\partial/\partial x^i$, where $x^i \equiv (x, y, z)$.

We conclude this subsection by noting that the curl and divergence of a vector field $\mathbf{A}(\mathbf{r})$, although important derivative operations, *do not completely capture how a vector field changes as you move from point to point*. A simple counting argument shows that, in three-dimensions, we need $3 \times 3 = 9$ components to completely specify how a vector field changes from point to point (three components of $\mathbf{A}$ times the three directions in which to take the derivative). The curl and divergence supply $3 + 1 = 4$ of those components. So we are missing 5 components, which turn out to have the geometrical interpretation of **shear** (See, e.g., Romano and Price 2012). The shear can be calculated in terms of the **directional derivative of a vector field**, which we shall discuss in Appendix A.4. Figure A.4 shows an example of a vector field that has zero curl and zero divergence, but is clearly not a constant. This is an example of a *pure-shear* field (See Exercise A.11).

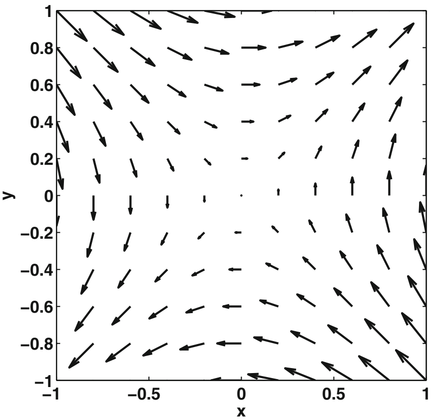



**Fig. A.4** Example of a vector field $\mathbf{A}(\mathbf{r}) = y\,\hat{\mathbf{x}} + x\,\hat{\mathbf{y}}$ that has both $\nabla \times \mathbf{A} = 0$ and $\nabla \cdot \mathbf{A} = 0$, but non-zero shear

### A.3.1 Product Rules

It turns out that the product rule

$$\frac{\mathrm{d}}{\mathrm{d}x}(fg) = \frac{\mathrm{d}f}{\mathrm{d}x}g + f\frac{\mathrm{d}g}{\mathrm{d}x} \tag{A.24}$$

for ordinary functions of one variable, $f(x)$ and $g(x)$, extends to the gradient, curl, and divergence operations, although the resulting expressions are more complicated. Since there are four different ways of combining a pair of scalar and/or vector fields (i.e., $fg$, $\mathbf{A}\cdot\mathbf{B}$, $f\mathbf{A}$, $\mathbf{A}\times\mathbf{B}$) and two different ways of taking derivatives of vector fields (either curl or divergence), there are six different product rules: here are six different product

$$\nabla(fg) = (\nabla f)g + f(\nabla g)\,, \tag{A.25a}$$
$$\nabla(\mathbf{A}\cdot\mathbf{B}) = \mathbf{A}\times(\nabla\times\mathbf{B}) + \mathbf{B}\times(\nabla\times\mathbf{A}) + (\mathbf{A}\cdot\nabla)\mathbf{B} + (\mathbf{B}\cdot\nabla)\mathbf{A}\,, \tag{A.25b}$$
$$\nabla\times(f\mathbf{A}) = (\nabla f)\times\mathbf{A} + f\nabla\times\mathbf{A}\,, \tag{A.25c}$$
$$\nabla\times(\mathbf{A}\times\mathbf{B}) = (\mathbf{B}\cdot\nabla)\mathbf{A} - (\mathbf{A}\cdot\nabla)\mathbf{B} + \mathbf{A}(\nabla\cdot\mathbf{B}) - \mathbf{B}(\nabla\cdot\mathbf{A})\,, \tag{A.25d}$$
$$\nabla\cdot(f\mathbf{A}) = (\nabla f)\cdot\mathbf{A} + f\nabla\cdot\mathbf{A}\,, \tag{A.25e}$$
$$\nabla\cdot(\mathbf{A}\times\mathbf{B}) = (\nabla\times\mathbf{A})\cdot\mathbf{B} - \mathbf{A}\cdot(\nabla\times\mathbf{B})\,. \tag{A.25f}$$

We will discuss some of these product rules in more detail in Appendix A.4.

**Exercise A.4** Prove the above product rules. (*Hint*: Do the calculations in Cartesian coordinates where the expressions for gradient, curl, and divergence are the simplest.)

### A.3.2 Second Derivatives

It is also possible to take *second* (and higher-order) derivatives of scalar and vector fields. Since $\nabla U$ and $\nabla\times\mathbf{A}$ are vector fields, we can take either their divergence or curl. Since $\nabla\cdot\mathbf{A}$ is a scalar field, we can take only its gradient. Thus, there are five such second derivatives:

$$\nabla\cdot\nabla U \equiv \nabla^2 U\,, \tag{A.26a}$$
$$\nabla\times\nabla U = 0\,, \tag{A.26b}$$
$$\nabla(\nabla\cdot\mathbf{A}) = \text{a vector field}\,, \tag{A.26c}$$
$$\nabla\cdot(\nabla\times\mathbf{A}) = 0\,, \tag{A.26d}$$
$$\nabla\times(\nabla\times\mathbf{A}) \equiv \nabla(\nabla\cdot\mathbf{A}) - \nabla^2\mathbf{A}\,. \tag{A.26e}$$

Note that the curl of a gradient, $\nabla \times \nabla U$, and the divergence of a curl, $\nabla \cdot (\nabla \times \mathbf{A})$, are both identically zero. The divergence of a gradient defines the **Laplacian** of a scalar field, $\nabla^2 U$, and the curl of a curl defines the Laplacian of a vector field, $\nabla^2 \mathbf{A}$ (second term on the right-hand side of (A.26e)). In Cartesian coordinates $x^i \equiv (x, y, z)$, the scalar and vector Laplacians are given by

$$\begin{aligned} \nabla^2 U &= \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2}\,, \\ \left(\nabla^2 \mathbf{A}\right)_i &= \frac{\partial^2 A_i}{\partial x^2} + \frac{\partial^2 A_i}{\partial y^2} + \frac{\partial^2 A_i}{\partial z^2}\,, \qquad i = 1, 2, 3\,. \end{aligned} \tag{A.27}$$

The gradient of a divergence is a non-zero vector field in general, but it has no special name, as it does not appear as frequently as the Laplacian operator.

***Example A.1*** Prove that $\nabla \times \nabla U = \mathbf{0}$ and $\nabla \cdot (\nabla \times \mathbf{A}) = 0$.
*Solution*: Since these are vector equations, we can do the proof in any coordinate system. For simplicity, we will use Cartesian coordinates $x^i \equiv (x, y, z)$ where the gradient, curl, and divergence are given by (A.23). Then

$$[\nabla \times \nabla U]_i = \sum_{j,k} \varepsilon_{ijk} \partial_j \partial_k U = 0\,, \tag{A.28}$$

since partial derivatives commute and $\varepsilon_{ijk}$ is totally anti-symmetric. Similarly,

$$\nabla \cdot (\nabla \times \mathbf{A}) = \sum_i \partial_i \left( \sum_{j,k} \varepsilon_{ijk} \partial_j A_k \right) = 0\,, \tag{A.29}$$

again since partial derivatives commute and $\varepsilon_{ijk}$ is totally anti-symmetric. □

**Exercise A.5** Verify $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$ in Cartesian coordinates.

## A.4 Directional Derivatives

You might have noticed that the right-hand sides of (A.25b) and (A.25d) for $\nabla(\mathbf{A} \cdot \mathbf{B})$ and $\nabla \times (\mathbf{A} \times \mathbf{B})$ involve quantities of the form $(\mathbf{B} \cdot \nabla)\mathbf{A}$, which are *not* gradients, curls, or divergences of a vector or scalar field. Geometrically, $(\mathbf{B} \cdot \nabla)\mathbf{A}$ represents the directional derivative of the vector field **A** in the direction of **B**, which generalizes the definition of the directional derivative of a scalar field. To calculate $(\mathbf{B} \cdot \boldsymbol{\nabla})\mathbf{A}$, we need to evaluate the directional derivatives of the components $A_i$ with respect to a basis $\hat{\mathbf{e}}_i$, as well as the directional derivatives of the *basis vectors* themselves. But before doing that calculation, it is worthwhile to remind ourselves about directional derivatives of scalar fields, and also how to calculate **coordinate basis vectors** in arbitrary curvilinear coordinates $(u, v, w)$.

### *A.4.1 Directional Derivative of a Function; Coordinate Basis Vectors*

Suppose we are given a curve $x^i = x^i(\lambda)$ parametrized by $\lambda$, with tangent vector $v^i \equiv \mathrm{d}x^i/\mathrm{d}\lambda$ defined along the curve. Then the directional derivative of a (scalar) function $f$ evaluated at any point along the curve is given by

$$\frac{\mathrm{d}f}{\mathrm{d}\lambda} \equiv \sum_i \frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\partial f}{\partial x^i} = \sum_i v^i \frac{\partial f}{\partial x^i} \equiv \mathbf{v}(f)\,. \tag{A.30}$$

The notation $\mathbf{v}(f)$ should be thought of as $\mathbf{v}$ "acting on" $f$. Note that if we abstract away the function $f$, we have $\mathbf{v} = \sum_i v^i \partial/\partial x^i$, with the partial derivative operators $\partial/\partial x^i$ playing the role of **coordinate basis vectors**. Denoting these basis vectors by the boldface symbol $\boldsymbol{\partial}_i$, we have[4]

$$\mathbf{v} = \sum_i v^i \boldsymbol{\partial}_i\,. \tag{A.31}$$

Thus, the directional derviative of a scalar field sets up a one-to-one correspondence between vectors $\mathbf{v}$ and directional derivative operators $\sum_i v^i \partial/\partial x^i$.

One nice feature about this correspondence between vectors and directional derivative operators is that it suggests how to calculate the coordinate basis vectors for arbitrary curvilinear coordinates $(u, v, w)$ in terms of the Cartesian basis vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$. One simply takes the chain rule

$$\frac{\partial}{\partial u} = \frac{\partial x}{\partial u}\frac{\partial}{\partial x} + \frac{\partial y}{\partial u}\frac{\partial}{\partial y} + \frac{\partial z}{\partial u}\frac{\partial}{\partial z}\,, \qquad \text{etc.}\,, \tag{A.32}$$

[4] A particular coordinate basis vector $\boldsymbol{\partial}_{\underline{i}}$ points along the $x^{\underline{i}}$ coordinate line, with all other coordinates (i.e., $x^j$ with $j \neq \underline{i}$) constant.

and formally converts it to a vector equation

$$\boldsymbol{\partial}_u = \frac{\partial x}{\partial u}\boldsymbol{\partial}_x + \frac{\partial y}{\partial u}\boldsymbol{\partial}_y + \frac{\partial z}{\partial u}\boldsymbol{\partial}_z\,, \qquad \text{etc.}\,, \tag{A.33}$$

with partial derivative operators replaced everywhere by coordinate basis vectors. But since the coordinate basis vectors in Cartesian coordinates are orthogonal and have unit norm, with $\boldsymbol{\partial}_x = \hat{\mathbf{x}}$, etc., it follows that

$$\begin{aligned} \boldsymbol{\partial}_u &= \frac{\partial x}{\partial u}\,\hat{\mathbf{x}} + \frac{\partial y}{\partial u}\,\hat{\mathbf{y}} + \frac{\partial z}{\partial u}\,\hat{\mathbf{z}}\,, \\ \boldsymbol{\partial}_v &= \frac{\partial x}{\partial v}\,\hat{\mathbf{x}} + \frac{\partial y}{\partial v}\,\hat{\mathbf{y}} + \frac{\partial z}{\partial v}\,\hat{\mathbf{z}}\,, \\ \boldsymbol{\partial}_w &= \frac{\partial x}{\partial w}\,\hat{\mathbf{x}} + \frac{\partial y}{\partial w}\,\hat{\mathbf{y}} + \frac{\partial z}{\partial w}\,\hat{\mathbf{z}}\,. \end{aligned} \tag{A.34}$$

The norms of these coordinate basis vectors are then given by

$$\begin{aligned} N_u &\equiv |\boldsymbol{\partial}_u| = \sqrt{\left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2}\,, \\ N_v &\equiv |\boldsymbol{\partial}_v| = \sqrt{\left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2}\,, \\ N_w &\equiv |\boldsymbol{\partial}_w| = \sqrt{\left(\frac{\partial x}{\partial w}\right)^2 + \left(\frac{\partial y}{\partial w}\right)^2 + \left(\frac{\partial z}{\partial w}\right)^2}\,, \end{aligned} \tag{A.35}$$

which we can then use to calculate *unit* vectors

$$\hat{\mathbf{u}} = N_u^{-1}\boldsymbol{\partial}_u\,, \qquad \hat{\mathbf{v}} = N_v^{-1}\boldsymbol{\partial}_v\,, \qquad \hat{\mathbf{w}} = N_w^{-1}\boldsymbol{\partial}_w\,. \tag{A.36}$$

In general, these unit vectors will *not* be orthogonal, although they will be for several common coordinate systems, including spherical coordinates $(r, \theta, \phi)$ and cylindrical coordinates $(\rho, \phi, z)$ (See Appendix A.5 for details). We will use the above results in the next section when calculating the directional derivative of a vector field in non-Cartesian coordinates.

### A.4.2 Directional Derivative of a Vector Field

Let's return now to the problem of calculating $(\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{A}$, which started this discussion of directional derivatives. In Cartesian coordinates, it is natural to define $(\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{A}$ in terms of its components via

$$(\mathbf{B} \cdot \nabla)\mathbf{A} \equiv \sum_{i,j} \left[(B_i \partial_i) A_j\right] \hat{\mathbf{e}}_j \tag{A.37}$$

since the orthonormal basis vectors $\hat{\mathbf{e}}_i = \{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$ are *constant* vector fields. In non-Cartesian coordinates, where the coordinate basis vectors change from point to point, we would need to make the appropriate coordinate transformations for both the vector components and the partial derivative operators. Although straightforward, this is usually a rather long and tedious process.

A simpler method for calculating $(\mathbf{B} \cdot \nabla)\mathbf{A}$ in non-Cartesian coordinates $x^i$ is to expand both $\mathbf{A}$ and $\mathbf{B}$ in terms of the orthonormal basis vectors $\hat{\mathbf{e}}_i$,

$$(\mathbf{B} \cdot \nabla)\mathbf{A} = \sum_{i,j} (B_i \hat{\mathbf{e}}_i \cdot \nabla)(A_j \hat{\mathbf{e}}_j) = \sum_{i,j} B_i (\nabla_{\hat{\mathbf{e}}_i} A_j)\hat{\mathbf{e}}_j + \sum_{i,j} B_i A_j (\nabla_{\hat{\mathbf{e}}_i} \hat{\mathbf{e}}_j)\,, \tag{A.38}$$

and then evaluate $\nabla_{\hat{\mathbf{e}}_i} \hat{\mathbf{e}}_j$ by further expanding $\hat{\mathbf{e}}_j$ as a linear combination of the Cartesian basis vectors $\hat{\mathbf{e}}_{i'} = \{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$. (Here we are using the notation $\nabla_{\hat{\mathbf{e}}_i} \equiv \hat{\mathbf{e}}_i \cdot \nabla$, and we are using a prime to distinguish the Cartesian basis vectors $\hat{\mathbf{e}}_{i'}$ from the non-Cartesian basis vectors $\hat{\mathbf{e}}_i$.) This leads to

$$\nabla_{\hat{\mathbf{e}}_i} \hat{\mathbf{e}}_j = \nabla_{\hat{\mathbf{e}}_i} \left( \sum_{k'} \Lambda_{jk'} \hat{\mathbf{e}}_{k'} \right) \equiv \sum_{k'} (\nabla_{\hat{\mathbf{e}}_i} \Lambda_{jk'}) \hat{\mathbf{e}}_{k'}\,, \tag{A.39}$$

where we have applied the derivatives only to the expansion coefficients $\Lambda_{jk'}$, since the Cartesian basis vectors $\hat{\mathbf{e}}_{k'}$ are constants. For example, for the spherical coordinate basis vectors $\hat{\mathbf{e}}_i = \{\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}\}$, we have

$$\begin{aligned}
\Lambda &= \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{bmatrix}, \\
\Lambda^{-1} &= \begin{bmatrix} \cos\phi\sin\theta & \cos\phi\cos\theta & -\sin\phi \\ \sin\phi\sin\theta & \sin\phi\cos\theta & \cos\phi \\ \cos\theta & -\sin\theta & 0 \end{bmatrix},
\end{aligned} \tag{A.40}$$

for the matrix of expansion coefficients $\Lambda_{jk'}$ and its inverse $(\Lambda^{-1})_{k'l}$ (See Example A.2 for details). If we then re-express the Cartesian basis vectors in terms of the original non-Cartesian basis vectors using the inverse transformation matrix $(\Lambda^{-1})_{k'l}$, we obtain

$$\nabla_{\hat{\mathbf{e}}_i} \hat{\mathbf{e}}_j = \sum_{k'} (\nabla_{\hat{\mathbf{e}}_i} \Lambda_{jk'}) \sum_{l} (\Lambda^{-1})_{k'l} \hat{\mathbf{e}}_l = \sum_{l} C_{ijl} \hat{\mathbf{e}}_l\,, \tag{A.41}$$

where

$$C_{ijl} \equiv \sum_{k'} (\nabla_{\hat{\mathbf{e}}_i} \Lambda_{jk'})(\Lambda^{-1})_{k'l}\,. \tag{A.42}$$

The $C_{ijl}$ are often called **connection coefficients**. Thus,

$$(\mathbf{B}\cdot\nabla)\mathbf{A} = \sum_{i,j} B_i(\nabla_{\hat{\mathbf{e}}_i} A_j)\hat{\mathbf{e}}_j + \sum_{i,j,l} B_i A_j C_{ijl}\hat{\mathbf{e}}_l\,. \tag{A.43}$$

Finally, if the non-Cartesian coordinates $x^i$ are *orthogonal*, as is the case for spherical coordinates $(r, \theta, \phi)$ and cylindrical coordinates $(\rho, \phi, z)$, then $\nabla_{\hat{\mathbf{e}}_i} = N_i^{-1}\partial/\partial x^i$, where $N_i$ is a normalization factor relating the (in general, unnormalized) coordinate basis vectors $\boldsymbol{\partial}_i$ to the orthonormal basis vectors $\hat{\mathbf{e}}_i$. For example, in spherical coordinates $\hat{\mathbf{r}} = \boldsymbol{\partial}_r$, $\hat{\boldsymbol{\theta}} = r^{-1}\boldsymbol{\partial}_\theta$, and $\hat{\boldsymbol{\phi}} = (r\sin\theta)^{-1}\boldsymbol{\partial}_\phi$.

Although this might seem like a complicated procedure when discussed abstractly, in practice it is relatively easy to carry out, as the following example shows.

***Example A.2*** Calculate $\nabla_{\hat{\mathbf{e}}_i}\hat{\mathbf{e}}_j$ in spherical coordinates $(r, \theta, \phi)$.
*Solution*: Recall that spherical coordinates $(r, \theta, \phi)$ are related to Cartesian coordinates $(x, y, z)$ via

$$x = r\sin\theta\cos\phi\,, \qquad y = r\sin\theta\sin\phi\,, \qquad z = r\cos\theta\,. \tag{A.44}$$

Using the chain rule to relate partial derivatives, e.g.,

$$\frac{\partial}{\partial r} = \frac{\partial x}{\partial r}\frac{\partial}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial}{\partial y} + \frac{\partial z}{\partial r}\frac{\partial}{\partial z}\,, \qquad \text{etc.}\,, \tag{A.45}$$

it follows that

$$\begin{aligned}
\hat{\mathbf{r}} &= \boldsymbol{\partial}_r = \sin\theta\cos\phi\,\hat{\mathbf{x}} + \sin\theta\sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}}\,,\\
\hat{\boldsymbol{\theta}} &= r^{-1}\boldsymbol{\partial}_\theta = \cos\theta\cos\phi\,\hat{\mathbf{x}} + \cos\theta\sin\phi\,\hat{\mathbf{y}} - \sin\theta\,\hat{\mathbf{z}}\,,\\
\hat{\boldsymbol{\phi}} &= (r\sin\theta)^{-1}\boldsymbol{\partial}_\phi = -\sin\phi\,\hat{\mathbf{x}} + \cos\phi\,\hat{\mathbf{y}}\,,
\end{aligned} \tag{A.46}$$

using the one-to-one correspondence between vectors and directional derivative operators discussed in Appendix A.4.1. The inverse transformation is given by

$$\begin{aligned}
\hat{\mathbf{x}} &= \boldsymbol{\partial}_x = \cos\phi\sin\theta\,\hat{\mathbf{r}} + \cos\phi\cos\theta\,\hat{\boldsymbol{\theta}} - \sin\phi\,\hat{\boldsymbol{\phi}}\,,\\
\hat{\mathbf{y}} &= \boldsymbol{\partial}_y = \sin\phi\sin\theta\,\hat{\mathbf{r}} + \sin\phi\cos\theta\,\hat{\boldsymbol{\theta}} + \cos\phi\,\hat{\boldsymbol{\phi}}\,,\\
\hat{\mathbf{z}} &= \boldsymbol{\partial}_z = \cos\theta\,\hat{\mathbf{r}} - \sin\theta\,\hat{\boldsymbol{\theta}}\,.
\end{aligned} \tag{A.47}$$

Performing the derivatives as described above, we find

$$\begin{aligned}
\nabla_{\hat{\mathbf{r}}}\hat{\mathbf{r}} &= \partial_r \left(\sin\theta\cos\phi\,\hat{\mathbf{x}} + \sin\theta\sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}}\right) = 0\,,\\
\nabla_{\hat{\boldsymbol{\theta}}}\hat{\mathbf{r}} &= r^{-1}\partial_\theta \left(\sin\theta\cos\phi\,\hat{\mathbf{x}} + \sin\theta\sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}}\right)\\
&= r^{-1}\left(\cos\theta\cos\phi\,\hat{\mathbf{x}} + \cos\theta\sin\phi\,\hat{\mathbf{y}} - \sin\theta\,\hat{\mathbf{z}}\right) = r^{-1}\,\hat{\boldsymbol{\theta}}\,,\\
\nabla_{\hat{\boldsymbol{\phi}}}\hat{\mathbf{r}} &= (r\sin\theta)^{-1}\partial_\phi \left(\sin\theta\cos\phi\,\hat{\mathbf{x}} + \sin\theta\sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}}\right)\\
&= (r\sin\theta)^{-1}\left(-\sin\theta\sin\phi\,\hat{\mathbf{x}} + \sin\theta\cos\phi\,\hat{\mathbf{y}}\right) = r^{-1}\,\hat{\boldsymbol{\phi}}\,.
\end{aligned} \tag{A.48}$$

Continuing in this fashion:

$$\begin{aligned}
&\nabla_{\hat{\mathbf{r}}}\hat{\mathbf{r}} = 0\,, \quad \nabla_{\hat{\boldsymbol{\theta}}}\hat{\mathbf{r}} = r^{-1}\,\hat{\boldsymbol{\theta}}\,, \quad \nabla_{\hat{\boldsymbol{\phi}}}\hat{\mathbf{r}} = r^{-1}\,\hat{\boldsymbol{\phi}}\,,\\
&\nabla_{\hat{\mathbf{r}}}\hat{\boldsymbol{\theta}} = 0\,, \quad \nabla_{\hat{\boldsymbol{\theta}}}\hat{\boldsymbol{\theta}} = -r^{-1}\,\hat{\mathbf{r}}\,, \quad \nabla_{\hat{\boldsymbol{\phi}}}\hat{\boldsymbol{\theta}} = r^{-1}\cot\theta\,\hat{\boldsymbol{\phi}}\,,\\
&\nabla_{\hat{\mathbf{r}}}\hat{\boldsymbol{\phi}} = 0\,, \quad \nabla_{\hat{\boldsymbol{\theta}}}\hat{\boldsymbol{\phi}} = 0\,, \quad \nabla_{\hat{\boldsymbol{\phi}}}\hat{\boldsymbol{\phi}} = -r^{-1}\,\hat{\mathbf{r}} - r^{-1}\cot\theta\,\hat{\boldsymbol{\theta}}\,.
\end{aligned} \tag{A.49}$$

□

**Exercise A.6** Calculate $\nabla_{\hat{\mathbf{e}}_i}\hat{\mathbf{e}}_j$ in cylindrical coordinates $(\rho, \phi, z)$.
You should find

$$x = \rho\cos\phi\,, \qquad y = \rho\sin\phi\,, \qquad z = z\,. \tag{A.50}$$

Relation between basis vectors:

$$\begin{aligned}
\hat{\boldsymbol{\rho}} &= \boldsymbol{\partial}_\rho = \cos\phi\,\hat{\mathbf{x}} + \sin\phi\,\hat{\mathbf{y}}\,,\\
\hat{\boldsymbol{\phi}} &= \rho^{-1}\boldsymbol{\partial}_\phi = -\sin\phi\,\hat{\mathbf{x}} + \cos\phi\,\hat{\mathbf{y}}\,,\\
\hat{\mathbf{z}} &= \boldsymbol{\partial}_z = \hat{\mathbf{z}}\,,
\end{aligned} \tag{A.51}$$

with inverse relations:

$$\begin{aligned}
\hat{\mathbf{x}} &= \cos\phi\,\hat{\boldsymbol{\rho}} - \sin\phi\,\hat{\boldsymbol{\phi}}\,,\\
\hat{\mathbf{y}} &= \sin\phi\,\hat{\boldsymbol{\rho}} + \cos\phi\,\hat{\boldsymbol{\phi}}\,,\\
\hat{\mathbf{z}} &= \hat{\mathbf{z}}\,,
\end{aligned} \tag{A.52}$$

and directional derivatives:

$$\begin{aligned}
&\nabla_{\hat{\rho}}\hat{\boldsymbol{\rho}} = 0\,, \quad \nabla_{\hat{\boldsymbol{\phi}}}\hat{\boldsymbol{\rho}} = \rho^{-1}\,\hat{\boldsymbol{\phi}}\,, \quad \nabla_{\hat{\mathbf{z}}}\hat{\boldsymbol{\rho}} = 0\,,\\
&\nabla_{\hat{\rho}}\hat{\boldsymbol{\phi}} = 0\,, \quad \nabla_{\hat{\boldsymbol{\phi}}}\hat{\boldsymbol{\phi}} = -\rho^{-1}\,\hat{\boldsymbol{\rho}}\,, \quad \nabla_{\hat{\mathbf{z}}}\hat{\boldsymbol{\phi}} = 0\,,\\
&\nabla_{\hat{\rho}}\hat{\mathbf{z}} = 0\,, \quad \nabla_{\hat{\boldsymbol{\phi}}}\hat{\mathbf{z}} = 0\,, \quad \nabla_{\hat{\mathbf{z}}}\hat{\mathbf{z}} = 0\,.
\end{aligned} \tag{A.53}$$

## A.5 Orthogonal Curvilinear Coordinates

In this section we derive expressions for the gradient, curl, and divergence in general orthogonal curvilinear coordinates $(u, v, w)$. Our starting point will be the definitions of gradient, curl, and divergence given in (A.19), (A.20), and (A.21). Examples of orthogonal curvilinear coordinates include Cartesian coordinates $(x, y, z)$, spherical coordinates $(r, \theta, \phi)$, and cylindrical coordinates $(\rho, \phi, z)$. These are the three main coordinate systems that we will be using most in this text.

Recall that in Cartesian coordinates $(x, y, z)$, the **line element** or infinitesimal squared distance between two nearby points is given by

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \qquad \text{(Cartesian)}\,. \tag{A.54}$$

Using the transformation equations (A.44) and (A.50), it is fairly easy to show that in spherical coordinates $(r, \theta, \phi)$ and in cylindrical coordinates $(\rho, \phi, z)$:

$$\begin{aligned} \mathrm{d}s^2 &= \mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2 \sin^2\theta\,\mathrm{d}\phi^2 \qquad &\text{(spherical)}\,,\\ \mathrm{d}s^2 &= \mathrm{d}\rho^2 + \rho^2\,\mathrm{d}\phi^2 + \mathrm{d}z^2 \qquad &\text{(cylindrical)}\,. \end{aligned} \tag{A.55}$$

More generally, in orthogonal curvilinear coordinates $(u, v, w)$:

$$\mathrm{d}s^2 = f^2\,\mathrm{d}u^2 + g^2\,\mathrm{d}v^2 + h^2\,\mathrm{d}w^2\,, \tag{A.56}$$

where $f$, $g$, and $h$ are functions of $(u, v, w)$ in general. The fact that there are no cross terms, like $\mathrm{d}u\,\mathrm{d}v$, is a consequence of the coordinates being *orthogonal*. Note that $f = 1$, $g = r$, and $h = r\sin\theta$ for spherical coordinates and $f = 1$, $g = \rho$, and $h = 1$ for cylindrical coordinates. These results are summarized in Table A.1.

For completely arbitrary curvilinear coordinates $x^i \equiv (x^1, x^2, x^3)$, the line element, (A.56), has the more general form

$$\mathrm{d}s^2 = \sum_{i,j} g_{ij}\,\mathrm{d}x^i\mathrm{d}x^j\,, \tag{A.57}$$

**Table A.1** Coordinates $(u, v, w)$ and functions $f, g, h$ for different orthogonal curvilinear coordinate systems

| Coordinates | $u$ | $v$ | $w$ | $f$ | $g$ | $h$ |
|---|---|---|---|---|---|---|
| Cartesian | $x$ | $y$ | $z$ | 1 | 1 | 1 |
| Spherical | $r$ | $\theta$ | $\phi$ | 1 | $r$ | $r\sin\theta$ |
| Cylindrical | $\rho$ | $\phi$ | $z$ | 1 | $\rho$ | 1 |

where $g_{ij} \equiv g_{ij}(x^1, x^2, x^3)$. The quantities $g_{ij}$ are called the components of the **metric**, and they can be represented in this case by a $3 \times 3$ matrix (See Appendix D.4.3 for a general discussion of matrix calculations). The metric components arise, for example, when finding the *geodesic* curves (shortest distance paths) between two points in a general curved space (See e.g., (C.14) and Exercise (C.7) in the context of the calculus of variations, Appendix C).

To make connection with the definitions of gradient, curl, and divergence given in (A.19), (A.20), and (A.21), we need expressions for the infinitesimal displacement vector, volume element, and area elements in general orthogonal curvilinear coordinates $(u, v, w)$. From the line element (A.56), we can conclude that the infinitesimal displacement vector $\mathrm{d}\mathbf{s}$ connecting two nearby points is

$$\mathrm{d}\mathbf{s} = f\,\mathrm{d}u\,\hat{\mathbf{u}} + g\,\mathrm{d}v\,\hat{\mathbf{v}} + h\,\mathrm{d}w\,\hat{\mathbf{w}}\,. \tag{A.58}$$

This is illustrated graphically in panel (a) of Fig. A.5. It is also easy to see from this figure that the infinitesimal volume element $\mathrm{d}V$ is given by

$$\mathrm{d}V = fgh\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\,. \tag{A.59}$$

The infinitesimal area elements are

$$\hat{\mathbf{n}}\,\mathrm{d}a = \begin{cases} \pm\hat{\mathbf{u}}\,gh\,\mathrm{d}v\,\mathrm{d}w \\ \pm\hat{\mathbf{v}}\,hf\,\mathrm{d}w\,\mathrm{d}u \\ \pm\hat{\mathbf{w}}\,fg\,\mathrm{d}u\,\mathrm{d}v \end{cases} \tag{A.60}$$

with the $\pm$ sign depending on whether the unit normals to the area elements point in the direction of increasing (or decreasing) coordinate value. One such area element is illustrated graphically in panel (b) of Fig. A.5.

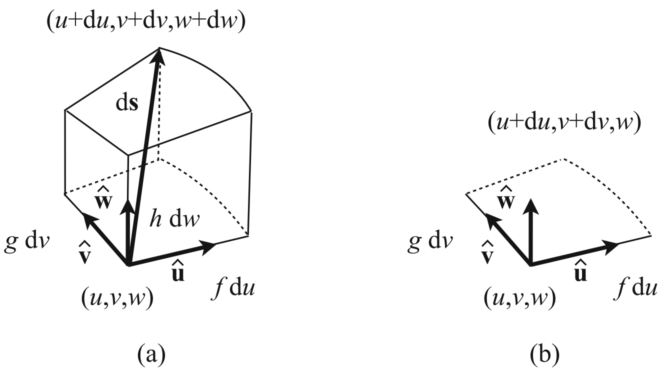



**Fig. A.5** Panel (a) Infinitesimal displacement vector $\mathrm{d}\mathbf{s}$ and volume element $\mathrm{d}V = fgh\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$ in general orthogonal curvilinear coordinates. Panel (b) The infinitesimal area element corresponding to the bottom ($w = $ const) surface of the volume element shown in panel (a)

For arbitrary curvilinear coordinates $x^i \equiv (x^1, x^2, x^3)$ with non-zero off-diagonal terms, we note that the above expressions generalize to

$$\mathrm{d}V = \sqrt{\det \mathsf{g}}\, \mathrm{d}x^1\, \mathrm{d}x^2\, \mathrm{d}x^3\,, \tag{A.61}$$

where $\det \mathsf{g}$ is the determinant of the matrix $\mathsf{g}$ of metric components $g_{ij}$ (See Appendix D.4.3.2), and

$$\hat{\mathbf{n}}\, \mathrm{d}a = \begin{cases} \pm\hat{\mathbf{n}}_1 \sqrt{g_{22}g_{33} - (g_{23})^2}\, \mathrm{d}x^2\, \mathrm{d}x^3 \\ \pm\hat{\mathbf{n}}_2 \sqrt{g_{33}g_{11} - (g_{31})^2}\, \mathrm{d}x^3\, \mathrm{d}x^1 \\ \pm\hat{\mathbf{n}}_3 \sqrt{g_{11}g_{22} - (g_{12})^2}\, \mathrm{d}x^1\, \mathrm{d}x^2 \end{cases} \tag{A.62}$$

where

$$\hat{\mathbf{n}}_1 = \frac{\boldsymbol{\partial}_2 \times \boldsymbol{\partial}_3}{|\boldsymbol{\partial}_2 \times \boldsymbol{\partial}_3|}\,, \quad \text{etc.} \tag{A.63}$$

Of particular relevance for both arbitrary curvilinear coordinates $(x^1, x^2, x^3)$ and orthogonal curvilinear coordinates $(u, v, w)$ is the distinction between the 3-dimensional and 2-dimensional *coordinate* volume and area elements, e.g., $\mathrm{d}^3x \equiv \mathrm{d}x^1\, \mathrm{d}x^2\, \mathrm{d}x^3$ and $\mathrm{d}^2x \equiv \mathrm{d}x^2\, \mathrm{d}x^3$, etc. (which are just products of coordinate differentials), and the *invariant* volume and area elements, $\mathrm{d}V$ and $\hat{\mathbf{n}}\, \mathrm{d}a$, which include the appropriate factors of the metric components $g_{ij}$.

**Exercise A.7** Verify (A.61) and (A.62).

### *A.5.1 Gradient*

Using the definition (A.19), it follows that

$$(\boldsymbol{\nabla} U) \cdot \mathrm{d}\mathbf{s} = \mathrm{d}U\,. \tag{A.64}$$

From (A.58), the left-hand side of the above equation can be written as

$$(\boldsymbol{\nabla} U) \cdot \mathrm{d}\mathbf{s} = (\boldsymbol{\nabla} U)_u\, f\, \mathrm{d}u + (\boldsymbol{\nabla} U)_v\, g\, \mathrm{d}v + (\boldsymbol{\nabla} U)_w\, h\, \mathrm{d}w\,, \tag{A.65}$$

while the right-hand side can be written as

$$\mathrm{d}U = \frac{\partial U}{\partial u}\, \mathrm{d}u + \frac{\partial U}{\partial v}\, \mathrm{d}v + \frac{\partial U}{\partial w}\, \mathrm{d}w\,. \tag{A.66}$$

By equating these last two equations, we can read off the components of the gradient, from which we obtain

$$\nabla U = \frac{1}{f}\frac{\partial U}{\partial u}\,\hat{\mathbf{u}} + \frac{1}{g}\frac{\partial U}{\partial v}\,\hat{\mathbf{v}} + \frac{1}{h}\frac{\partial U}{\partial w}\,\hat{\mathbf{w}}\,. \tag{A.67}$$

**Exercise A.8** Consider a particle of mass $m$ moving in the potential

$$U(x, y, z) = \frac{1}{2}k(x^2 + y^2) + mgz\,. \tag{A.68}$$

Calculate the force $\mathbf{F} = -\nabla U$ in (a) spherical coordinates and (b) cylindrical coordinates. You should find:

$$\begin{aligned} &\text{(a)} \quad \mathbf{F} = (-kr\sin^2\theta - mg\cos\theta)\,\hat{\mathbf{r}} + (-kr\sin\theta\cos\theta + mg\sin\theta)\,\hat{\boldsymbol{\theta}}\,, \\ &\text{(b)} \quad \mathbf{F} = -k\rho\,\hat{\boldsymbol{\rho}} - mg\,\hat{\mathbf{z}}\,. \end{aligned} \tag{A.69}$$

### A.5.2 Curl

Using the definition (A.20), it follows that

$$(\nabla \times \mathbf{A}) \cdot \hat{\mathbf{n}}\,\mathrm{d}a = \oint_C \mathbf{A} \cdot \mathrm{d}\mathbf{s}\,, \tag{A.70}$$

where $C$ is the *infinitesimal* closed curve bounding the area element $\hat{\mathbf{n}}\,\mathrm{d}a$, with orientation given by the right-hand rule relative to $\hat{\mathbf{n}}$. To calculate the components of $\nabla \times \mathbf{A}$, we take (in turn) the three different infinitesimal area elements given in (A.60). Starting with $\hat{\mathbf{n}}\,\mathrm{d}a = \hat{\mathbf{u}}\,gh\,\mathrm{d}v\,\mathrm{d}w$, the left-hand side of (A.70) becomes

$$(\nabla \times \mathbf{A}) \cdot \hat{\mathbf{n}}\,\mathrm{d}a = (\nabla \times \mathbf{A})_u\,gh\,\mathrm{d}v\,\mathrm{d}w\,. \tag{A.71}$$

Since this area element lies in a $u$ = const surface, $\mathrm{d}u = 0$, for which

$$\mathbf{A} \cdot \mathrm{d}\mathbf{s} = A_v\,g\,\mathrm{d}v + A_w\,h\,\mathrm{d}w\,. \tag{A.72}$$

Integrating this around the corresponding boundary curve $C$ shown in Fig. A.6, we find that the right-hand side of (A.70) becomes

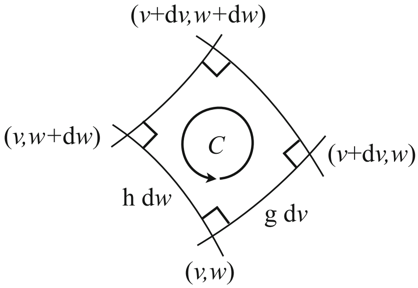



**Fig. A.6** Infinitesimal closed curve $C$ in the $u = \text{const}$ surface bounding the area element $\hat{\mathbf{n}}\,\mathrm{d}a = \hat{\mathbf{u}}\,gh\,\mathrm{d}v\,\mathrm{d}w$

$$\begin{aligned}\oint_C \mathbf{A}\cdot\mathrm{d}\mathbf{s} &= (A_v g)|_w\,\mathrm{d}v + (A_w h)|_{v+dv}\,\mathrm{d}w - (A_v g)|_{w+dw}\,\mathrm{d}v - (A_w h)|_v\,\mathrm{d}w\\ &= \frac{\partial}{\partial v}(A_w h)\,\mathrm{d}v\,\mathrm{d}w - \frac{\partial}{\partial w}(A_v g)\,\mathrm{d}w\,\mathrm{d}v\,.\end{aligned} \tag{A.73}$$

Thus,

$$(\nabla\times\mathbf{A})_u = \frac{1}{gh}\left[\frac{\partial}{\partial v}(A_w\,h) - \frac{\partial}{\partial w}(A_v\,g)\right]. \tag{A.74}$$

Repeating the above calculation for the other two components yields

$$(\nabla\times\mathbf{A})_v = \frac{1}{hf}\left[\frac{\partial}{\partial w}(A_u\,f) - \frac{\partial}{\partial u}(A_w\,h)\right], \tag{A.75}$$

and

$$(\nabla\times\mathbf{A})_w = \frac{1}{fg}\left[\frac{\partial}{\partial u}(A_v\,g) - \frac{\partial}{\partial v}(A_u\,f)\right]. \tag{A.76}$$

### *A.5.3 Divergence*

Using the definition (A.21), it follows that

$$(\nabla\cdot\mathbf{A})\,\mathrm{d}V = \oint_S \mathbf{A}\cdot\hat{\mathbf{n}}\,\mathrm{d}a\,, \tag{A.77}$$

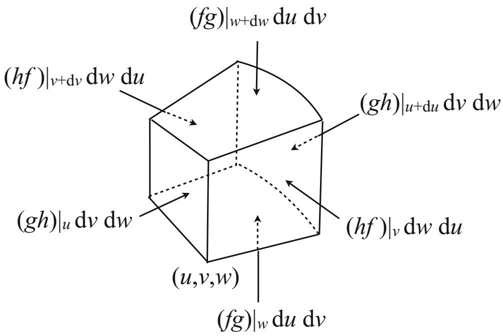



**Fig. A.7** Infinitesimal closed surface $S$ bounding the volume element $\mathrm{d}V = fgh\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$. The magnitudes of the infinitesimal area elements comprising $S$ are also given; the notation $\big|_u$ means evaluated at $u$, etc

where $S$ is the *infinitesimal* closed surface bounding the volume element $\mathrm{d}V$, with outward pointing normal $\hat{\mathbf{n}}$. From (A.59), the left-hand side of the above equation can be written as

$$(\nabla \cdot \mathbf{A})\,\mathrm{d}V = (\nabla \cdot \mathbf{A})\, fgh\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\,. \tag{A.78}$$

From (A.60), the integrand of the right-hand side of (A.77) contains the terms

$$\mathbf{A}\cdot\hat{\mathbf{n}}\,\mathrm{d}a = \begin{cases} \pm A_u\, gh\,\mathrm{d}v\,\mathrm{d}w \\ \pm A_v\, hf\,\mathrm{d}w\,\mathrm{d}u \\ \pm A_w\, fg\,\mathrm{d}u\,\mathrm{d}v \end{cases}. \tag{A.79}$$

Integrating over the boundary surface $S$ shown in Fig. A.7, we obtain

$$\begin{aligned} \oint_S \mathbf{A}\cdot\hat{\mathbf{n}}\,\mathrm{d}a = {} & (A_u\, gh)|_{u+du}\;\mathrm{d}v\,\mathrm{d}w - (A_u\, gh)|_u\;\mathrm{d}v\,\mathrm{d}w \\ & +(A_v\, hf)|_{v+dv}\;\mathrm{d}w\,\mathrm{d}u - (A_v\, hf)|_v\;\mathrm{d}w\,\mathrm{d}u \\ & +(A_w\, fg)|_{w+dw}\;\mathrm{d}u\,\mathrm{d}v - (A_w\, fg)|_w\;\mathrm{d}u\,\mathrm{d}v \\ = {} & \left[\frac{\partial}{\partial u}(A_u\, gh) + \frac{\partial}{\partial v}(A_v\, hf) + \frac{\partial}{\partial w}(A_w\, fg)\right]\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w\,. \end{aligned} \tag{A.80}$$

Thus,

$$\nabla\cdot\mathbf{A} = \frac{1}{fgh}\left[\frac{\partial}{\partial u}(A_u\, gh) + \frac{\partial}{\partial v}(A_v\, hf) + \frac{\partial}{\partial w}(A_w\, fg)\right]. \tag{A.81}$$

### A.5.4 Laplacian

Since the Laplacian of a scalar field is defined as the divergence of the gradient, it immediately follows from (A.67) and (A.81) that

$$\nabla^2 U = \frac{1}{fgh}\left[\frac{\partial}{\partial u}\left(\frac{gh}{f}\frac{\partial U}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{hf}{g}\frac{\partial U}{\partial v}\right) + \frac{\partial}{\partial w}\left(\frac{fg}{h}\frac{\partial U}{\partial w}\right)\right]. \tag{A.82}$$

**Exercise A.9** Show that in spherical coordinates $(r, \theta, \phi)$:

$$\begin{aligned}
\nabla U &= \frac{\partial U}{\partial r}\,\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial U}{\partial \theta}\,\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial U}{\partial \phi}\,\hat{\boldsymbol{\phi}}\,,\\
\nabla \times \mathbf{A} &= \frac{1}{r\sin\theta}\left[\frac{\partial}{\partial \theta}(A_\phi \sin\theta) - \frac{\partial A_\theta}{\partial \phi}\right]\hat{\mathbf{r}}\\
&\quad + \frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial A_r}{\partial \phi} - \frac{\partial}{\partial r}(A_\phi r)\right]\hat{\boldsymbol{\theta}} + \frac{1}{r}\left[\frac{\partial}{\partial r}(A_\theta r) - \frac{\partial A_r}{\partial \theta}\right]\hat{\boldsymbol{\phi}}\,,\\
\nabla \cdot \mathbf{A} &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 A_r\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}(\sin\theta\, A_\theta) + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial \phi}\,,\\
\nabla^2 U &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial U}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial U}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 U}{\partial \phi^2}\,.
\end{aligned} \tag{A.83}$$

**Exercise A.10** Show that in cylindrical coordinates $(\rho, \phi, z)$:

$$\begin{aligned}
\nabla U &= \frac{\partial U}{\partial \rho}\,\hat{\boldsymbol{\rho}} + \frac{1}{\rho}\frac{\partial U}{\partial \phi}\,\hat{\boldsymbol{\phi}} + \frac{\partial U}{\partial z}\,\hat{\mathbf{z}}\,,\\
\nabla \times \mathbf{A} &= \left[\frac{1}{\rho}\frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z}\right]\hat{\boldsymbol{\rho}} + \left[\frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial \rho}\right]\hat{\boldsymbol{\theta}} + \frac{1}{\rho}\left[\frac{\partial}{\partial \rho}(A_\phi \rho) - \frac{\partial A_\rho}{\partial \phi}\right]\hat{\mathbf{z}}\,,\\
\nabla \cdot \mathbf{A} &= \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho A_\rho\right) + \frac{1}{\rho}\frac{\partial A_\phi}{\partial \phi} + \frac{\partial A_z}{\partial z}\,,\\
\nabla^2 U &= \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial U}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 U}{\partial \phi^2} + \frac{\partial^2 U}{\partial z^2}\,.
\end{aligned} \tag{A.84}$$

## A.6 Integral Theorems of Vector Calculus

Using the definitions of gradient, curl, and divergence given above, one can prove the following fundamental theorems of integral vector calculus:

**Theorem A.1 Fundamental theorem for gradients***:*

$$\int_C (\nabla U) \cdot \mathrm{d}\mathbf{s} = U(\mathbf{r}_b) - U(\mathbf{r}_a)\,, \tag{A.85}$$

*where* $\mathbf{r}_a$, $\mathbf{r}_b$ *are the endpoints of* $C$.

**Theorem A.2 Stokes' theorem***:*

$$\int_S (\nabla \times \mathbf{A}) \cdot \hat{\mathbf{n}}\, \mathrm{d}a = \oint_C \mathbf{A} \cdot \mathrm{d}\mathbf{s}\,, \tag{A.86}$$

*where* $C$ *is the closed curved bounding the surface* $S$*, with orientation given by the right-hand rule relative to* $\hat{\mathbf{n}}$.

**Theorem A.3 Divergence theorem***:*

$$\int_V (\nabla \cdot \mathbf{A})\, \mathrm{d}V = \oint_S \mathbf{A} \cdot \hat{\mathbf{n}}\, \mathrm{d}a\,, \tag{A.87}$$

*where* $S$ *is the closed surface bounding the volume* $V$*, with outward pointing normal* $\hat{\mathbf{n}}$.

For infinitesimal volume elements, area elements, and path lengths, the proofs of these theorems follow trivially from the definitions given in (A.19), (A.20), and (A.21). For *finite* size volumes, areas, and path lengths, one simply adds together the contribution from infinitesimal elements. The neighboring surfaces, edges, and endpoints of these infinitesimal elements have *oppositely-directed* normals, tangent vectors, etc., and hence yield terms that cancel out when forming the sum. For detailed proofs, we recommend Schey (1996), Boas (2006), or Griffiths (1999).

## A.7 Some Additional Theorems for Vector Fields

Here we state (without proof) some additional theorems for vector fields. These make use of the identities $\nabla \times \nabla U = 0$ and $\nabla \cdot (\nabla \times \mathbf{A}) = 0$, which we derived earlier (See Example A.1), and the integral theorems of vector calculus from the previous subsection.

**Theorem A.4** *Any vector field* $\mathbf{F}$ *can be written in the form*

$$\mathbf{F} = -\nabla U + \nabla \times \mathbf{W} \,. \tag{A.88}$$

Note that this decomposition is not unique as the transformations

$$\begin{aligned} &U \rightarrow U + C \,, \quad \text{where } C = \text{const} \,, \\ &\mathbf{W} \rightarrow \mathbf{W} + \nabla \Lambda \,, \end{aligned} \tag{A.89}$$

leave $\mathbf{F}$ unchanged.

**Theorem A.5** *Curl-free vector fields:*

$$\nabla \times \mathbf{F} = 0 \quad \Leftrightarrow \quad \mathbf{F} = -\nabla U \quad \Leftrightarrow \quad \oint_C \mathbf{F} \cdot \mathrm{d}\mathbf{s} = 0 \,. \tag{A.90}$$

**Theorem A.6** *Divergence-free vector fields:*

$$\nabla \cdot \mathbf{F} = 0 \quad \Leftrightarrow \quad \mathbf{F} = \nabla \times \mathbf{W} \quad \Leftrightarrow \quad \oint_S \mathbf{F} \cdot \hat{\mathbf{n}} \, \mathrm{d}a = 0 \,. \tag{A.91}$$

Both of the above theorems require that: (i) $\mathbf{F}$ be differentiable, and (ii) the region of interest be simply-connected (i.e., that there are not any holes; see the discussion in Appendix B.2). We will assume that both of these conditions are always satisfied.

Theorem A.5 is particularly relevant in the context of **conservative forces**, which we encounter often in the main text. Recall that $\mathbf{F}$ is conservative if and only if the work done by $\mathbf{F}$ in moving a particle from $\mathbf{r}_a$ to $\mathbf{r}_b$ is *independent* of the path connecting the two points. But path-independence is equivalent to the condition that $\oint_C \mathbf{F} \cdot \mathrm{d}\,\mathbf{s} = 0$ for any closed curve $C$. Thus, from Theorem A.5, we can conclude that a conservative force is curl-free, i.e., $\nabla \times \mathbf{F} = 0$, and that it can always be written as the gradient of a scalar field.

**Exercise A.11** In two dimensions, consider the vector fields

$$\mathbf{A} = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} \,, \qquad \mathbf{B} = -y\,\hat{\mathbf{x}} + x\,\hat{\mathbf{y}} \,, \qquad \mathbf{C} = y\,\hat{\mathbf{x}} + x\,\hat{\mathbf{y}} \,. \tag{A.92}$$

(a) Show that $\nabla \times \mathbf{A} = 0$, $\nabla \cdot \mathbf{B} = 0$, and $\nabla \times \mathbf{C} = 0$, $\nabla \cdot \mathbf{C} = 0$.
(b) Make plots of these vector fields.
(c) Show that $\mathbf{A} = \rho\,\hat{\boldsymbol{\rho}}$ and $\mathbf{B} = \rho\,\hat{\boldsymbol{\phi}}$, where $(\rho, \phi)$ are plane polar coordinates related to $(x, y)$ via $x = \rho\cos\phi$ and $y = \rho\sin\phi$.

(d) Show that $\mathbf{C} = \frac{1}{2}\nabla V$, where $(U, V)$ are orthogonal hyperbolic coordinates on the plane defined by $U \equiv x^2 - y^2$ and $V \equiv 2xy$.

A non-constant vector field like **C**, which is both curl-free and divergence-free, is said to be a **pure-shear** vector field. The "shearing pattern" of the **C** will look like that in Fig. A.4.

## A.8 Dirac Delta Function

The **Dirac delta function** $\delta(\mathbf{r} - \mathbf{r}_0)$ is a mathematical representation of a "spike"—i.e., a quantity that is zero at all points except at the spike, where it is infinite,

$$\delta(\mathbf{r} - \mathbf{r}_0) = \begin{cases} 0\,, & \text{if } \mathbf{r} \neq \mathbf{r}_0 \\ \infty\,, & \text{if } \mathbf{r} = \mathbf{r}_0 \end{cases} \tag{A.93}$$

and such that[5]

$$\int_V \mathrm{d}V\, \delta(\mathbf{r} - \mathbf{r}_0) = 1 \tag{A.94}$$

for any volume $V$ containing the spike. The Dirac delta function is not an ordinary mathematical function. It is what mathematicians call a **generalized function** or **distribution**. An example of a Dirac delta function is the mass density of an idealized point particle, $\mu(\mathbf{r}) = m\delta(\mathbf{r} - \mathbf{r}_0)$.

In one dimension, the Dirac delta function $\delta(x - x_0)$ can be represented as the limit of a sequence of functions $f_n(x)$ all of which have unit area, but which get narrower and higher as $n \to \infty$. Some simple example sequences are:

(i) A sequence of top-hat functions centered at $x_0$ with width $2/n$:

$$f_n(x) = \begin{cases} n/2 & x_0 - 1/n < x < x_0 + 1/n \\ 0 & \text{otherwise} \end{cases} \tag{A.95}$$

(ii) A sequence of Gaussian probability distributions with mean $\mu = x_0$ and standard deviation $\sigma = 1/n$:

[5] We are adopting here the standard "physicist's" definition of a 3-dimensional Dirac delta function (See e.g., Griffiths 1999), where we integrate it against the 3-dimensional volume element $\mathrm{d}V$. But note that we could also define a 3-dimensional Dirac delta function $\tilde{\delta}(\mathbf{r} - \mathbf{r}_0)$ with respect to the *coordinate* volume element $\mathrm{d}^3x$ via $\int_V \mathrm{d}^3x\, \tilde{\delta}(\mathbf{r} - \mathbf{r}_0) = 1$. (Recall that $\mathrm{d}^3x = \mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$ while $\mathrm{d}V = fgh\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$ for orthogonal curvilinear coordinates $(u, v, w)$.) The difference between these two definitions of the Dirac delta function shows up in their transformation properties under a coordinate transformation, see Footnote 7.

$$f_n(x) = \frac{n}{\sqrt{2\pi}} e^{-n^2(x-x_0)^2/2} \tag{A.96}$$

(iii) A sequence of sinc functions[6] centered at $x_0$ of the form:

$$f_n(x) = \frac{n}{\pi} \operatorname{sinc}\left[n(x - x_0)\right] \tag{A.97}$$

Equivalently, the Dirac delta function can be defined in terms of its action on a set of **test functions** $f(x)$, which are infinitely differentiable and which vanish as $x \to \pm\infty$. The defining property of a 1-dimensional Dirac delta function is then

$$\int_a^b \mathrm{d}x\; f(x)\delta(x - x') = \begin{cases} f(x') & a < x' < b \\ 0 & \text{otherwise} \end{cases} \tag{A.98}$$

for any test function $f(x)$.

**Exercise A.12** Prove the following properties of the 1-dimensional Dirac delta function, which follow from the defining property (A.98):

$$\begin{aligned}
\delta(x - a) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(u(x - a)\right) , \\
\delta'(-x) &= -\delta'(x) , \\
\delta(-x) &= \delta(x) , \\
\delta(ax) &= \frac{1}{|a|}\delta(x) , \\
\delta\left[f(x)\right] &= \sum_i \frac{\delta(x - x_i)}{|f'(x_i)|} .
\end{aligned} \tag{A.99}$$

In the above expressions, $u(x)$ is the unit step function,

$$u(x) = \begin{cases} 0\,, & x < 0 \\ 1\,, & x \geq 0 \end{cases} \tag{A.100}$$

and $f(x)$ is such that $f(x_i) = 0$ and $f'(x_i) \neq 0$. The last two properties indicate that the one-dimensional Dirac delta function $\delta(x)$ transforms like a *density* under a change of variables—i.e., $\delta(x)\,\mathrm{d}x = \delta(y)\,\mathrm{d}y$.

[6]The sinc function, $\operatorname{sinc} x$, is defined by $\operatorname{sinc} x \equiv \sin x/x$.

In three dimensions, the defining property of the Dirac delta function is

$$\int_V \mathrm{d}V\ f(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}') = \begin{cases} f(\mathbf{r}') & \text{if } \mathbf{r}' \in V \\ 0 & \text{otherwise} \end{cases} \tag{A.101}$$

for any test function $f(\mathbf{r})$. Note that this definition implies

$$\begin{aligned} \delta(\mathbf{r}-\mathbf{r}') &= \delta(x-x')\delta(y-y')\delta(z-z')\,, \\ \delta(\mathbf{r}-\mathbf{r}') &= \frac{1}{r^2\sin\theta}\delta(r-r')\delta(\theta-\theta')\delta(\phi-\phi')\,, \\ \delta(\mathbf{r}-\mathbf{r}') &= \frac{1}{\rho}\delta(\rho-\rho')\delta(\phi-\phi')\delta(z-z')\,, \end{aligned} \tag{A.102}$$

in order that

$$\begin{aligned} \int \mathrm{d}V\ \delta(\mathbf{r}-\mathbf{r}') &= \int\!\!\int\!\!\int \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\ \delta(x-x')\delta(y-y')\delta(z-z') \\ &= \int\!\!\int\!\!\int \mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi\ \delta(r-r')\delta(\theta-\theta')\delta(\phi-\phi') \\ &= \int\!\!\int\!\!\int \mathrm{d}\rho\,\mathrm{d}\phi\,\mathrm{d}z\ \delta(\rho-\rho')\delta(\phi-\phi')\delta(z-z') \end{aligned} \tag{A.103}$$

be *independent* of the choice of coordinates. For general orthogonal curvilinear coordinates $(u, v, w)$,

$$\delta(\mathbf{r}-\mathbf{r}') = \frac{1}{fgh}\delta(u-u')\delta(v-v')\delta(w-w') \tag{A.104}$$

as a consequence of $\mathrm{d}V = fgh\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w$.[7]

There is also an integral representation of the 1-dimensional Dirac delta function, which can be heuristically "derived" by taking a limit of sinc functions:

$$\delta(x) = \lim_{L\to\infty}\frac{L}{\pi}\mathrm{sinc}(Lx) = \lim_{L\to\infty}\frac{1}{2\pi}\int_{-L}^{L}\mathrm{d}k\ \mathrm{e}^{\mathrm{i}kx} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}k\ \mathrm{e}^{\mathrm{i}kx}\,, \tag{A.105}$$

where we used (A.97). Thus,

$$\delta(x-x') = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}k\ \mathrm{e}^{\pm\mathrm{i}k(x-x')}\,. \tag{A.106}$$

[7] If we used the alternative definition of the 3-dimensional Dirac delta function $\tilde{\delta}(\mathbf{r}-\mathbf{r}_0)$ discussed in Footnote 5, then $\tilde{\delta}(\mathbf{r}-\mathbf{r}') = \delta(u-u')\delta(v-v')\delta(w-w')$, without the factor of $fgh$.

Similarly, in 3-dimensions,

$$\delta(\mathbf{r}-\mathbf{r}') = \frac{1}{(2\pi)^3}\int_{\text{all space}} \mathrm{d}V_{\mathbf{k}}\, \mathrm{e}^{\pm i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\,, \tag{A.107}$$

where $\mathrm{d}V_{\mathbf{k}}$ is the 3-dimensional volume element in **k**-space.

***Example A.3*** Recall that in Newtonian gravity the gravitational potential $\Phi(\mathbf{r}, t)$ satisfies Poisson's equation

$$\nabla^2\Phi(\mathbf{r}, t) = 4\pi G\mu(\mathbf{r}, t)\,, \tag{A.108}$$

where $G$ is Newton's constant and $\mu(\mathbf{r}, t)$ is the mass density of the source distribution. Note that the left-hand side of the above equation is just the Laplacian of $\Phi$. We now show that for a stationary point source $\mu(\mathbf{r}, t) = m\delta(\mathbf{r}-\mathbf{r}_0)$, the potential is given by the well-known formula

$$\Phi(\mathbf{r}) = -\frac{Gm}{|\mathbf{r}-\mathbf{r}_0|}\,. \tag{A.109}$$

We begin by noting that

$$\boldsymbol{\nabla}\cdot\left(\frac{\hat{\mathbf{r}}}{r^2}\right) = 0\,, \qquad \text{for } r \neq 0\,, \tag{A.110}$$

which follows from the expression for the divergence in spherical coordinates (See Exercise A.9). To determine its behavior at $r = 0$, we consider the volume integral of $\boldsymbol{\nabla}\cdot(\hat{\mathbf{r}}/r^2)$ over a spherical volume of radius $R$ centered at the origin. Using the divergence theorem (A.87), we obtain

$$\int_V \boldsymbol{\nabla}\cdot\left(\frac{\hat{\mathbf{r}}}{r^2}\right)\mathrm{d}V = \oint_S \frac{\hat{\mathbf{r}}}{r^2}\cdot\hat{\mathbf{n}}\,\mathrm{d}a = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\frac{1}{R^2}\,R^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi = 4\pi\,, \tag{A.111}$$

*independent* of the radius $R$. Thus, by comparison with the definition of the Dirac delta function, (A.94), we can conclude that

$$\boldsymbol{\nabla}\cdot\left(\frac{\hat{\mathbf{r}}}{r^2}\right) = 4\pi\delta(\mathbf{r})\,. \tag{A.112}$$

But since

$$\boldsymbol{\nabla}\left(\frac{1}{r}\right) = -\frac{\hat{\mathbf{r}}}{r^2}\,, \tag{A.113}$$

we can also write

$$\nabla^2\left(\frac{1}{r}\right) = -4\pi\delta(\mathbf{r})\,. \tag{A.114}$$

Finally, by simpling shifting the origin, we have

$$\nabla\cdot\left(\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}\right) = 4\pi\delta(\mathbf{r}-\mathbf{r}')\,,\quad \nabla^2\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) = -4\pi\delta(\mathbf{r}-\mathbf{r}')\,. \tag{A.115}$$

Thus, $\Phi(\mathbf{r}) = -Gm/|\mathbf{r}-\mathbf{r}_0|$ as claimed. □

**Exercise A.13** Show that

$$\begin{aligned} \nabla\times(r^n\hat{\mathbf{r}}) &= 0 \quad \text{for all } n\,, \\ \nabla\cdot(r^n\hat{\mathbf{r}}) &= (n+2)r^{n-1} \quad \text{for } n\neq -2\,. \end{aligned} \tag{A.116}$$

Thus, it is only for $n=-2$ that we get a Dirac delta function.

## Suggested References

*Full references are given in the bibliography at the end of the book.*

Boas (2006): Chapter 6 is devoted to vector algebra and vector calculus, especially suited for undergraduates.

Griffiths (1999): Chapter 1 provides an excellent review of vector algebra and vector calculus, at the same level as this appendix. Our discussion of orthogonal curvilinear coordinates in Appendix A.5 is a summary of Appendix A in Griffiths, which has more detailed derivations and discussion.

Schey (1996): An excellent introduction to vector calculus emphasizing the geometric nature of the divergence, gradient, and curl operations.

# Appendix B
# Differential Forms

Although we will not need to develop the full machinery of tensor calculus for the applications to classical mechanics covered in this book, the concept of a **differential form** and the associated operations of **exterior derivative** and **wedge product** will come in handy from time to time. For example, they are particulary useful for determining whether certain differential equations or constraints on a mechanical system (See e.g., Sect. 2.2.3) are *integrable* or not. They are also helpful in understanding the geometric structure underlying Poisson brackets (Sect. 3.5). More generally, differential forms are actually the quantities that you integrate on a manifold, with the integral theorems of vector calculus (Appendix A.6) being special cases of a more general (differential-form version) of Stokes' theorem.

In broad terms, the exterior derivative is a generalization of the total derivative (or gradient) of a function, and the curl of a vector field in three dimensions. The wedge product is a generalization of the cross-product of two vectors. And differential forms are quantities constructed from a sum of wedge products of coordinate differentials $dx^i$. Readers interested to learn more about differential forms and related topics should see e.g., Flanders (1963) and Schutz (1980).

## B.1 Definitions

Since we have not developed a general framework for working with tensors, our presentation of differential forms will be somewhat heuristic, starting with familar examples for 0-forms and 1-forms, and then adding mathematical operations as needed (e.g., wedge product and exterior derivative) to construct higher-order differential forms. To keep things sufficiently general, we will consider an $n$-dimensional manifold $M$ with coordinates $x^i \equiv (x^1, x^2, \ldots, x^n)$. From time to time we will consider ordinary 3-dimensional space to make connection with more familiar mathematical objects and operations.



M.J. Benacquista and J.D. Romano, *Classical Mechanics*, Undergraduate Lecture Notes in Physics, https://doi.org/10.1007/978-3-319-68780-3

### B.1.1 0-Forms, 1-Forms, and Exterior Derivative

To begin, a **0-form** is just a function

$$\alpha \equiv \alpha(x^1, x^2, \ldots, x^n)\,, \tag{B.1}$$

while a **1-form** is a linear combination of the coordinate differentials,

$$\beta \equiv \sum_i \beta_i \, \mathrm{d}x^i\,, \tag{B.2}$$

for which the components $\beta_i \equiv \beta_i(x^1, x^2, \ldots, x^n)$ transform according to

$$\beta_{i'} = \sum_i \frac{\partial x^i}{\partial x^{i'}} \beta_i\,, \qquad i' = 1', 2', \ldots, n'\,, \tag{B.3}$$

under a coordinate transformation $x^i \to x^{i'}(x^i)$. We impose this requirement on the components in order that

$$\beta \equiv \sum_i \beta_i \, \mathrm{d}x^i = \sum_{i'} \beta_{i'} \, \mathrm{d}x^{i'} \tag{B.4}$$

be invariant under a coordinate transformation. The set $\{\mathrm{d}x^1, \mathrm{d}x^2, \ldots, \mathrm{d}x^n\}$ is a coordinate *basis* for the $n$-dimensional space of 1-forms on $M$. A simple example of a 1-form is the **exterior derivative** of a 0-form $\alpha$,

$$\mathrm{d}\alpha \equiv \sum_i (\partial_i \alpha) \, \mathrm{d}x^i\,, \tag{B.5}$$

where $\partial_i \alpha \equiv \partial \alpha / \partial x^i$. Note that the exterior derivative of a 0-form is just the usual total differential (or gradient) of a function.

**Exercise B.1** Verify (B.4) using (B.3) and the transformation property of the coordinate differentials $\mathrm{d}x^i$.

### B.1.2 2-Forms and Wedge Product

To construct a **2-form** from two 1-forms, we introduce the **wedge product** of two forms. We require this product to be *anti-symmetric*,

$$\mathrm{d}x^i \wedge \mathrm{d}x^j = -\mathrm{d}x^j \wedge \mathrm{d}x^i \,, \tag{B.6}$$

and *linear* with respect to its arguments,

$$\alpha \wedge (f\beta + g\gamma) = f(\alpha \wedge \beta) + g(\alpha \wedge \gamma) \,, \tag{B.7}$$

where $f$ and $g$ are any two functions. Given this definition, it immediately follows that the wedge product of two 1-forms $\alpha$ and $\beta$ can be written as

$$\alpha \wedge \beta = \sum_{i,j} \alpha_i \beta_j \, \mathrm{d}x^i \wedge \mathrm{d}x^j = \sum_{i<j} (\alpha_i \beta_j - \alpha_j \beta_i) \, \mathrm{d}x^i \wedge \mathrm{d}x^j \,, \tag{B.8}$$

where we used the anti-symmetry of $\mathrm{d}x^i \wedge \mathrm{d}x^j$ to get the last equality. Note that in three dimensions

$$\alpha_i \beta_j - \alpha_j \beta_i = \sum_k \varepsilon_{ijk} (\boldsymbol{\alpha} \times \boldsymbol{\beta})_k \,, \tag{B.9}$$

where on the right-hand side we are treating $\alpha_i$ and $\beta_i$ as the components of two vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Thus, the wedge product of two 1-forms generalizes the *cross product* of two vectors in three dimensions. The most general 2-form on $M$ will have the form

$$\gamma \equiv \sum_{i<j} \gamma_{ij} \, \mathrm{d}x^i \wedge \mathrm{d}x^j \,, \tag{B.10}$$

where the components $\gamma_{ij} \equiv \gamma_{ij}(x^1, x^2, \ldots, x^n)$ are totally anti-symmetric under interchange of $i$ and $j$ (i.e., $\gamma_{ij} = -\gamma_{ji}$ for all $i$ and $j$).

The exterior derivative can also be extended to an arbitrary 1-form $\alpha$. We simply take the exterior derivative of the components $\alpha_j$, for $j = 1, 2, \ldots, n$, and then wedge those 1-forms $\mathrm{d}\alpha_j = \sum_i (\partial_i \alpha_j) \, \mathrm{d}x^i$ with the coordinate differentials $\mathrm{d}x^j$. This leads to the 2-form

$$\mathrm{d}\alpha \equiv \sum_{i,j} (\partial_i \alpha_j) \, \mathrm{d}x^i \wedge \mathrm{d}x^j = \sum_{i<j} (\partial_i \alpha_j - \partial_j \alpha_i) \, \mathrm{d}x^i \wedge \mathrm{d}x^j \,. \tag{B.11}$$

Note that in three dimensions

$$\partial_i \alpha_j - \partial_j \alpha_i = \sum_k \varepsilon_{ijk} (\nabla \times \boldsymbol{\alpha})_k \,, \tag{B.12}$$

where on the right-hand side we are treating $\alpha_i$ as the components of a vector field $\boldsymbol{\alpha}$. So the exterior derivative of a 1-form generalizes the *curl* of a vector field in three dimensions.

### *B.1.3 3-Forms and Higher-Order Forms*

We can continue in this fashion to construct 3-forms, 4-forms, etc., by requiring that the wedge product be *associative*,

$$\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma = \alpha \wedge \beta \wedge \gamma\,. \tag{B.13}$$

Thus, a general $p$-form $\alpha$ can be written as

$$\alpha = \sum_{i_1 < i_2 < \cdots < i_p} \alpha_{i_1 i_2 \cdots i_p}\, \mathrm{d}x^{i_1} \wedge \mathrm{d}x^{i_2} \wedge \cdots \wedge \mathrm{d}x^{i_p}\,, \tag{B.14}$$

where the components $\alpha_{i_1 i_2 \cdots i_p} \equiv \alpha_{i_1 i_2 \cdots i_p}(x^1, x^2, \ldots, x^n)$ are totally anti-symmetric under interchange of the indices $i_1, i_2, \ldots, i_p$. Similarly, the exterior derivative of a $p$-form $\alpha$ is the $(p+1)$ form

$$\mathrm{d}\alpha = \sum_{i_1 < i_2 < \cdots < i_{p+1}} \left(\partial_{i_1}\alpha_{i_2 i_3 \cdots i_{p+1}} - \partial_{i_2}\alpha_{i_1 i_3 \cdots i_{p+1}} \cdots - \partial_{i_{p+1}}\alpha_{i_2 i_3 \cdots i_p i_1}\right) \mathrm{d}x^{i_1} \wedge \mathrm{d}x^{i_2} \wedge \cdots \wedge \mathrm{d}x^{i_p+1}\,. \tag{B.15}$$

Note that $(n+1)$ and higher-rank forms in an $n$-dimensional space are identically zero due to the anti-symmetry of the wedge product, i.e., $\mathrm{d}x^i \wedge \mathrm{d}x^j = 0$ for $i = j$.

### *B.1.4 Total Anti-Symmetrization*

By introducing a notation for totally anti-symmetrizing a set of indices, e.g.,

$$\begin{aligned} [ij] &\equiv \frac{1}{2!}(ij - ji)\,, \\ [ijk] &\equiv \frac{1}{3!}(ijk - ikj + jki - jik + kij - kji)\,, \\ &\text{etc.}\,, \end{aligned} \tag{B.16}$$

we can write down the general expressions for the components of the wedge product and exterior derivative in compact form:

$$(\alpha \wedge \beta)_{i_1 \cdots i_p j_1 \cdots j_q} = \frac{(p+q)!}{p!q!}\, \alpha_{[i_1 \cdots i_p} \beta_{j_1 \cdots j_q]} \tag{B.17}$$

and

$$(\mathrm{d}\alpha)_{i_1 i_2 \cdots i_{p+1}} = (p+1)\partial_{[i_1}\alpha_{i_2 \cdots i_{p+1}]}\,, \tag{B.18}$$

where $\alpha$ and $\beta$ denote a $p$-form and $q$-form, respectively.

**Exercise B.2** Let $\alpha$ be a $p$-form and $\beta$ be a $q$-form. Show that

$$\alpha \wedge \beta = (-1)^{pq}\beta \wedge \alpha\,. \tag{B.19}$$

**Exercise B.3** Let $\alpha$ be a $p$-form and $\beta$ be a $q$-form. Show that the exterior derivative of the wedge product $\alpha \wedge \beta$ satisfies

$$\mathrm{d}(\alpha \wedge \beta) = (\mathrm{d}\alpha) \wedge \beta + (-1)^p \alpha \wedge (\mathrm{d}\beta)\,, \tag{B.20}$$

as a consequence of the ordinary product rule for partial derivatives and the anti-symmetry of the differential forms.

## B.2 Closed and Exact Forms

Given the above definitions, we can introduce some additional terminology:

- A $p$-form $\alpha$ is said to be **exact** if there exists a $(p-1)$-form $\beta$ for which $\alpha = \mathrm{d}\beta$.
- A $p$-form $\alpha$ is said to be **closed** if $\mathrm{d}\alpha = 0$.

It is easy to show that all exact forms are closed—i.e.,

$$\mathrm{d}(\mathrm{d}\beta) = 0\,, \tag{B.21}$$

as a consequence of the commutativity of partial derivatives, $\partial_i \partial_j = \partial_j \partial_i$. This result is called the **Poincaré lemma**. But what about the converse? Are all closed forms also exact?

The answer is that all closed forms are *locally* exact, but *globally* this need not be true (See e.g., Schutz 1980 for a proof). Global exactness of a closed form requires that the space be *topologically trivial* (i.e., simply-connected), in the sense that the space shouldn't contain any "holes". More precisely, this means that any closed loop in the space should be (smoothly) contractible to a point. Ordinary 3-dimensional space with no points removed or the surface of a 2-sphere are examples of simply-connected spaces. The *punctured plane* ($\mathbb{R}^2$ with the origin removed) or the surface of a torus are examples of spaces that are not simply-connected. Any closed curve

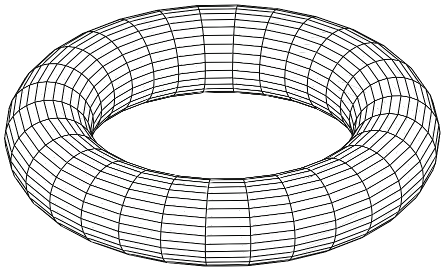


**Fig. B.1** The coordinate lines on a torus are examples of closed curves that are not contractible to a point

encircling the origin of the punctured plane, and any of the coordinates lines on the torus shown in Fig. B.1 are not contractible to a point. (See Schutz 1980 or Flanders 1963 for more details.)

**Exercise B.4** In three dimensions, show that $\mathrm{d}(\mathrm{d}\alpha) = 0$ corresponds to

(a) $\nabla \times \nabla\alpha = \mathbf{0}$ if $\alpha$ is a 0-form;
(b) $\nabla \cdot (\nabla \times \mathbf{A}) = 0$ if $\alpha$ is a 1-form $\alpha = \sum_i \alpha_i \, \mathrm{d}x^i$ with $A_i \equiv \alpha_i$.

**Exercise B.5** Consider the 1-form

$$\alpha \equiv \frac{1}{x^2 + y^2} (-y \, \mathrm{d}x + x \, \mathrm{d}y) \tag{B.22}$$

defined on the punctured plane. Show that $\alpha$ is closed, but globally is not exact. Find a function $f(x, y)$ for which $\alpha = \mathrm{d}f$ locally. (*Hint*: Plane polar coordinates $(r, \phi)$ might be useful for this.)

## B.3 Frobenius' Theorem

You may recall from a math methods class trying to determine if a 1st-order differential equation of the form

$$A(x, y) \, \mathrm{d}x + B(x, y) \, \mathrm{d}y = 0 \tag{B.23}$$

is integrable or not. You may also remember that if

$$\partial_y A = \partial_x B \,, \tag{B.24}$$

then (at least locally) there exists a function $\varphi \equiv \varphi(x, y)$ for which

$$\mathrm{d}\varphi = A(x, y)\,\mathrm{d}x + B(x, y)\,\mathrm{d}y\,, \quad \text{with} \quad A = \partial_x \varphi\,, \quad B = \partial_y \varphi\,. \tag{B.25}$$

But requiring that the differential equation be exact is actually *too strong* a requirement for integrability. More generally, (B.23) is integrable if and only if there exists a function $\mu \equiv \mu(x, y)$, called an **integrating factor**, for which

$$\mu(x, y)\,[A(x, y)\,\mathrm{d}x + B(x, y)\,\mathrm{d}y] \tag{B.26}$$

is exact, so that[1]

$$\partial_y(\mu A) = \partial_x(\mu B)\,. \tag{B.27}$$

It turns out that in two dimensions one can *always* find such an integrating factor. Thus, *all* 1st-order differential equations of the form given in (B.23) are integrable (Exercise B.6). But explicitly finding an integrating factor in practice is not an easy task in general.

Now in three and higher dimensions not all 1st-order differential equations are integrable, so testing for integrability is a necessary and important task. Writing the differential equation in $n$ dimensions as

$$\alpha \equiv \sum_i \alpha_i\,\mathrm{d}x^i = 0\,, \tag{B.28}$$

where $\alpha_i \equiv \alpha_i(x^1, x^2, \ldots, x^n)$, the question of integrability again becomes does there exist an integrating factor $\mu \equiv \mu(x^1, x^2, \ldots, x^n)$ for which

$$\mathrm{d}\varphi = \mu\alpha \tag{B.29}$$

for some $\varphi$. This would imply

$$\partial_i(\mu\alpha_j) = \partial_j(\mu\alpha_i) \quad \text{for all} \quad i, j = 1, 2, \ldots, n\,, \tag{B.30}$$

which in the language of differential forms becomes

$$0 = \mathrm{d}\mu \wedge \alpha + \mu\,\mathrm{d}\alpha \quad \Leftrightarrow \quad \mathrm{d}\alpha = -\mu^{-1}\mathrm{d}\mu \wedge \alpha\,. \tag{B.31}$$

---

[1] Recall from thermodynamics that heat flow is described by an *inexact* differential $đQ$ (notationally, the bar on the 'd' is to indicate that it is not the total differential of a function $Q$). But $đQ$ becomes exact when multiplied by an integrating factor, i.e., $\mathrm{d}S = đQ/T$, where $T$ is the temperature and $S$ is the entropy.

But since $\alpha \wedge \alpha = 0$, it follows that

$$\mathrm{d}\alpha \wedge \alpha = 0\,. \tag{B.32}$$

This is a *necessary* condition for (B.28) to be integrable. That it is also a *sufficient* condition was proven by Frobenius in 1877. Thus, **Frobenius' theorem** tells us that (B.32) is necessary and sufficient for the integrability of the 1st-order differential equation (B.28).

Frobenius's theorem can also be extended to the case of a *system* of 1st-order differential equations:

$$\alpha^A \equiv \sum_i \alpha_i^A\, \mathrm{d}x^i = 0\,, \qquad A = 1, 2, \ldots, M\,, \tag{B.33}$$

where $M < n$ and $\alpha_i^A \equiv \alpha_i^A(x^1, x^2, \ldots, x^n)$. We would like to know if this system is integrable in the sense of defining an $(n - M)$-dimensional hypersurface in the original $n$-dimensional space of coordinates. The necessary and sufficient condition for this to be true is the existence of an invertible transformation from the $\alpha^A$ to a set of exact 1-forms (i.e., total differentials):

$$\mathrm{d}\varphi^A = \sum_B \mu_{AB}\alpha^B \quad \Leftrightarrow \quad \alpha^A = \sum_B \left(\mu^{-1}\right)_{AB} \mathrm{d}\varphi^B\,, \tag{B.34}$$

where $\varphi^A \equiv \varphi^A(x^1, x^2, \ldots, x^n)$ is a set of functions, and $\mu_{AB} \equiv \mu_{AB}(x^1, x^2, \ldots, x^n)$ are the components of an invertible matrix (which is a generalization of the integrating factor $\mu$ for a single equation). In this context, Frobenius' theorem states that the set of constraints given by (B.33) is integrable if and only if

$$\mathrm{d}\alpha^A \wedge \alpha^1 \wedge \alpha^2 \wedge \cdots \wedge \alpha^M = 0\,, \qquad A = 1, 2, \ldots, M\,. \tag{B.35}$$

For a single differential equation, we recover (B.32).

**Exercise B.6** Using Frobenius' theorem, prove that *any* 1st-order differential equation in two dimensions,

$$\alpha \equiv A(x, y)\, \mathrm{d}x + B(x, y)\, \mathrm{d}y = 0\,, \tag{B.36}$$

is integrable.

**Exercise B.7** (*Adapted from Flanders* 1963.) Consider the 1st-order differential equation

$$\alpha \equiv yz\,\mathrm{d}x + xz\,\mathrm{d}y + \mathrm{d}z = 0\,, \tag{B.37}$$

in three dimensions.

(a) Use Frobenius' theorem to show that this equation is integrable.
(b) Verify that $\mu = \mathrm{e}^{xy}$ is an integrating factor for $\alpha$ with $\varphi = z\mathrm{e}^{xy}$.

## B.4 Integration of Differential Forms

Although you may not have thought about it this way, the things that you integrate on a manifold are really just differential forms. Indeed, the integrand $f(x)\,\mathrm{d}x$ of the familiar integral

$$\int_{x_1}^{x_2} f(x)\,\mathrm{d}x \tag{B.38}$$

from calculus is, in the language of this appendix, a 1-form. And the transformation property of $f(x)$ under a change of variables $x \to y(x)$,

$$f(x) \to f(y) \equiv \left.\frac{f(x)}{\mathrm{d}y/\mathrm{d}x}\right|_{x=x(y)}\,, \tag{B.39}$$

is just what you need in order for

$$f(x)\,\mathrm{d}x = f(y)\,\mathrm{d}y \tag{B.40}$$

to be *independent* of the choice of coordinates (compare these last two equations with (B.3) and (B.4)).

More generally, a 1-form field $\alpha$ on an $n$-dimensional manifold $M$ can be thought of as mapping from a 1-dimensional curve $C$ (with parameter $\lambda$ and tangent vector $\mathrm{d}x^i/\mathrm{d}\lambda$) to the value of the line integral

$$\int_C \alpha = \int_C \sum_i \alpha_i\,\mathrm{d}x^i \equiv \int_{\lambda_1}^{\lambda_2} \sum_i \alpha_i \frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\,\mathrm{d}\lambda\,. \tag{B.41}$$

Similarly, a 2-form field $\beta$ can be thought of as a mapping from a 2-dimensional surface $S$ (with coordinates $(u, v)$ and tangent vectors $\partial x^i/\partial u$, $\partial x^i/\partial v$) to the value of the surface integral

$$\begin{aligned}
\int_S \beta &= \int_S \sum_{i<j} \beta_{ij}\, \mathrm{d}x^i \wedge \mathrm{d}x^j \\
&\equiv \int_{u_1}^{u_2} \int_{v_1}^{v_2} \sum_{i<j} \beta_{ij} \left( \frac{\partial x^i}{\partial u} \frac{\partial x^j}{\partial v} - \frac{\partial x^j}{\partial u} \frac{\partial x^i}{\partial v} \right) \mathrm{d}u\, \mathrm{d}v \\
&= \int_{u_1}^{u_2} \int_{v_1}^{v_2} \sum_{i<j} \beta_{ij} \frac{\partial(x^i, x^j)}{\partial(u, v)}\, \mathrm{d}u\, \mathrm{d}v\,,
\end{aligned} \tag{B.42}$$

where the **Jacobian** of the transformation from $(x^i, x^j)$ to $(u, v)$ is[2]

$$\frac{\partial(x^i, x^j)}{\partial(u, v)} \equiv \begin{vmatrix} \frac{\partial x^i}{\partial u} & \frac{\partial x^i}{\partial v} \\ \frac{\partial x^j}{\partial u} & \frac{\partial x^j}{\partial v} \end{vmatrix} = \frac{\partial x^i}{\partial u} \frac{\partial x^j}{\partial v} - \frac{\partial x^j}{\partial u} \frac{\partial x^i}{\partial v}\,. \tag{B.43}$$

The extension to 3-form, 4-form, $\cdots$, $n$-form fields follows by noting that the above integrands are special cases of the general result that a $p$-form $\gamma$ maps a set of $p$ vectors with components $\{A^i, B^j, \ldots, C^k\}$ to the real number

$$\sum_{i<j<\cdots<k} \gamma_{ij\cdots k} \left( A^i B^j \cdots C^k - A^j B^i \cdots C^k - \ \cdots \ - A^k B^j \cdots C^i \right)\,. \tag{B.44}$$

By taking

$$A^i = \mathrm{d}u\, \frac{\partial x^i}{\partial u}\,, \qquad B^j = \mathrm{d}v\, \frac{\partial x^j}{\partial v}\,, \qquad \cdots\,, \qquad C^k = \mathrm{d}w\, \frac{\partial x^k}{\partial w}\,, \tag{B.45}$$

where $(u, v, \ldots, w)$ are the coordinates for a $p$-dimensional hypersurface, we are able to generalize (B.41) and (B.42) to arbitrary $p$-forms, with the appropriate Jacobians entering these expressions. Figure B.2 shows the tangent vectors and infinitesimal coordinate area element for a 2-dimensional surface spanned by the coordinates $(u, v)$.

Note that all of these integrals are *oriented* in the sense that swapping the order of the coordinates, e.g., $(u, v) \to (v, u)$, in the parametrization of the 2-dimensional surface $S$, changes the sign of the Jacobian and hence the sign of the integral. In addition, if one decides to change coordinates to do a particular integral, the Jacobian of the transformation enters automatically via the wedge product of the coordinate differentials. For example, in two dimensions, if one transforms from $(x, y)$ to $(u, v)$ it follows that

[2]The vertical lines in (B.43) mean you should take the *determinant* of the $2 \times 2$ matrix of partial derivatives. See Appendix D.4.3.2 for more details, if needed.

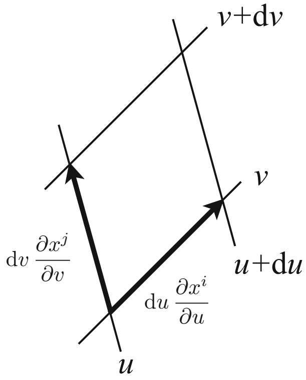



**Fig. B.2** Infinitesimal coordinate area element for a 2-dimensional surface spanned by the coordinates $(u, v)$

$$
\begin{aligned}
\mathrm{d}x \wedge \mathrm{d}y &= \left(\frac{\partial x}{\partial u}\,\mathrm{d}u + \frac{\partial x}{\partial v}\,\mathrm{d}v\right) \wedge \left(\frac{\partial y}{\partial u}\,\mathrm{d}u + \frac{\partial y}{\partial v}\,\mathrm{d}v\right) \\
&= \left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right)\mathrm{d}u \wedge \mathrm{d}v = \frac{\partial(x, y)}{\partial(u, v)}\,\mathrm{d}u \wedge \mathrm{d}v\,,
\end{aligned}
\tag{B.46}
$$

where we used the chain rule and the anti-symmetry of the wedge product to get the first and second equalities above. In $n$ dimensions, for a coordinate transformation from $x^i \equiv (x^1, x^2, \ldots, x^n)$ to $x^{i'} = (x^{1'}, x^{2'}, \ldots, x^{n'})$, we have

$$
\begin{aligned}
\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n &= \sum_{i'_1} \cdots \sum_{i'_n} \frac{\partial x^1}{\partial x^{i'_1}} \cdots \frac{\partial x^n}{\partial x^{i'_n}} \mathrm{d}x^{i'_1} \wedge \cdots \wedge \mathrm{d}x^{i'_n} \\
&= \sum_{i'_1} \cdots \sum_{i'_n} \frac{\partial x^1}{\partial x^{i'_1}} \cdots \frac{\partial x^n}{\partial x^{i'_n}} \varepsilon^{i'_1 \cdots i'_n}\, \mathrm{d}x^{1'} \wedge \cdots \wedge \mathrm{d}x^{n'} \\
&= \frac{\partial(x^1, \ldots, x^n)}{\partial(x^{1'}, \ldots, x^{n'})}\,\mathrm{d}x^{1'} \wedge \cdots \wedge \mathrm{d}x^{n'}\,,
\end{aligned}
\tag{B.47}
$$

where we made use of the $n$-dimensional Levi-Civita symbol and used (D.81) to get the last equality. Note that this is precisely the *inverse* transformation of the components of an $n$-form

$$
\omega_{1\cdots n} = \frac{\partial(x^{1'}, \ldots, x^{n'})}{\partial(x^1, \ldots, x^n)}\,\omega_{1'\cdots n'}\,, \tag{B.48}
$$

which is needed for

$$
\omega = \omega_{1\cdots n}\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n = \omega_{1'\cdots n'}\,\mathrm{d}x^{1'} \wedge \cdots \wedge \mathrm{d}x^{n'} \tag{B.49}
$$

to be invariant under a coordinate transformation. Given the presence of the Jacobian in (B.47), it is natural to interpret the $n$-dimensional coordinate element $\mathrm{d}^n x$ as the wedge product

$$\mathrm{d}^n x \equiv \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \cdots \wedge \mathrm{d}x^n \,, \tag{B.50}$$

and the integral of $\omega_{1\cdots n} \equiv \omega_{1\cdots n}(x^1, x^2, \ldots, x^n)$ as the integral of the $n$-form $\omega$:

$$\int \omega_{1\cdots n}\, \mathrm{d}^n x \equiv \int \omega_{1\cdots n}\, \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n = \int \omega \,. \tag{B.51}$$

### *B.4.1 Stokes' Theorem for Differential Forms*

Finally, to end this section, we note that the integral theorems of vector calculus (Appendix A.6) are actually special cases of an all-inclusive **Stokes' theorem**, written in terms of differential forms,

$$\int_U \mathrm{d}\alpha = \oint_{\partial U} \alpha \,, \tag{B.52}$$

where $\alpha$ is a $(p-1)$ form and $U$ is $p$-dimensional region in $M$ with boundary $\partial U$. We will not prove (B.52) here (See e.g., Flanders 1963). Rather we leave it as an exercise (Exercise B.9) to show that in three dimensions (B.52) reduces to the fundamental theorem for gradients (A.85), Stokes' theorem (A.86), and the divergence theorem (A.87), if one makes the appropriate identification of 1-forms and 2-forms with vector fields, and exterior derivative with either the gradient, curl, or divergence. Thus, (B.52) unifies the integral theorems of vector calculus.

**Exercise B.8** (a) Show explicitly by taking partial derivatives that a coordinate transformation from Cartesian coordinates $(x, y)$ to plane polar coordinates $(r, \phi)$ leads to

$$\mathrm{d}x \wedge \mathrm{d}y = r\, \mathrm{d}r \wedge \mathrm{d}\phi \,. \tag{B.53}$$

(b) Similarly, show that a coordinate transformation from Cartesian coordinates $(x, y, z)$ to spherical coordinates $(r, \theta, \phi)$ leads to

$$\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z = r^2 \sin\theta\, \mathrm{d}r \wedge \mathrm{d}\theta \wedge \mathrm{d}\phi \,. \tag{B.54}$$

**Exercise B.9** Show that in three dimensions (B.52) reduces to the fundamental theorem for gradients (A.85), Stokes' theorem (A.86), and the divergence theorem (A.87), by making the following identifications:

(a) For $p = 1$, identify the 0-form $\alpha$ with the function $U$, and the exterior derivative $\mathrm{d}\alpha$ with the gradient $\boldsymbol{\nabla} U$. Also, identify

$$\frac{\mathrm{d}x^i}{\mathrm{d}s}\,\mathrm{d}s \tag{B.55}$$

with the line element $\mathrm{d}\mathbf{s}$, where $\mathrm{d}x^i/\mathrm{d}s$ is the tangent vector to the curve $C$ parameterized by the arc length $s$.

(b) For $p = 2$, identify the 1-form $\alpha$ with the vector field $A_i \equiv \alpha_i$, and use (B.12) to identify $\mathrm{d}\alpha$ with $\boldsymbol{\nabla} \times \mathbf{A}$. Also, identify

$$\sum_{j<k} \varepsilon_{ijk} \frac{\partial(x^j, x^k)}{\partial(u, v)}\,\mathrm{d}u\,\mathrm{d}v \tag{B.56}$$

with the area element $\hat{\mathbf{n}}\,\mathrm{d}a$, where $(u, v)$ are coordinates on $S$.

(c) For $p = 3$, identifty the 2-form $\alpha$ with the vector field $A_i \equiv \sum_{j<k} \varepsilon_{ijk}\alpha_{jk}$, and the exterior derivative $\mathrm{d}\alpha$ with $\varepsilon_{ijk}\boldsymbol{\nabla} \cdot \mathbf{A}$. Also, identify

$$\sum_{i<j<k} \varepsilon_{ijk} \frac{\partial(x^i, x^j, x^k)}{\partial(u, v, w)}\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w \tag{B.57}$$

with the volume element $\mathrm{d}V$, where $(u, v, w)$ are coordinates in $V$.

## Suggested References

*Full references are given in the bibliography at the end of the book.*

Flanders (1963): A classic text about differential forms, appropriate for graduate students or advanced undergraduates comfortable with abstract mathematics.

Schutz (1980): A introduction to differential geometry, including tensor calculus and differential forms, with an emphasis on geometrical methods. Appropriate for graduate students or advanced undergraduates comfortable with abstract mathematics.

# Appendix C
# Calculus of Variations

The calculus of variations is an extension of the standard procedure for finding the **extrema** (i.e., maxima and minima) of a function $f(x)$ of a single real variable $x$. But instead of extremizing a function $f(x)$, we extremize a **functional** $I[y]$, which is a "function of a function" $y = f(x)$. Classic problems that can be solved using the calculus of variations are: (i) finding the curve connecting two points in the plane that has the shortest distance (a *geodesic* problem), (ii) finding the shape of a closed curve of fixed length that encloses the maximum area (an *isoperimetric* problem), (iii) finding the shape of a wire joining two points such that a bead will slide along the wire under the influence of gravity in the shortest amount of time (the famous *brachistochrone* problem of Johann Bernoulli). The calculus of variations also provides an alternative way of obtaining the equations of motion for a particle, or a system of particles, in classical mechanics. In this appendix, we derive the Euler equations, discuss ways of solving these equations in certain simplified scenarios, and extend the formalism to deal with integral constraints. For a more thorough introduction to the calculus of variations, see, e.g., Boas (2006), Gelfand and Fomin (1963), and Lanczos (1949). Specific applications to classical mechanics will be given in Chap. 3.

## C.1 Functionals

In its simplest form, a **functional** $I = I[y]$ is a mapping from some specified set of functions $\{y = f(x)\}$ to the set of real numbers $\mathbb{R}$. For the types of problems that we will be most interested in, the functions $y = f(x)$ are defined on some finite interval $x \in [x_1, x_2]$; they are single-valued and have continuous first derivatives; and they have *fixed endpoints* $\wp_1 \equiv (x_1, y_1)$, $\wp_2 \equiv (x_2, y_2)$. (Curves that cannot be described by a single-valued function can be put in parametric form, $x = x(t)$, $y = y(t)$, which we will discuss in detail in Appendix C.6.) A simple concrete example of a functional is the arc length of the curve traced out by a function $y = f(x)$ that connects $\wp_1$ and $\wp_2$:



M.J. Benacquista and J.D. Romano, *Classical Mechanics*, Undergraduate Lecture Notes in Physics, https://doi.org/10.1007/978-3-319-68780-3

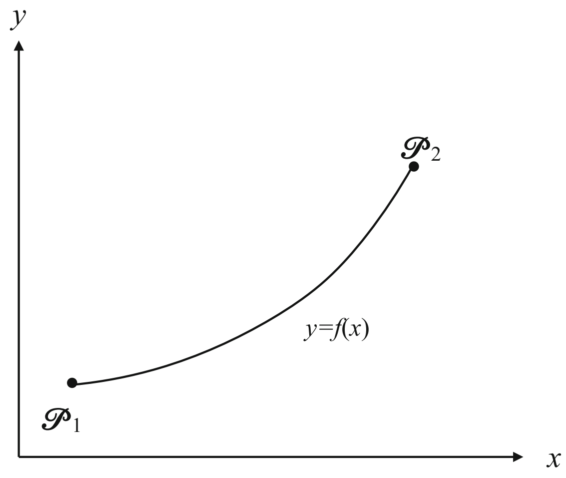



**Fig. C.1** The arc length of the curve traced out by the function $y = f(x)$ between $\wp_1$ and $\wp_2$ can be thought of as the value of a functional $I[y]$ evaluated for this particular function $y = f(x)$

$$I[y] \equiv \int_{\wp_1}^{\wp_2} \mathrm{d}s = \int_{\wp_1}^{\wp_2} \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} = \int_{x_1}^{x_2} \sqrt{1 + y'^2}\, \mathrm{d}x\,, \tag{C.1}$$

as shown in Fig. C.1. The corresponding calculus of variations problem is then to find the function $y = f(x)$ that minimizes the arc length between the two endpoints. We know the answer to this problem is a straight line, but to actually *prove* it requires some work. We will do this explicitly using the formalism of the calculus of variations in Example C.1 below.

More generally, we will consider functionals of the form

$$I[y] \equiv \int_{x_1}^{x_2} F(y, y', x)\, \mathrm{d}x\,, \tag{C.2}$$

where $x$ is the independent variable and the set of functions $\{y = f(x)\}$ is as before, but the integrand $F$ is now an *arbitrary* function of the three variables $(y, y', x)$. (For the arc-length functional defined previously, $F(y, y', x) = \sqrt{1 + y'^2}$, which is independent of $x$ and $y$, but that does not have to be the case in general.) Note that to do the integral over $x$, we need to express both $y$ and $y'$ in terms of $f(x)$, but for the variational calculations that follow, we simply treat $F$ as an ordinary function of three *independent* variables. The fact that $y$ and $y'$ are related to one another only shows up later on, when we need to relate the variation $\delta y'$ to $\delta y$, cf. (C.6).

Toward the end of this appendix, in Appendix C.7, we will extend our definition of a functional to $n$-degrees of freedom:

$$I[y_1, y_2, \ldots, y_n] \equiv \int_{x_1}^{x_2} F(y_1, y_2, \ldots, y_n; y_1', y_2', \ldots, y_n'; x)\, \mathrm{d}x\,, \tag{C.3}$$

where $y_i \equiv f_i(x)$, $i = 1, 2, \ldots, n$, are $n$ functions of the independent variable $x \in [x_1, x_2]$, which is a form more appropriate for classical mechanics problems with $x$ replaced by the time $t$. But for most of this appendix, we will work with the simpler functional given by (C.2).

## C.2 Deriving the Euler Equation

Given $I[y]$, we now want to find its extrema—i.e., those functions $y = f(x)$ for which $I[y]$ has a local maximum or minimum. Similar to ordinary calculus, a *necessary* (but not sufficient) condition for $y$ to be an extremum is that the *1st-order* change in $I[y]$ vanish for arbitrary variations to $y = f(x)$ that preserve the boundary conditions. We define such a variation to $y = f(x)$ by

$$\delta y \equiv \bar{f}(x) - f(x)\,, \tag{C.4}$$

where $\bar{f}(x)$ is a function that differs infinitesimally from $f(x)$ at each value of $x$ in the domain $[x_1, x_2]$ (See Fig. C.2). In terms of $\delta y$, the variation of the functional is then given by $\delta I[y] \equiv I[y+\delta y] - I[y]$, where we ignore all terms that are 2nd-order or higher in $\delta y$, $\delta y'$. The condition $\delta I[y] = 0$ determines the **stationary values** of the functional. These include maxima and minima, but also points of inflection or *saddle points*. To check if a stationary value is an extremum, we need to calculate the change in $I[y]$ to *2nd-order* in $\delta y$. If the 2nd-order contribution $\delta^2 I[y]$ is positive, then we have a minimum; if it is negative, a maximum; and if it is zero, a saddle point. However, in the calculations that follow, we will stop at 1st order, as it will usually be obvious from the context of the problem whether our stationary solution is a maximum or minimum, without having to explicitly carry out the 2nd-order variation.

Given definition (C.2) of the functional $I[y]$, it follows that

$$\begin{aligned}
\delta I[y] &\equiv I[y + \delta y] - I[y] \\
&= \int_{x_1}^{x_2} F(y + \delta y, y' + \delta y', x)\,\mathrm{d}x - \int_{x_1}^{x_2} F(y, y', x)\,\mathrm{d}x \\
&= \int_{x_1}^{x_2} \left\{ \left(\frac{\partial F}{\partial y}\right)\delta y + \left(\frac{\partial F}{\partial y'}\right)\delta y' \right\} \mathrm{d}x\,,
\end{aligned} \tag{C.5}$$

where we ignored all 2nd-order terms to the get the last line. As mentioned earlier, the variations $\delta y$ and $\delta y'$ are not independent of one another, but are related by

$$\delta y' \equiv \delta\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right) = \frac{\mathrm{d}}{\mathrm{d}x}\delta y\,. \tag{C.6}$$

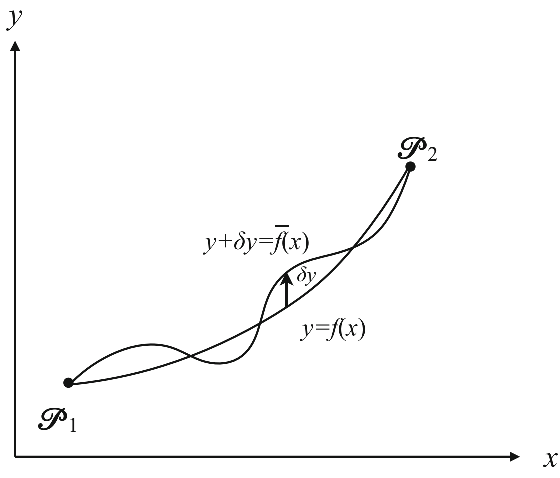



**Fig. C.2** Graphical illustration of a variation $\delta y \equiv \bar{f}(x) - f(x)$ to the function $y = f(x)$. Note that the variation must vanish at the end points in order to preserve the boundary conditions

Making this substitution and then integrating the term involving $\delta y'$ by parts, we find

$$\delta I[y] = \left(\frac{\partial F}{\partial y'}\right)\delta y\Big|_{x_1}^{x_2} + \int_{x_1}^{x_2}\left\{\left(\frac{\partial F}{\partial y}\right) - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right)\right\}\delta y\,\mathrm{d}x\,. \tag{C.7}$$

But since the variation $\delta y$ must vanish at the endpoints, i.e.,

$$\delta y|_{x_1} = 0\,, \quad \delta y|_{x_2} = 0\,, \tag{C.8}$$

the first term on the right-hand side of (C.7) is zero. Then, since the variation $\delta y$ is otherwise arbitrary, it follows that

$$\delta I[y] = 0 \quad \Leftrightarrow \quad \frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) = 0\,. \tag{C.9}$$

The equation on the right-hand side is called the **Euler equation**. (In the context of classical mechanics, where $F$ is the Lagrangian of the system, the above equation is called the **Euler-Lagrange equation**. See Chap. 3 for details.)

***Example C.1*** Using the Euler equation, show that the curve that minimizes the distance between two fixed points in a plane is a straight line.

*Proof* From (C.1) we have $F(y, y', x) = \sqrt{1 + y'^2}$, which is independent of both $x$ and $y$. Thus, the Euler equation (C.9) simplifies to

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) = 0 \quad \Leftrightarrow \quad \frac{\partial F}{\partial y'} = \text{const}\,. \tag{C.10}$$

Performing the derivative for our particular $F$, we get

$$\frac{\partial F}{\partial y'} = \frac{y'}{\sqrt{1 + y'^2}} = \text{const}\,, \tag{C.11}$$

which, after rearranging, gives $y' = A$ (another constant). So $y = Ax + B$, which is the equation of a straight line. The integration constants $A$ and $B$ are determined by the fixed endpoint conditions $y_1 = f(x_1)$ and $y_2 = f(x_2)$. □

A straight line is an example of a **geodesic**—i.e., the shortest distance path between two points. In the following two exercises, you are asked to determine the geodesics on the surface of the cylinder and the surface of a sphere. Recall that the line element on the surface of a cylinder of radius $R$ is

$$\mathrm{d}s^2 = R^2\,\mathrm{d}\phi^2 + \mathrm{d}z^2\,, \tag{C.12}$$

and the line element on the surface of a sphere of radius $R$ is

$$\mathrm{d}s^2 = R^2\,(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\,. \tag{C.13}$$

In general, for an $n$-dimensional space with coordinates $x^i \equiv (x^1, x^2, \ldots, x^n)$, the line element can be written as

$$\mathrm{d}s^2 = \sum_{i,j=1}^{n} g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j\,, \tag{C.14}$$

where $g_{ij} \equiv g_{ij}(x^1, x^2, \ldots, x^n)$.

**Exercise C.1** Show that a geodesic on the surface of a cylinder of radius $\rho = R$ is a *helix*—i.e., $z = A\phi + B$, where $A$, $B$ are constants, determined by the location of the endpoints $\wp_1 = (\phi_1, z_1)$ and $\wp_2 = (\phi_2, z_2)$. This result should not be surprising given that the surface of a cylinder is *intrinsically flat*, just like a plane in two dimensions.

**Exercise C.2** Show that a geodesic on the surface of a sphere is an arc of a *great circle*—i.e., the intersection of the surface of the sphere with a plane passing through the center of the sphere, $A\cos\phi + B\sin\phi + \cot\theta = 0$, where $A$ and $B$ are constants, determined by the end points of the curve. (*Hint*: Take $\theta$ as the independent variable for this calculation.)

**Exercise C.3** Consider the functionals

$$I_1[y] \equiv \int_{x_1}^{x_2} F(y, y', x)\,\mathrm{d}x\,, \qquad I_2[y] \equiv \int_{x_1}^{x_2} F^2(y, y', x)\,\mathrm{d}x\,, \tag{C.15}$$

where $F(y, y', x)$ is everywhere positive. As usual, assume that the functions $y = f(x)$ are fixed at the end points $x_1$ and $x_2$. Show that the Euler equations for $I_1[y]$ and $I_2[y]$ agree if and only if

$$\frac{\mathrm{d}F}{\mathrm{d}x} = 0 \quad \text{or} \quad \frac{\partial F}{\partial y'} = 0\,. \tag{C.16}$$

Thus, unlike varying an ordinary function $f(x) > 0$ for which the stationary values of $f(x)$ and $g(x) \equiv f^2(x)$ are identical, the stationary values of the functionals defined by $F(y, y', x)$ and $F^2(y, y', x)$ differ in general.

## C.3 A More Formal Discussion of the Variational Process

We can give a more formal derivation of Euler's equation and the associated variational process by writing the variation $\delta y$ of the function $y = f(x)$ as

$$\delta y(x) \equiv \bar{f}(x) - f(x) = \varepsilon\eta(x)\,, \tag{C.17}$$

where $\eta(x)$ is a function satisfying the appropriate boundary conditions (e.g., it vanishes at the endpoints), and $\varepsilon$ is a real variable, which we take to be infinitesimal for $\delta y$ to represent an infinitesimal variation of $y$. The variation of a functional $I[y]$ resulting from the above variation of $y$ is then

$$\delta I[y] \equiv I[y + \varepsilon\eta] - I[y]\,. \tag{C.18}$$

Note that since $I[y+\varepsilon\eta]$ is an *ordinary* function of the real variable $\varepsilon$, we can Taylor expand $I[y + \varepsilon\eta]$ or take its derivatives with respect to $\varepsilon$ in the usual way. For our applications, we will be particularly interested in the first derivative of $I[y + \varepsilon\eta]$ with respect to $\varepsilon$ evaluated at $\varepsilon = 0$:

$$\left.\frac{\mathrm{d}I[y+\varepsilon\eta]}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} \equiv \lim_{\varepsilon\to 0}\frac{I[y+\varepsilon\eta]-I[y]}{\varepsilon}\,. \tag{C.19}$$

In terms of this derivative, we can define the **functional derivative** of $I$, denoted $\delta I[y]/\delta y(x)$ or more simply $\delta I[y]/\delta y$, as[1]

$$\left.\frac{\mathrm{d}I[y+\varepsilon\eta]}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = \int_{x_1}^{x_2}\mathrm{d}x\,\frac{\delta I[y]}{\delta y(x)}\,\eta(x)\,, \tag{C.20}$$

where the integration is over the domain $x\in[x_1,x_2]$ of the functions $y=f(x)$ on which the functional $I[y]$ is defined. Note that the above definition is the *functional analogue* of the definition of the *directional derivative* of a function $\varphi(x^1,x^2,\ldots,x^n)$ in the direction of $\boldsymbol{\eta}$:

$$\left.\frac{\mathrm{d}\varphi(\mathbf{x}+\varepsilon\boldsymbol{\eta})}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} \equiv \lim_{\varepsilon\to 0}\frac{\varphi(\mathbf{x}+\varepsilon\boldsymbol{\eta})-\varphi(\mathbf{x})}{\varepsilon} = \sum_{i=1}^{n}\frac{\partial\varphi}{\partial x^i}\eta^i\,, \tag{C.21}$$

where integration over the continuous variable $x$ in (C.20) replaces the summation over the discrete index $i$ in (C.21); see also Appendix A.4.1 and (A.30).

If the functional $I[y]$ has the form given in (C.2), i.e.,

$$I[y]\equiv\int_{x_1}^{x_2}\mathrm{d}x\;F(y,y',x)\,, \tag{C.22}$$

where the functions $y=f(x)$ are fixed at $x_1$ and $x_2$, then the variational procedure described above leads to the same results that we found in the previous section, namely

$$\left.\frac{\mathrm{d}I[y+\varepsilon\eta]}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = \left(\frac{\partial F}{\partial y'}\right)\eta\Big|_{x_1}^{x_2} + \int_{x_1}^{x_2}\left\{\left(\frac{\partial F}{\partial y}\right)-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right)\right\}\eta\,\mathrm{d}x\,. \tag{C.23}$$

But since the function $\eta(x)$ vanishes at the endpoints, the above expression simplifies to

$$\left.\frac{\mathrm{d}I[y+\varepsilon\eta]}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = \int_{x_1}^{x_2}\left\{\frac{\partial F}{\partial y}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right)\right\}\eta\,\mathrm{d}x\,, \tag{C.24}$$

for which

[1] *A word of caution.* The functional derivative $\delta I[y]/\delta y(x)$ is a *density* in $x$, being defined *inside* an integral, (C.20). As such, the dimensions of $\mathrm{d}x\,\delta I[y]/\delta y(x)$ are the same as the dimensions of $I$ divided by the dimensions of $y$. For example, if $I[y]$ is the arc length functional, then $\delta I[y]/\delta y(x)$ has dimension of 1/length.

$$\frac{\delta I[y]}{\delta y} = \frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) . \tag{C.25}$$

Thus, in terms of a functional derivative, the Euler equation (C.9) can be written as $\delta I[y]/\delta y(x) = 0$.

**Exercise C.4** Let $I[y]$ be a functional that depends only on the value of $y$ at a particular value of $x$, e.g.,

$$I[y] \equiv y(x_0) . \tag{C.26}$$

Show that for this case the functional derivative is the Dirac delta function:

$$\frac{\delta I[y]}{\delta y(x)} = \delta(x - x_0) . \tag{C.27}$$

## C.4 Alternate Form of the Euler Equation

When $F$ does not explicitly depend upon $x$, it is convenient to work with the Euler equation in an alternative form. If we simply take the total derivative of $F$ with respect to $x$ we have

$$\frac{\mathrm{d}F}{\mathrm{d}x} = \frac{\partial F}{\partial y} y' + \frac{\partial F}{\partial y'} y'' + \frac{\partial F}{\partial x} . \tag{C.28}$$

But since

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(y' \frac{\partial F}{\partial y'}\right) = y'' \frac{\partial F}{\partial y'} + y' \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) , \tag{C.29}$$

we can rewrite (C.28) as

$$\frac{\mathrm{d}F}{\mathrm{d}x} = \frac{\partial F}{\partial x} + \frac{\mathrm{d}}{\mathrm{d}x}\left(y' \frac{\partial F}{\partial y'}\right) + y' \left[\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right)\right] . \tag{C.30}$$

Thus, using the Euler equation, (C.9), we have

$$\delta I[y] = 0 \quad \Leftrightarrow \quad 0 = \frac{\partial F}{\partial x} + \frac{\mathrm{d}}{\mathrm{d}x}\left(y' \frac{\partial F}{\partial y'} - F\right) . \tag{C.31}$$

## C.5 Possible Simplifications

The Euler equation (C.9) or its alternate form (C.31) is a 2nd-order ordinary differential equation with respect to the independent variable $x$. This equation may simplify depending on the form of $F(y, y', x)$:

1. If $F$ is independent of $y$, then $\partial F/\partial y = 0$ and the Euler equation (C.9) can be integrated to yield
$$\frac{\partial F}{\partial y'} = \text{const}\,. \tag{C.32}$$

2. If $F$ does not depend explicitly on $x$, then $\partial F/\partial x = 0$ and the alternate form of the Euler equation (C.31) can be integrated to yield
$$y'\frac{\partial F}{\partial y'} - F = \text{const}\,. \tag{C.33}$$

It turns out that simplification (2) is equivalent to making a change of the independent variable in the integrand of the functional from $x$ to $y$ using

$$\mathrm{d}x = x'\,\mathrm{d}y \quad \Leftrightarrow \quad y' = 1/x'\,, \tag{C.34}$$

so that

$$I[y] = \int_{x_1}^{x_2} F(y, y')\,\mathrm{d}x = \int_{y_1}^{y_2} F(y, 1/x')\,x'\,\mathrm{d}y \equiv \int_{y_1}^{y_2} \tilde{F}(x, x', y)\,\mathrm{d}y \equiv \tilde{I}[x]\,. \tag{C.35}$$

But since
$$\tilde{F}(x, x', y) \equiv x'\,F(y, 1/x') \tag{C.36}$$

is independent of $x$, the Euler equation for $\tilde{I}[x]$ simplifies to $\partial \tilde{F}/\partial x' = \text{const}$. But note that

$$\frac{\partial \tilde{F}}{\partial x'} = \frac{\partial}{\partial x'}\left[x'\,F(y, 1/x')\right] = F + x'\frac{\partial F}{\partial y'}\frac{\partial(1/x')}{\partial x'} = F - \frac{1}{x'}\frac{\partial F}{\partial y'} = F - y'\frac{\partial F}{\partial y'}\,, \tag{C.37}$$

which means that
$$\frac{\partial \tilde{F}}{\partial x'} = \text{const} \quad \Leftrightarrow \quad y'\frac{\partial F}{\partial y'} - F = \text{const} \tag{C.38}$$

as claimed. (Note: The change of independent variables from $x$ to $y$ assumes that the function $y = f(x)$ is invertible, so that $x = f^{-1}(y)$ and $\tilde{F}(x, x', y)$ are well-defined.)

***Example C.2*** A soap film is suspended between two circular loops of wire, as shown in Fig. C.3. Ignoring the effects of gravity, the soap film takes the shape of a surface of revolution, which has *minimimum* surface area. Thus, in terms of the function $y = f(x)$, the functional that we need to minimize is the surface area of revolution

$$I[y] = \int_{\wp_1}^{\wp_2} 2\pi y \, \mathrm{d}s = 2\pi \int_{x_1}^{x_2} y\sqrt{1 + y'^2}\, \mathrm{d}x \,, \tag{C.39}$$

which has

$$F(y, y', x) = 2\pi y\sqrt{1 + y'^2} \,. \tag{C.40}$$

But since $F$ does not depend explicitly on $x$, we can use simplification (2) to write

$$y'\frac{\partial F}{\partial y'} - F = y'\frac{2\pi y y'}{\sqrt{1 + y'^2}} - 2\pi y\sqrt{1 + y'^2} = -\frac{2\pi y}{\sqrt{1 + y'^2}} = \text{const} \,. \tag{C.41}$$

Rewriting this constant as $-2\pi A$ and solving for $y'$ yields

$$y' \equiv \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{A}\sqrt{y^2 - A^2} \,. \tag{C.42}$$

This is a separable equation, which can be integrated using the hyperbolic trig substitutions $y = A\cosh u$, recalling that $\cosh^2 u - \sinh^2 u = 1$, and $\mathrm{d}\cosh u = \sinh u \, \mathrm{d}u$. Thus,

$$x = A\int \frac{\mathrm{d}y}{\sqrt{y^2 - A^2}} + B = A\int \frac{A\sinh u \, \mathrm{d}u}{A\sinh u} + B = Au + B = A\cosh^{-1}(y/A) + B \,, \tag{C.43}$$

or, equivalently,

$$y = A\cosh\left(\frac{x - B}{A}\right) . \tag{C.44}$$

Such a curve is called a **catenary**. As usual, the integration constants $A$ and $B$ can be determined by the boundary conditions for $y = f(x)$, which are related to the radii of the two circular loops of wire. Unfortunately, solving for $A$ and $B$ involves solving a transcendental equation. See Chap. 17 of Arfken (1970) for a discussion of special cases of this problem. □

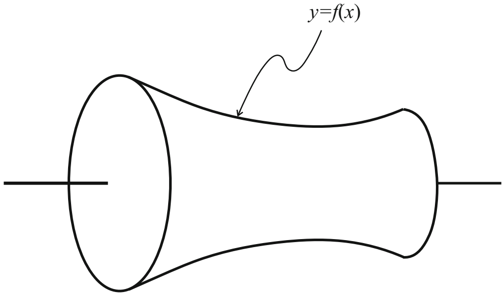



**Fig. C.3** Soap film suspended between two circular loops of wire. If we ignore the effects of gravity, the soap film takes the shape of a surface of revolution about the horizontal axis, which has minimum surface area

**Exercise C.5** Find the shape of a wire joining two points such that a bead will slide along the wire under the influence of gravity (without friction) in the shortest amount of time. (See Fig. C.4.) Assume that the bead is released from rest at $y = 0$. Such a curve is called a **brachistochrone**, which in Greek means "shortest time."

*Hint*: You should extremize the functional

$$I[y] = \int_{\wp_1}^{\wp_2} \frac{ds}{v} = \int_{x_1}^{x_2} dx\, \frac{\sqrt{1+y'^2}}{\sqrt{2gy}}\,, \tag{C.45}$$

where conservation of energy

$$\frac{1}{2}mv^2 - mgy = 0 \quad \Rightarrow \quad v = \sqrt{2gy} \tag{C.46}$$

was used to yield an expression for the speed $v$ in terms of $y$. By using simplification (2) or changing the independent variable of the functional from $x$ to $y$, you should find

$$y' \equiv \frac{dy}{dx} = \sqrt{\frac{1-Ay}{Ay}}\,. \tag{C.47}$$

which has solution

$$x = \frac{1}{2A}(\theta - \sin\theta)\,, \qquad y = \frac{1}{2A}(1-\cos\theta)\,. \tag{C.48}$$

This is the parametric representation of a **cycloid** (i.e., the path traced out by a point on the rim of a wheel as it rolls without slipping across a horizontal surface). See Fig. C.5.

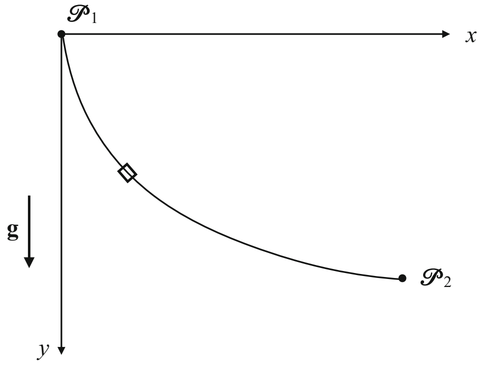



**Fig. C.4** Geometrical set-up for the brachistochrone problem, Exercise C.5. The goal is to find the shape of the wire connecting points $\wp_1$ and $\wp_2$ such that a bead slides along the wire under the influence of gravity (and in the absence of friction) in the shortest amount of time. Note that we have chosen the $y$-axis to increase in the downward direction

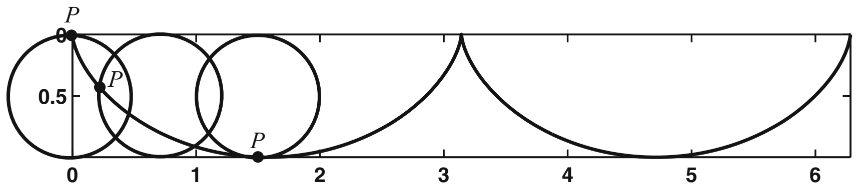



**Fig. C.5** A cycloid is the path traced out by a point on the rim of a wheel as it rolls without slipping across a flat surface. It is also the shape of the wire that solves the brachistochrone problem, Exercise C.5. To be consistent with the geometry of Exercise C.5 shown in Fig. C.4, we are considering the wheel as rolling to the right in contact with the *top* horizontal surface

**Exercise C.6** (*Adapted from Kuchǎr* 1995.) Consider a two-dimensional surface of revolution obtained by rotating the curve $z = f(\rho)$ around the $z$-axis, where $\rho \equiv \sqrt{x^2 + y^2}$. An example of such a surface is shown in panel (a) of Fig. C.6, which is a *paraboloid*, defined by $f(\rho) = \rho^2$. Surfaces of revolution are most conveniently described by embedding equations

$$x = \rho \cos\phi\,, \qquad y = \rho \sin\phi\,, \qquad z = f(\rho)\,, \tag{C.49}$$

where $\phi$ is the standard azimuthal angle in the $xy$-plane.

(a) Write down the line element $\mathrm{d}s^2$ on the surface of revolution in terms of the coordinates $(\rho, \phi)$ by simply substituting the embedding equations into the 3-dimensional line element $\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$.

(b) By varying the arc length functional, obtain the geodesic equation for a curve $\rho = \rho(\phi)$ on the surface of revolution, and show that it can be solved via quadratures,

$$\phi - \phi_0 = c_1 \int_{\rho_0}^{\rho} \frac{\mathrm{d}\rho}{\rho} \sqrt{\frac{1 + [f'(\rho)]^2}{\rho^2 - c_1^2}}\,, \tag{C.50}$$

where $c_1$ is a constant.

(c) Evaluate the above integral for the case of a surface of a cone with half-angle $\alpha$, which is defined by $f(\rho) = \rho \cot \alpha$. (See panel (b) of Fig. C.6.) You should find

$$\rho = \frac{c_1}{\cos(\phi \sin \alpha + c_2)}\,, \tag{C.51}$$

where $c_1$, $c_2$ are constants determined by the boundary conditions.

(d) Show that the above solution is equivalent to a straight line in a flat 2-dimensional space with Cartesian coordinates

$$\bar{x} \equiv \rho \cos(\phi \sin \alpha)\,, \qquad \bar{y} \equiv \rho \sin(\phi \sin \alpha)\,. \tag{C.52}$$

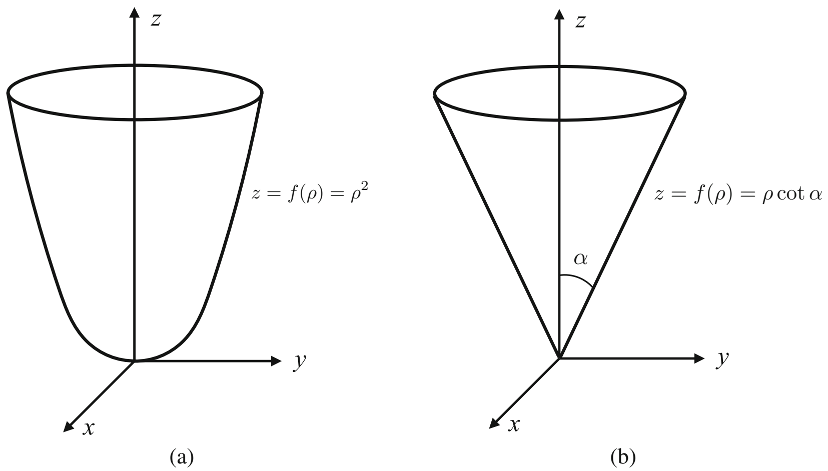



**Fig. C.6** Examples of surfaces of revolution, obtained by rotating a curve $z = f(\rho)$ around the $z$-axis. Panel (a) A *paraboloid* defined by $f(\rho) \equiv \rho^2$. Panel (b) A *cone* with half-angle $\alpha$, defined by $f(\rho) = \rho \cot \alpha$

## C.6 Variational Problem in Parametric Form

Although we have been considering functionals of the form given by (C.2), where the curve is explicitly described by the function $y = f(x)$, there may be cases where it is more convenient (or even necessary) to describe the curve in parametric form, e.g., $x = x(t)$, $y = y(t)$. Such an example is the variational problem to find the shape of a *closed* curve of fixed length that encloses the greatest area. (We will revisit this in more detail in Appendix C.8.) Here we derive the necessary and sufficient conditions for a functional to depend only on the curve in the $xy$-plane and not on the choice of parametric representation of the curve. The relevant theorem is:

**Theorem C.1** *The necessary and sufficient conditions for the functional*

$$I[x, y] = \int_{t_1}^{t_2} G(x, y, \dot{x}, \dot{y}, t)\, dt \tag{C.53}$$

*to depend only on the curve in the $xy$-plane and not on the choice of parametric representation of the curve is that $G$ not depend explicitly on $t$ and be a positive-homogeneous function of degree one in $\dot{x}$ and $\dot{y}$—i.e.,*

$$G(x, y, \lambda\dot{x}, \lambda\dot{y}) = \lambda G(x, y, \dot{x}, \dot{y}) \tag{C.54}$$

*for all $\lambda > 0$.*

*Proof* Start with the functional[2]

$$I[y] = \int_{x_1}^{x_2} F(y, y', x)\, \mathrm{d}x\,, \tag{C.55}$$

which by its definition depends only the curve traced out by the function $y = f(x)$. We then introduce a parameter $t$ so that $x = x(t)$, $y = y(t)$. Then

$$\mathrm{d}x = \dot{x}\, \mathrm{d}t\,, \quad \mathrm{d}y = \dot{y}\, \mathrm{d}t\,, \quad x' = \frac{\mathrm{d}x}{\mathrm{d}y} = \frac{\dot{x}}{\dot{y}}\,, \quad y' = \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\dot{y}}{\dot{x}}\,, \tag{C.56}$$

and

$$\int_{x_1}^{x_2} F(y, y', x)\, \mathrm{d}x = \int_{t_1}^{t_2} F(y, \dot{y}/\dot{x}, x)\dot{x}\, \mathrm{d}t\,. \tag{C.57}$$

Thus,

[2] This derivation closely follows that given in Sect. 10 of Gelfand and Fomin (1963).

$$I[y] = \int_{x_1}^{x_2} F(y, y', x)\, \mathrm{d}x = \int_{t_1}^{t_2} G(x, y, \dot{x}, \dot{y})\, \mathrm{d}t \equiv I[x, y]\,, \tag{C.58}$$

where

$$G(x, y, \dot{x}, \dot{y}) = F(y, \dot{y}/\dot{x}, x)\dot{x}\,. \tag{C.59}$$

Note that $G$ has the properties that it does not depend explicitly on $t$, and that

$$G(x, y, \lambda\dot{x}, \lambda\dot{y}) = \lambda G(x, y, \dot{x}, \dot{y}) \tag{C.60}$$

for all $\lambda > 0$. Thus, $G$ is a *positive-homogeneous function of degree one* in $\dot{x}$ and $\dot{y}$.

Conversely, suppose that we have a functional of the form

$$I[x, y] = \int_{t_1}^{t_2} G(x, y, \dot{x}, \dot{y})\, \mathrm{d}t\,, \tag{C.61}$$

where $G$ does not explicitly depend on $t$ and is a positive-homogeneous function of degree one in $\dot{x}$ and $\dot{y}$. Then if we change the parametrization from $t$ to a new parameter $\tau$, we have

$$\mathrm{d}t = \frac{\mathrm{d}t}{\mathrm{d}\tau}\, \mathrm{d}\tau\,, \quad \dot{x} = \frac{\mathrm{d}x}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}t}\,, \quad \dot{y} = \frac{\mathrm{d}y}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}t}\,, \tag{C.62}$$

and

$$G(x, y, \dot{x}, \dot{y}) = G\left(x, y, \frac{\mathrm{d}x}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}t}, \frac{\mathrm{d}y}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}t}\right) = G\left(x, y, \frac{\mathrm{d}x}{\mathrm{d}\tau}, \frac{\mathrm{d}y}{\mathrm{d}\tau}\right)\frac{\mathrm{d}\tau}{\mathrm{d}t}\,, \tag{C.63}$$

where the last equality used the positive-homogeneous-of-degree-one property of $G$. Thus,

$$I[x, y] = \int_{t_1}^{t_2} G(x, y, \dot{x}, \dot{y})\, \mathrm{d}t = \int_{\tau_1}^{\tau_2} G\left(x, y, \frac{\mathrm{d}x}{\mathrm{d}\tau}, \frac{\mathrm{d}y}{\mathrm{d}\tau}\right) \mathrm{d}\tau\,, \tag{C.64}$$

which shows that the functional is independent of the parametric representation of the curve. □

***Example C.3*** Here we show explicitly that the Euler equations obtained from the parametrized functional

$$I[x, y] \equiv \int_{t_1}^{t_2} G(x, y, \dot{x}, \dot{y})\, \mathrm{d}t = \int_{t_1}^{t_2} F(y, \dot{y}/\dot{x}, x)\dot{x}\, \mathrm{d}t \tag{C.65}$$

reduce to the standard Euler equation (C.9) obtained from

$$I[y] = \int_{x_1}^{x_2} F(y, y', x)\, \mathrm{d}x\,. \tag{C.66}$$

*Proof* The Euler equations obtained from (C.65) by varying both $x$ and $y$ are

$$\frac{\partial G}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial G}{\partial \dot{x}}\right) = 0\,, \tag{C.67a}$$

$$\frac{\partial G}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial G}{\partial \dot{y}}\right) = 0\,. \tag{C.67b}$$

Since $G$ does not depend explicitly on $t$, then by an extension of (C.33) to two variables (See also Appendix C.7.3), we also have

$$\dot{x}\frac{\partial G}{\partial \dot{x}} + \dot{y}\frac{\partial G}{\partial \dot{y}} - G = \text{const}\,. \tag{C.68}$$

Differentiating this last equation with respect to $t$ yields

$$\dot{x}\left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial G}{\partial \dot{x}}\right) - \frac{\partial G}{\partial x}\right] + \dot{y}\left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial G}{\partial \dot{y}}\right) - \frac{\partial G}{\partial y}\right] = 0\,, \tag{C.69}$$

where we have cancelled out the terms involving $\ddot{x}$ and $\ddot{y}$. This last equation shows that the two equations (C.67a) and (C.67b) are *not* independent, but follow one from the other. So, without loss of generality, let's consider (C.67b). Then by writing $G$ in terms of $F$ and performing the derivatives, we find

$$\begin{aligned}
0 &= \frac{\partial G}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial G}{\partial \dot{y}}\right) \\
&= \frac{\partial}{\partial y}\left[F(y, \dot{y}/\dot{x}, x)\dot{x}\right] - \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial}{\partial \dot{y}}\left[F(y, \dot{y}/\dot{x}, x)\dot{x}\right]\right) \\
&= \frac{\partial F}{\partial y}\dot{x} - \dot{x}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\frac{1}{\dot{x}}\dot{x}\right) \\
&= \dot{x}\left[\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right)\right],
\end{aligned} \tag{C.70}$$

which is proportional to the standard Euler equation (C.9). □

**Exercise C.7** Consider the parametrized form of the arc length functional in two dimensions,

$$I[x_1, x_2] \equiv \int_{\wp_1}^{\wp_2} \mathrm{d}s = \int_{t_1}^{t_2} G(x_1, x_2, \dot{x}_1, \dot{x}_2)\,\mathrm{d}t\,, \tag{C.71}$$

where

$$G(x_1, x_2, \dot{x}_1, \dot{x}_2) = \frac{\mathrm{d}s}{\mathrm{d}t} = \sqrt{\sum_{i,j} g_{ij}\,\dot{x}_i\,\dot{x}_j}\,. \tag{C.72}$$

Note that by writing the arc length in this form (See (C.14)), we are allowing for the possibility that the 2-dimensional space be curved (e.g., the surface of a sphere) and that the coordinates need not be Cartesian, so $g_{ij} \equiv g_{ij}(x_1, x_2)$ in general.

(a) Show that the Euler equations for this functional are

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_j g_{ij}\dot{x}_j\right) - \frac{1}{2}\sum_{j,k}\frac{\partial g_{jk}}{\partial x_i}\dot{x}_j\dot{x}_k = \frac{1}{G}\left(\frac{\mathrm{d}G}{\mathrm{d}t}\right)\sum_j g_{ij}\dot{x}_j\,. \tag{C.73}$$

These are the geodesic equations in an *arbitrary* parametrization.

(b) Show that if we choose the parameter $t$ to be linearly related to the arc length $s$ along the curve,

$$t = as + b\,, \qquad a, b = \text{const}\,, \tag{C.74}$$

then $G = \text{const}$, and the geodesic equation simplifies to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_j g_{ij}\dot{x}_j\right) - \frac{1}{2}\sum_{j,k}\frac{\partial g_{jk}}{\partial x_i}\dot{x}_j\dot{x}_k = 0\,. \tag{C.75}$$

Such a parametrization of the curve is called an **affine parametrization**.

(c) Show that one obtains the same simplified form of the geodesic equation by varying instead the *kinetic energy functional*,

$$J[x_1, x_2] \equiv \int_{t_1}^{t_2} K(x_1, x_2, \dot{x}_1, \dot{x}_2)\,\mathrm{d}t\,, \tag{C.76}$$

where

$$K(x_1, x_2, \dot{x}_1, \dot{x}_2) \equiv \frac{1}{2}\sum_{i,j} g_{ij}\,\dot{x}_i\,\dot{x}_j = \frac{1}{2}G^2(x_1, x_2, \dot{x}_1, \dot{x}_2)\,. \tag{C.77}$$

The equivalence of these two approaches follows from the fact that $G$ and $K$ differ by an overall multiplicative constant when $t$ is an affine parameter. If $t$ is *not* an affine parameter, then the two functionals $I$ and $J$ lead to the different equations of motion, consistent with the results of Exercise C.3. (Note that these results hold, in general, in $n$ dimensions.)

## C.7 Generalizations

The standard calculus of variations problem (C.2) discussed in the preceding sections can be extended in several ways. Here we describe three such extensions.

### *C.7.1 Functionals that Depend on Higher-Order Derivatives*

Consider a functional of the form

$$I[y] = \int_{x_1}^{x_2} F(y, y', y'', x)\,\mathrm{d}x\,, \tag{C.78}$$

where the set of functions $\{y = f(x)\}$ is now restricted so that *both* $y$ and $y'$ are fixed at the endpoints $\wp_1$ and $\wp_2$. Then proceeding in a manner similar to that in Appendix C.2, we find

$$\delta I[y] = 0 \quad \Leftrightarrow \quad \frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) + \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{\partial F}{\partial y''}\right) = 0\,. \tag{C.79}$$

Note that the Euler equation for this case may contain *3rd* or even *4th-order* derivatives of $y = f(x)$. Although most problems in classical mechanics involve *2nd-order* differential equations, 3rd or higher-order differential equations have applications in certain areas of chaos theory (See, e.g., Goldstein et al. 2002).

### *C.7.2 Allowing Variations with Free Endpoints*

The standard variational problem (C.2) with fixed endpoints can be generalized by allowing the variations $\delta y$ to be *non-zero* at either one or both endpoints $\wp_1$, $\wp_2$. The derivation given in Appendix C.2 then leads to

$$\delta I[y]=0 \quad \Leftrightarrow \quad \frac{\partial F}{\partial y}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right)=0\,, \quad \left.\frac{\partial F}{\partial y'}\right|_{x_1}=0\,, \quad \left.\frac{\partial F}{\partial y'}\right|_{x_2}=0\,. \tag{C.80}$$

The conditions

$$\left.\frac{\partial F}{\partial y'}\right|_{x_1}=0\,, \qquad \left.\frac{\partial F}{\partial y'}\right|_{x_2}=0\,, \tag{C.81}$$

are sometimes called **natural boundary conditions** for the curve. In the context of classical mechanics, the natural boundary conditions for a particle moving in response to a velocity-independent conservative force correspond to *zero velocity* at the endpoints. The only solution to the equations of motion that satisfies these boundary conditions at *both* endpoints is the trivial solution, where the particle just sits at one location forever. Imposing the natural boundary condition at just *one* endpoint and fixed boundary conditions at the other allows for non-trivial solutions, in general.

**Exercise C.8** Redo the brachistochrone problem (Exercise C.5), but this time allowing the second endpoint at $x_2$ to be free—i.e., $\delta y|_{x_2} \neq 0$. You should find that the solution is again a cycloid, but which intersects the line $x = x_2$ at a *right angle*.

### *C.7.3 Generalization to Several Dependent Variables*

The derivation of the Euler equation given in Appendix C.2 can be easily be extended to functionals of the form

$$I[y_1,\dots,y_n] \equiv \int_{x_1}^{x_2} F(y_1,\dots,y_n;y_1',\dots,y_n';x)\,\mathrm{d}x\,, \tag{C.82}$$

where $y_i \equiv f_i(x)$, $i = 1, 2, \dots, n$, are $n$ functions of the independent variable $x \in [x_1, x_2]$, which we require to be fixed at the endpoints $\wp_1$ and $\wp_2$. The corresponding Euler equations obtained from the variational principle $\delta I[y_1,\dots,y_n] = 0$ with respect to variations $\delta y_i$ that vanish at the endpoints are

$$\frac{\partial F}{\partial y_i}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y_i'}\right)=0\,, \qquad i=1,2,\dots,n\,. \tag{C.83}$$

These equations form a system of $n$ 2nd-order ordinary differential equations for the functions $y_i = f_i(x)$ with respect to the independent variable $x$.

The simplifications discussed in Appendix C.5 carry over to the general case of $n$ degrees of freedom:

1. If $F$ is independent of a particular $y_{\underline{i}}$, then $\partial F/\partial y_{\underline{i}} = 0$ and the corresponding Euler equation can be integrated to yield

$$\frac{\partial F}{\partial y'_{\underline{i}}} = \text{const} . \tag{C.84}$$

2. If $F$ does not depend explicitly on $x$, then

$$h \equiv \sum_{i=1}^{n} y'_i \frac{\partial F}{\partial y'_i} - F = \text{const} . \tag{C.85}$$

To prove the second result above:

$$\frac{\mathrm{d}h}{\mathrm{d}x} = \sum_{i=1}^{n} y''_i \frac{\partial F}{\partial y'_i} + \sum_{i=1}^{n} y'_i \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'_i}\right) - \sum_{i=1}^{n}\left[\frac{\partial F}{\partial y_i} y'_i + \frac{\partial F}{\partial y'_i} y''_i\right] = 0 , \tag{C.86}$$

where we assumed that $F$ does not depend explicitly on $x$ (i.e., $\partial F/\partial x = 0$) and used the Euler equations (C.83) to get the last equality.

There is also an additional simplification if $F$ does not depend on the derivative of one of the variables:

3. If $F$ is independent of a particular derivative, which we will take (without loss of generality) to be $y'_n$, then the Euler equation for $y_n$ becomes $\partial F/\partial y_n = 0$, which can be solved algebraically for $y_n$ in terms of all of the other variables and their derivatives (assuming $\partial^2 F/\partial y_n^2 \neq 0$). The Euler equations for all the other variables can then be obtained from the *reduced* functional

$$\underline{I}[y_1, \dots, y_{n-1}] \equiv \int_{x_1}^{x_2} \underline{F}(y_1, \dots, y_{n-1}; y'_1, \dots, y'_{n-1}; x)\, \mathrm{d}x , \tag{C.87}$$

where

$$\begin{aligned} &\underline{F}(y_1, \dots, y_{n-1}; y'_1, \dots, y'_{n-1}; x) \\ &\quad \equiv F(y_1, \dots, y_n; y'_1, \dots, y'_{n-1}; x)\big|_{y_n = y_n(y_1,\dots,y_{n-1}; y'_1,\dots,y'_{n-1}; x)} \end{aligned} \tag{C.88}$$

**Exercise C.9** Verify simplification (3) above by showing that

$$\frac{\partial \underline{F}}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial \underline{F}}{\partial y_i'}\right) = 0 \quad \Rightarrow \quad \frac{\partial F}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y_i'}\right) = 0\,, \tag{C.89}$$

for $i = 1, 2, \ldots, n-1$ as a consequence of $\partial F/\partial y_n = 0$.

## C.8 Isoperimetric Problems

So far, we've been considering variational problems of the form $\delta I[y] = 0$, where the functions $y = f(x)$ have been subject only to *boundary conditions*, e.g., fixed at the endpoints $x = x_1$ and $x_2$. But there might also exist situations where the functions $y = f(x)$ are subject to an *integral constraint*

$$J[y] \equiv \int_{x_1}^{x_2} G(y, y', x)\,\mathrm{d}x = J_0\,, \tag{C.90}$$

where $J_0$ is a constant. Such problems are called **isoperimetric problems**, since the classic example of such a problem is to find the shape of a closed curve of *fixed* length (perimeter) $\ell$ that encloses the maximum area. Due to the constraint, the variations of $y$ in $\delta I = 0$ are not free, but are subject to the condition that $\delta J[y] = 0$. We can incorporate this condition into the variational problem by using the method of *Lagrange multipliers* (See Sect. 2.4). This amounts to adding to $\delta I = 0$ a multiple of $\delta J = 0$,

$$\delta I[y] + \lambda \delta J[y] = 0\,, \tag{C.91}$$

where $\lambda$ is an undetermined constant (the Lagrange multiplier for this problem). Note that (C.91) can be recast as finding the stationary values of the functional

$$\bar{I}[y, \lambda] \equiv I[y] + \lambda\,(J[y] - J_0) = \int_{x_1}^{x_2} (F + \lambda G)\;\mathrm{d}x - \lambda\, J_0 \tag{C.92}$$

with respect to *unconstrained* variations of both $y$ and $\lambda$. (The variation with respect to $\lambda$ recovers the integral constraint $J[y] - J_0 = 0$.) Performing the variations give rise to two equations, which can be solved for the two unknowns $y = f(x)$ and $\lambda$ (if desired).

***Example C.4*** Here we will find the shape of a closed curve of fixed length $\ell$ that encloses the largest area. Since the curve is closed, we will not be able to represent it as a single-valued function $y = f(x)$ or $x = g(y)$. Instead, we have to represent it *parametrically*—i.e., by $x = x(t)$, $y = y(t)$, where $t \in [t_1, t_2]$ is a parameter along the curve. Without loss of generality, we can orient the curve in the $xy$-plane so that

$$x|_{t_1,t_2} = 0\,, \quad y|_{t_1,t_2} = 0\,, \quad \dot{x}|_{t_1,t_2} = -v\,, \quad \dot{y}|_{t_1,t_2} = 0\,, \tag{C.93}$$

with the derivatives so chosen as to avoid a kink at the origin. (See Fig. C.7.)

The functional that we want to extremize is the area under the curve

$$I[x, y] = \oint y\,\mathrm{d}x = \int_{t_1}^{t_2} y\dot{x}\,\mathrm{d}t\,, \tag{C.94}$$

subject to the constraint

$$J[x, y] = \oint \mathrm{d}s = \int_{t_1}^{t_2} \sqrt{\dot{x}^2 + \dot{y}^2}\,\mathrm{d}t = \ell\,. \tag{C.95}$$

The curve is traversed clockwise so that the area obtained is enclosed by the curve.

Using the method of Lagrange multipliers discussed above, we extremize the combined functional

$$\bar{I}[x, y, \lambda] = \int_{t_1}^{t_2} (F + \lambda G)\,\mathrm{d}t - \lambda\ell\,, \tag{C.96}$$

where

$$F + \lambda G = y\dot{x} + \lambda\sqrt{\dot{x}^2 + \dot{y}^2}\,. \tag{C.97}$$

Note that $F + \lambda G$ does not explicitly depend on $t$ and is a positive-homogeneous function of degree one in $\dot{x}$ and $\dot{y}$. Thus, from the discussion of Appendix C.6, the solution to the variational problem will depend only on the curve in the $xy$-plane and not on a particular parametric representation of the curve.

The Euler equations obtained from $F + \lambda G$ by varying $x$ and $y$ are

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(y + \frac{\lambda\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}\right) = 0 \quad \Rightarrow \quad y + \frac{\lambda\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}} = A\,, \tag{C.98}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\lambda\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}\right) - \dot{x} = 0 \quad \Rightarrow \quad \frac{\lambda\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}} - x = B\,, \tag{C.99}$$

where $A$ and $B$ are constants. These equations can be simplified if we switch the parametric representation from $t$ to arc length $s$, noting that

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{1}{\sqrt{\dot{x}^2 + \dot{y}^2}}\,. \tag{C.100}$$

The right-hand sides of (C.98) and (C.99) then become

$$y + \lambda \frac{\mathrm{d}x}{\mathrm{d}s} = A\,, \qquad \lambda \frac{\mathrm{d}y}{\mathrm{d}s} - x = B\,. \tag{C.101}$$

We now apply the boundary conditions of (C.93) but in terms of arc length $s$,

$$x|_{s=0,\ell} = 0\,, \quad y|_{s=0,\ell} = 0\,, \quad \left.\frac{\mathrm{d}x}{\mathrm{d}s}\right|_{s=0,\ell} = -1\,, \quad \left.\frac{\mathrm{d}y}{\mathrm{d}s}\right|_{s=0,\ell} = 0\,, \tag{C.102}$$

which lead to $B = 0$ and $A = -\lambda$. The first equation in (C.101) can then be solved for $y$ in terms of $\mathrm{d}x/\mathrm{d}s$,

$$y = -\lambda \left(1 + \frac{\mathrm{d}x}{\mathrm{d}s}\right)\,, \tag{C.103}$$

and substituted back into the second equation. This leads to

$$\frac{\mathrm{d}^2 x}{\mathrm{d}s^2} = -\lambda^{-2}\, x\,, \tag{C.104}$$

with solution

$$x(s) = -|\lambda| \sin(s/|\lambda|)\,, \tag{C.105}$$

which satisfies the boundary conditions above. Using (C.103) we have

$$y(s) = -\lambda \left(1 - \cos(s/|\lambda|)\right)\,. \tag{C.106}$$

Note that these last two equations are the parametric representation of a circle

$$x^2 + (y + \lambda)^2 = \lambda^2\,, \tag{C.107}$$

with radius $R = |\lambda|$ and center $(0, -\lambda)$, In order that the circle lie above the $x$-axis, as suggested by Fig. C.7, we need $\lambda = -R$. Finally, since $2\pi R = \ell$ is the length of the curve, the Lagrange multiplier $\lambda = -\ell/2\pi$. Thus,

$$x = -\frac{\ell}{2\pi} \sin(2\pi s/l)\,, \qquad y = \frac{\ell}{2\pi}\left(1 - \cos(2\pi s/l)\right)\,. \tag{C.108}$$

The area enclosed by the circle is $A = \pi R^2 = \ell^2/4\pi$, which is the largest area enclosed by a closed curve of fixed length $\ell$. □

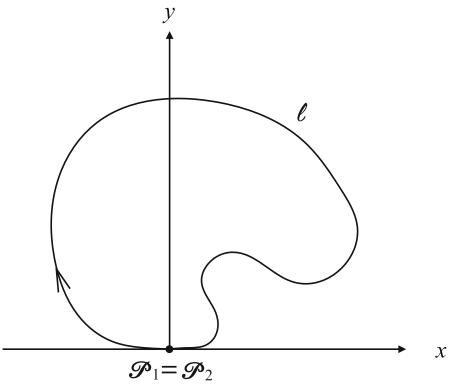



**Fig. C.7** Closed curve of fixed length $\ell$ in the $xy$-plane. The coordinates and parametrization are chosen so that the curve starts and stops at the origin and is tangent to the $x$-axis at the origin. The curve is traversed in the clockwise direction so that the area calculated in (C.94) is that *enclosed* by the curve

**Exercise C.10** Find the shape of the curve of fixed length $\ell$ and *free* endpoint $\wp_2$ on the $x$-axis that encloses the largest area between it and the $x$-axis. (See Fig. C.8.) You should find:

$$x = \frac{\ell}{\pi}\,(1 - \cos(\pi s/l)) \ , \qquad y = -\frac{\ell}{\pi}\,\sin(\pi s/l)\,, \tag{C.109}$$

which is the parametric representation of a *semi-circle* of radius $R = \ell/\pi$, length $\ell$, and center $(\ell/\pi, 0)$. Note that the area enclosed by this curve and the $x$-axis is $\ell^2/2\pi$, which is *twice* as large as that calculated in Example C.4 for the closed-curve variational problem.

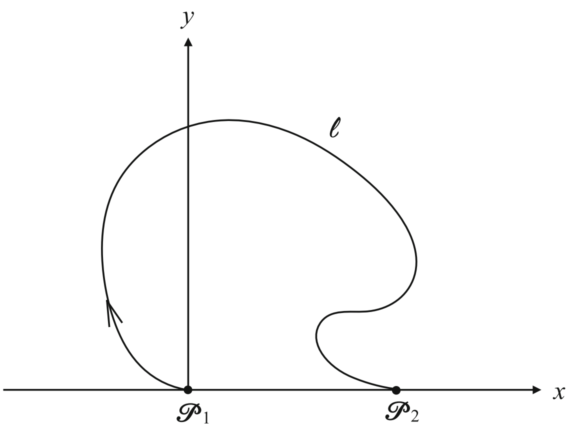



**Fig. C.8** Curve of fixed length $\ell$, having one endpoint $\wp_1$ fixed at the origin and the other endpoint $\wp_2$ *freely-variable* in the $x$-direction

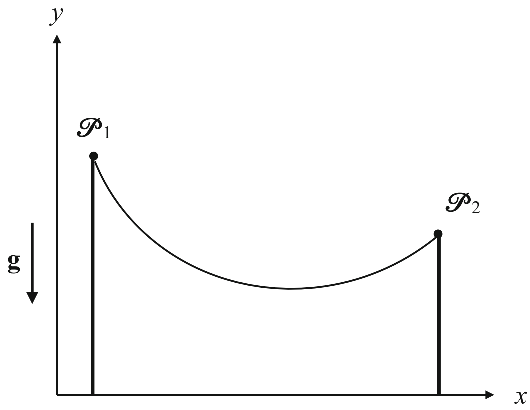



**Fig. C.9** Flexible hanging cable of fixed length $\ell$ supported at endpoints $\wp_1$ and $\wp_2$. A uniform gravitational field **g** points downward

**Exercise C.11** Find the shape of a flexible hanging cable of fixed length $\ell$, which is supported at endpoints $\wp_1$ and $\wp_2$ in a uniform gravitational field **g** pointing downward. (See Fig. C.9.)
*Hint*: The shape minimizes the gravitational potential energy of the cable

$$I[x] = \int_{\wp_1}^{\wp_2} \mathrm{d}m\, gy = \int_{\wp_1}^{\wp_2} \mu \mathrm{d}s\, gy = \mu g \int_{x_1}^{x_2} y\sqrt{1 + y'^2}\, \mathrm{d}x\,, \qquad \text{(C.110)}$$

where $\mu$ is the mass-per-unit-length of the cable (assumed constant), subject to the constraint that the cable has fixed length $\ell$, i.e.,

$$J[y] = \int_{\wp_1}^{\wp_2} \mathrm{d}s = \int_{x_1}^{x_2} \sqrt{1 + y'^2}\, \mathrm{d}x = \ell\,. \qquad \text{(C.111)}$$

You should find that the solution has the form of a catenary, $y \sim \cosh x$.

## Suggested References

*Full references are given in the bibliography at the end of the book.*

Boas (2006): Chapter 9 is devoted solely to the calculus of variations; an excellent introduction to the topic well-suited for undergraduates with many examples and problems. Solutions to several of the problems presented in this appendix can be found in Boas (2006).

Gelfand and Fomin (1963): A more rigorous mathematical treament of the calculus of variations. Our discussion of variational problems in parametric form follows closely the presentation given in Sect. 10 of this book.

Lanczos (1949): In our opinion, the best book on variational methods in the context of classical mechanics. It contains excellent descriptions/explanations of the calculus of variations, constrained systems, the method of Lagrange multipliers, etc. Suitable for either advanced undergraduates or graduate students.

# Appendix D
# Linear Algebra

Linear algebra can be thought of as an extension of the mathematical structure of ordinary (3-dimensional) vectors and matrices to an *arbitrary* number of dimensions. The general mathematical framework of linear algebra is particularly relevant for the matrix calculations required for describing rigid-body motion (Chaps. 6 and 7), and for calculating the normal modes associated with small oscillations (Chap. 8). Although not strictly necessary for classical mechanics, we will allow our vectors and matrices to be *complex-valued*, since this generalization requires limited additional work, and it turns out to be extremely useful for quantum mechanics, where complex numbers are the rule, not the exception. We will, however, restrict ourselves to a *finite* number of dimensions $n$, although it is also possible to have infinite-dimensional vector spaces (e.g., function spaces).

Since we will only be summarizing key results here and not giving detailed proofs, we encourage readers to refer to other texts, e.g., Boas (2006); Dennery and Kryzwicki (1967); Griffiths (2005); Halmos (1958) to fill in the missing details. Our approach in this appendix is similar to that of Appendix A in (Griffiths, 2005).

## D.1 Vector Space

An abstract vector space consists of two types of objects (**vectors** and **scalars**) and two types of operations (**vector addition** and **scalar multiplication**), which interact with one another and are subject to certain properties (enumerated below). We will denote vectors by boldface symbols, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\cdots$, and scalars (which we will take to be complex numbers) by italicized symbols, $a, b, c, \cdots$. Vector addition will be denoted by a $+$ sign between two vectors, e.g., $\mathbf{A}+\mathbf{B}$, and scalar multiplication by juxtaposition of a scalar and a vector, e.g., $a\mathbf{A}$. The properties obeyed by these operations are as follows.



M.J. Benacquista and J.D. Romano, *Classical Mechanics*, Undergraduate Lecture Notes in Physics, https://doi.org/10.1007/978-3-319-68780-3

### *D.1.1 Vector Addition*

1. Closure: The addition of two vectors is also a vector:

$$\mathbf{A} + \mathbf{B} = \mathbf{C} \tag{D.1}$$

2. Commutativity:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \tag{D.2}$$

3. Associativity:

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \tag{D.3}$$

4. Zero vector:

$$\exists \mathbf{0} \quad \text{such that} \quad \mathbf{A} + \mathbf{0} = \mathbf{A} \quad \forall \mathbf{A} \tag{D.4}$$

5. Inverse vector:

$$\forall \mathbf{A} \quad \exists - \mathbf{A} \quad \text{such that} \quad \mathbf{A} + (-\mathbf{A}) = \mathbf{0} \tag{D.5}$$

### *D.1.2 Scalar Multiplication*

1. Closure: The multiplication of a scalar and a vector is also a vector:

$$a\mathbf{A} = \mathbf{B} \tag{D.6}$$

2. Identity: The scalar 1 is the identity operator on vectors:

$$1\mathbf{A} = \mathbf{A} \tag{D.7}$$

3. Scalar multiplication is distributive with respect to scalar addition:

$$(a + b)\mathbf{A} = a\mathbf{A} + b\mathbf{A} \tag{D.8}$$

4. Scalar multiplication is distributive with respect to vector addition:

$$a(\mathbf{A} + \mathbf{B}) = a\mathbf{A} + a\mathbf{B} \tag{D.9}$$

5. Scalar multiplication is associative with respect to scalar multiplication:

$$a(b\mathbf{A}) = (ab)\mathbf{A} \tag{D.10}$$

Using the various properties given above, it is easy to see that $0\mathbf{A} = \mathbf{0}$ and $(-1)\mathbf{A} = -\mathbf{A}$, since

$$(\mathbf{A} + 0\mathbf{A}) = (1\mathbf{A} + 0\mathbf{A}) = (1+0)\mathbf{A} = 1\mathbf{A} = \mathbf{A}\,, \tag{D.11}$$

and

$$(\mathbf{A} + (-1)\mathbf{A}) = (1\mathbf{A} + (-1)\mathbf{A}) = (1+(-1))\mathbf{A} = 0\mathbf{A} = \mathbf{0}\,. \tag{D.12}$$

The above definitions and properties allow us to extend the familiar properties of ordinary 3-dimensional vectors and real numbers to other sets of objects. Although *most* properties of 3-dimensional vectors carry over to these higher-dimensional abstract vector spaces, some do not, such as the cross (vector) product of two vectors, e.g., $\mathbf{A} \times \mathbf{B}$, which is defined in 3-dimensions by (A.2). (But see the *wedge product* of differential forms described in Appendix B.)

## D.2 Basis Vectors

A key concept when working with vectors is that of a *basis*. But in order to define what we mean by a basis, we must first introduce some terminology:

- A **linear combination** of vectors $\mathbf{A}, \mathbf{B}, \cdots$ is any vector of the form

$$a\mathbf{A} + b\mathbf{B} + \cdots \tag{D.13}$$

  where $a, b, \cdots$ are scalars.
- A vector $\mathbf{C}$ is **linearly independent** of the set of vectors $\{\mathbf{A}, \mathbf{B}, \cdots\}$ if and only if $\mathbf{C}$ cannot be written as a linear combination of the vectors in the set—i.e.,

$$\neg\exists a, b, \cdots \quad \text{such that} \quad \mathbf{C} = a\mathbf{A} + b\mathbf{B} + \cdots \tag{D.14}$$

- A *set* of vectors $\{\mathbf{A}, \mathbf{B}, \cdots\}$ is a **linearly independent set** if and only if each vector in the set is linearly independent of all the other vectors in the set.
- A set of vectors $\{\mathbf{A}, \mathbf{B}, \cdots\}$ **spans** the vector space if and only if any vector $\mathbf{C}$ in the vector space can be written as a linear combinaton of the vectors in the set—i.e.,

$$\exists a, b, \cdots \quad \text{such that} \quad \mathbf{C} = a\mathbf{A} + b\mathbf{B} + \cdots \tag{D.15}$$

In terms of the above definitions, a **basis** for a vector space is defined to be any set of vectors which is (i) linearly independent and (ii) spans the vector space. The number of basis vectors is defined as the **dimension** of the vector space. Thus, an $n$-dimensional vector space has a basis consisting of $n$ vectors,

$$\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}\,. \tag{D.16}$$

This is, of course, consistent with what we know about the space of ordinary 3-dimensional vectors. The set of unit vectors $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$ is a basis for that space. (More about what *unit* means for a more general vector space in just a bit.)

### D.2.1 Components of a Vector

Given a set of basis vectors (D.16), we can write

$$\mathbf{A} = A_1\mathbf{e}_1 + A_2\mathbf{e}_2 + \cdots + A_n\mathbf{e}_n \equiv \sum_i A_i\mathbf{e}_i \, . \tag{D.17}$$

The scalars $A_1, A_2, \ldots, A_n$ are called the **components** of $\mathbf{A}$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$. The *decomposition* of $\mathbf{A}$ into its components is *unique* for a *given* basis as shown in Exercise D.1 below. But for a *different* set of basis vectors, e.g., $\{\mathbf{e}'_1, \mathbf{e}'_2, \ldots, \mathbf{e}'_n\}$, the components of $\mathbf{A}$ will be different.

**Exercise D.1** Prove that the decomposition (D.17) of $\mathbf{A}$ into its components $A_1, A_2, \ldots, A_n$ is unique. *Hint*: Use proof by contradiction—i.e., assume that there exist *other* components $A'_1, A'_2, \ldots, A'_n$, for which

$$\mathbf{A} = \sum_i A'_i\mathbf{e}_i \, . \tag{D.18}$$

Then show that this leads to a contradiction regarding the linear independence of the basis vectors unless $A'_i = A_i$ for all $i$.

Vector addition and scalar multiplication are what you might expect in terms of the components of the vectors. That is, if we denote the correspondence between vectors and components by[1]

$$\mathbf{A} \quad \leftrightarrow \quad \mathsf{A} \equiv [A_1, A_2, \ldots, A_n]^T \equiv \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} , \tag{D.19}$$

then it is easy to show that

[1]Our notation is such that $\mathbf{A}$ denotes the abstract vector, $A_i$ its $i$th component with respect to a basis, and $\mathsf{A}$ the collection of components $A_1, A_2, \ldots, A_n$ represented as an $n \times 1$ column matrix. The superscript $T$ denotes transpose, which converts a row matrix into a colum matrix, and vice versa.

$$\mathbf{A}+\mathbf{B} \quad \leftrightarrow \quad \mathsf{A}+\mathsf{B}=[A_1+B_1, A_2+B_2, \ldots, A_n+B_n]^T\,, \tag{D.20}$$

and

$$a\mathbf{A} \quad \leftrightarrow \quad a\mathsf{A}=[aA_1, aA_2, \ldots, aA_n]^T\,. \tag{D.21}$$

Thus, vector addition corresponds to ordinary addition of the (scalar) components of the vectors, and scalar multiplication of a vector corresponds to multiplying each of the components of the vector by that scalar.

**Exercise D.2** Show that the zero vector $\mathbf{0}$ and inverse vector $-\mathbf{A}$ can be written in terms of components as

$$\mathbf{0} \quad \leftrightarrow \quad \mathsf{0}=[0, 0, \ldots, 0]^T\,, \tag{D.22}$$

and

$$-\mathbf{A} \quad \leftrightarrow \quad -\mathsf{A}=[-A_1, -A_2, \ldots, -A_n]^T\,. \tag{D.23}$$

## D.3 Inner Product

An abstract $n$-dimensional vector space as defined above generalizes several key properties of vectors in ordinary 3-dimensional space. But by itself, the definition of a vector space does not specify how to calculate the *length* of a vector, or the *angle* that one vector makes with another vector.[2] In order to extend these concepts to higher-dimensional vector spaces, we need to introduce an additional mathematical structure, called an **inner product** on the space of vectors. As we shall see below, this inner product generalizes the notion of the ordinary *dot product* (or scalar product) $\mathbf{A}\cdot\mathbf{B}$ of 3-dimensional vectors, (A.1), and hence will allow us to talk about *unit* vectors and *orthogonality* of two vectors.

**Definition**: Given two vectors $\mathbf{A}$ and $\mathbf{B}$, the **inner product** of $\mathbf{A}$ and $\mathbf{B}$ (denoted $\mathbf{A}\cdot\mathbf{B}$) is a scalar (i.e., a complex number) that satisfies the following three properties:

$$\mathbf{A}\cdot\mathbf{B}=(\mathbf{B}\cdot\mathbf{A})^* \tag{D.24a}$$

$$\mathbf{A}\cdot\mathbf{A}\geq 0\,, \quad \mathbf{A}\cdot\mathbf{A}=0 \text{ if and only if } \mathbf{A}=\mathbf{0} \tag{D.24b}$$

$$\mathbf{A}\cdot(b\mathbf{B}+c\mathbf{C})=b(\mathbf{A}\cdot\mathbf{B})+c(\mathbf{A}\cdot\mathbf{C}) \tag{D.24c}$$

[2] As anybody who has taken freshman physics knows, length and angle are *key* concepts for ordinary (3-dimensional) vectors, which are sometimes *defined* as “arrows” having magnitude and direction!

Mathematicians call a vector space with the additional structure of an inner product an **inner product space**.

**Exercise D.3** Show that if $\mathbf{C} = b\mathbf{B}$ then $\mathbf{C} \cdot \mathbf{A} = b^* (\mathbf{B} \cdot \mathbf{A})$.

The fact that $\mathbf{A} \cdot \mathbf{A} \geq 0$ allows us to interpret $\mathbf{A} \cdot \mathbf{A}$ as the (squared) length or *norm* (or magnitude) of the vector $\mathbf{A}$. We will denote the norm of $\mathbf{A}$ as either $|\mathbf{A}|$ or $A$, so that

$$|\mathbf{A}| \equiv A \equiv \sqrt{\mathbf{A} \cdot \mathbf{A}}\,. \tag{D.25}$$

This leads to following definitions:

- A vector $\mathbf{A}$ is said to have *unit norm* (or to be **normalized**) if and only if $|\mathbf{A}| = 1$.
- Two vectors $\mathbf{A}$ and $\mathbf{B}$ are said to be **orthogonal** if and only if the inner product of $\mathbf{A}$ and $\mathbf{B}$ vanishes—i.e., $\mathbf{A} \cdot \mathbf{B} = 0$.
- A set of vectors $\{\mathbf{A}_1, \mathbf{A}_2, \cdots\}$ is said to be **orthonormal** if and only if $\mathbf{A}_i \cdot \mathbf{A}_j = \delta_{ij}$ for all $i, j = 1, 2, \ldots, n$, where $\delta_{ij}$ is the Kronecker delta symbol (which equals one if $i = j$, and equals zero otherwise).
- An **orthonormal basis** is an orthonormal set of basis vectors, which we will typically denote with hats, $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \ldots, \hat{\mathbf{e}}_n$. These vectors are thus linearly independent, span the vector space, and satisfy

$$\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij}\,. \tag{D.26}$$

### D.3.1 Gram-Schmidt Orthonormalization Procedure

As we already know from doing calculations with ordinary vectors in 3-dimensions, it is often convenient to work with a set of orthonormal basis vectors, e.g., $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$. Hence, it is good to know that there exists a general procedure, called the **Gram-Schmidt orthonormalization procedure**, for taking an arbitrary set of basis vectors in $n$ dimensions and converting it into an orthonormal set of basis vectors. As must be the case, this new set of basis vectors is formed by taking appropriate linear combinations of the original (non-orthonormal) basis vectors.

We start with a set of basis vectors

$$\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}\,, \tag{D.27}$$

which we will assume is not orthonormal. (If the basis vectors were already orthonormal, then there would be nothing that you need to do!) Take $\mathbf{e}_1$ and simply divide by its norm. The result is a unit vector that points in the same direction as $\mathbf{e}_1$,

$$\hat{\mathbf{f}}_1 \equiv \frac{\mathbf{e}_1}{|\mathbf{e}_1|} \,. \tag{D.28}$$

Now take $\mathbf{e}_2$, and subtract off its component in the direction of $\hat{\mathbf{f}}_1$:

$$\mathbf{f}_2 \equiv \mathbf{e}_2 - (\hat{\mathbf{f}}_1 \cdot \mathbf{e}_2)\,\hat{\mathbf{f}}_1 \,. \tag{D.29}$$

This makes $\mathbf{f}_2$ orthogonal to $\hat{\mathbf{f}}_1$ as one can easily check,

$$\hat{\mathbf{f}}_1 \cdot \mathbf{f}_2 = \hat{\mathbf{f}}_1 \cdot \mathbf{e}_2 - (\hat{\mathbf{f}}_1 \cdot \mathbf{e}_2)(\hat{\mathbf{f}}_1 \cdot \hat{\mathbf{f}}_1) = 0 \,. \tag{D.30}$$

Then normalize $\mathbf{f}_2$,

$$\hat{\mathbf{f}}_2 \equiv \frac{\mathbf{f}_2}{|\mathbf{f}_2|} \,. \tag{D.31}$$

Thus, both $\hat{\mathbf{f}}_1$ and $\hat{\mathbf{f}}_2$ have unit norm and they are orthogonal to one another. For $\hat{\mathbf{f}}_3$ we proceed in a similar fashion:

$$\mathbf{f}_3 \equiv \mathbf{e}_3 - (\hat{\mathbf{f}}_1 \cdot \mathbf{e}_3)\,\hat{\mathbf{f}}_1 - (\hat{\mathbf{f}}_2 \cdot \mathbf{e}_3)\,\hat{\mathbf{f}}_2 \,, \tag{D.32}$$

and

$$\hat{\mathbf{f}}_3 \equiv \frac{\mathbf{f}_3}{|\mathbf{f}_3|} \,. \tag{D.33}$$

Continue as above for $\hat{\mathbf{f}}_4, \hat{\mathbf{f}}_5, \ldots, \hat{\mathbf{f}}_n$.

Note that this procedure does not produce a *unique* orthonormal basis. The resulting set of orthonormal basis vectors depends on the ordering of the basis vectors as illustrated in the following exercise.

**Exercise D.4** (a) Use the Gram-Schmidt orthonormalization procedure to construct an orthonormal basis starting from

$$\mathbf{e}_1 \equiv \hat{\mathbf{x}} \,, \qquad \mathbf{e}_2 \equiv \hat{\mathbf{x}} + \hat{\mathbf{y}} \,, \qquad \mathbf{e}_3 \equiv \hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}} \,, \tag{D.34}$$

where $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ are the standard orthonormal basis vectors in ordinary 3-dimensional space. (b) Repeat the procedure, but this time with the basis vectors enumerated in the reverse order,

$$\mathbf{e}_1 \equiv \hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}} \,, \qquad \mathbf{e}_2 \equiv \hat{\mathbf{x}} + \hat{\mathbf{y}} \,, \qquad \mathbf{e}_3 \equiv \hat{\mathbf{x}} \,. \tag{D.35}$$

Do you get the same result as in part (a)?

### D.3.2 Component Form of the Inner Product

To illustrate that the inner product defined above generalizes the *dot* product of ordinary 3-dimensional vectors, it is simplest to show that we can recover the form of the dot product given in (A.4), which we rewrite here as

$$\mathbf{A}\cdot\mathbf{B} = A_1B_1 + A_2B_2 + A_3B_3 = \sum_i A_iB_i\,, \tag{D.36}$$

where $A_i$, $B_i$ with $i = 1, 2, 3$ are the components of the ordinary 3-dimensional vectors $\mathbf{A}$, $\mathbf{B}$ with respect to some orthonormal basis $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$.

So let's work now in an arbitrary $n$-dimensional vector space equipped with an inner product, and let $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \dots, \hat{\mathbf{e}}_n\}$ denote an orthonormal basis for this space. Then according to (D.17), we can write

$$\mathbf{A} = \sum_i A_i\hat{\mathbf{e}}_i\,, \qquad \mathbf{B} = \sum_i B_i\hat{\mathbf{e}}_i\,, \tag{D.37}$$

for any two vectors $\mathbf{A}$, $\mathbf{B}$, where $A_i$, $B_i$ with $i = 1, 2, \dots, n$ are the components of these vectors with respect to the given orthonormal basis. Since the basis vectors are orthonormal, it is easy to show that

$$A_i = \hat{\mathbf{e}}_i\cdot\mathbf{A}\,. \tag{D.38}$$

The proof is simply

$$\hat{\mathbf{e}}_i\cdot\mathbf{A} = \hat{\mathbf{e}}_i\cdot\left(\sum_j A_j\hat{\mathbf{e}}_j\right) = \sum_j A_j(\hat{\mathbf{e}}_i\cdot\hat{\mathbf{e}}_j) = \sum_j A_j\delta_{ij} = A_i\,, \tag{D.39}$$

where we used the linearity property (D.24c) of the inner product and the orthonormality (D.26) of the basis vectors to obtain the second and third equalities.[3] Using these results, it is then fairly straightforward to show that

$$\mathbf{A}\cdot\mathbf{B} = A_1^*B_1 + A_2^*B_2 + \cdots + A_n^*B_n = \sum_i A_i^*B_i\,, \tag{D.40}$$

and, as a consequence,

[3]Note that $A_i \neq \mathbf{A}\cdot\hat{\mathbf{e}}_i$ in general, since the components of a vector can be complex. Using Exercise (D.3), it follows that $\mathbf{A}\cdot\hat{\mathbf{e}}_i = A_i^*$.

$$|\mathbf{A}|^2 = \sum_i |A_i|^2 \,. \tag{D.41}$$

Note that (D.40) does indeed generalize the dot product (D.36) to arbitrary dimensions. Setting $n = 3$ and taking our vector components to be real-valued, we see that (D.40) reduces to (D.36).

**Exercise D.5** Prove the component form (D.40) of the inner product.

### *D.3.3 Schwarz Inequality*

Recall that for ordinary vectors in 3-dimensions, the dot product $\mathbf{A} \cdot \mathbf{B}$ can also be written as

$$\mathbf{A} \cdot \mathbf{B} = AB\cos\theta \,, \tag{D.42}$$

where $A \equiv |\mathbf{A}| \equiv \sqrt{\mathbf{A} \cdot \mathbf{A}}$ and $B \equiv |\mathbf{B}| \equiv \sqrt{\mathbf{B} \cdot \mathbf{B}}$ are the magnitudes of the two vectors, and $\theta$ is the angle between them; see (A.1). If we rewrite the above equation as

$$\cos\theta = \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}||\mathbf{B}|} \,, \tag{D.43}$$

then it can be thought of as the *definition* of the angle between the two vectors in terms of their dot products and their magnitudes. This suggests a way of generalizing the concept of "angle between two vectors" to an arbitrary $n$-dimensional vector space. Namely, simply interpret the expressions on the right-hand side of (D.43) in terms of the inner product defined by (D.24a), (D.24b), (D.24c). Unfortunately this won't work since $\mathbf{A} \cdot \mathbf{B}$ is a *complex number* in general, so the angle $\theta$ would not be real. But it turns out that there is a simple solution, which amounts to taking the *absolute value* of the right-hand side,

$$\cos\theta = \frac{|\mathbf{A} \cdot \mathbf{B}|}{|\mathbf{A}||\mathbf{B}|} = \sqrt{\frac{|\mathbf{A} \cdot \mathbf{B}|^2}{(\mathbf{A} \cdot \mathbf{A})(\mathbf{B} \cdot \mathbf{B})}} \,. \tag{D.44}$$

So this is now a real quantity, but the fact that this equation actually gives us something that we can interpret as an angle is thanks to the **Schwarz inequality**

$$|\mathbf{A} \cdot \mathbf{B}|^2 \le (\mathbf{A} \cdot \mathbf{A})(\mathbf{B} \cdot \mathbf{B}) \,, \tag{D.45}$$

which guarantees that the right-hand side of (D.44) has a value $\le 1$.

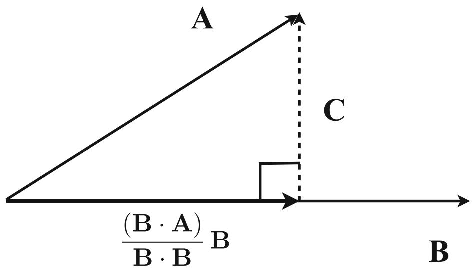



**Fig. D.1** Graphical illustration of the vector **C** used in the proof of the Schwarz inequality. Note that $((\mathbf{B}\cdot\mathbf{A})/\mathbf{B}\cdot\mathbf{B})\mathbf{B}$ is the projection of **A** onto **B**, so (by construction) **C** is orthogonal to **B**. Note that $\mathbf{C}=\mathbf{0}$ if and only if **A** and **B** are proportional to one another, which corresponds to the equal sign in the Schwarz inequality

**Exercise D.6** Prove the Schwarz inequality. (*Hint*: Consider the vector

$$\mathbf{C} \equiv \mathbf{A} - \frac{(\mathbf{B}\cdot\mathbf{A})}{\mathbf{B}\cdot\mathbf{B}}\mathbf{B}\,, \tag{D.46}$$

and then use $|\mathbf{C}|^2 \equiv \mathbf{C}\cdot\mathbf{C} \geq 0$. (See Fig. D.1.) Note that the Schwarz inequality becomes an equality when **A** and **B** are proportional (i.e., parallel or anti-parallel) to one another.)

## D.4 Linear Transformations

Given our abstract $n$-dimensional vector space, we would now like to define a certain class of operations (called *linear transformations*), which map vectors to other vectors in such a way that they preserve the linear property of vector addition and scalar multiplication of vectors. Rotations of ordinary 3-dimensional vectors, which play an important role in *all* branches of physics, are just one example of linear transformations.

**Definition**: A **linear transformation T** is a mapping that takes a vector **A** to another vector $\mathbf{A}' \equiv \mathbf{T}\mathbf{A}$ such that

$$\mathbf{T}(a\mathbf{A} + b\mathbf{B}) = a(\mathbf{T}\mathbf{A}) + b(\mathbf{T}\mathbf{B})\,. \tag{D.47}$$

This method of introducing additional structure on a space in such a way that it interacts "naturally" with other structures in the space (in this case scalar multiplication and vector addition) is common practice in mathematics.

Note that multiplying every vector in the space by the same scalar $c$ (i.e., $\mathbf{A} \mapsto c\mathbf{A}$) is an example of a linear transformation since

$$\begin{aligned} c(a\mathbf{A} + b\mathbf{B}) &= c(a\mathbf{A}) + c(b\mathbf{B}) \\ &= (ca)\mathbf{A} + (cb)\mathbf{B} \\ &= (ac)\mathbf{A} + (bc)\mathbf{B} \\ &= a(c\mathbf{A}) + b(c\mathbf{B}) , \end{aligned} \tag{D.48}$$

where we used the distributive property of scalar multiplication with respect vector addition (D.9); the associative property of scalar multiplication of vectors (D.10); the commutative property for multiplication of two scalars (complex numbers); and the associative property of scalar multiplication of vectors (again) to get the successive equalities above. But adding a constant vector $\mathbf{C}$ to every vector in the space is *not* an example of a linear transformation as you are asked to show in the following exercise.

**Exercise D.7** Prove that adding a constant vector $\mathbf{C}$ to every vector in the space, i.e., $\mathbf{A} \mapsto \mathbf{A} + \mathbf{C}$, is *not* an example of a linear transformation.

The set of linear transformations, by itself, has an interesting mathematical structure. You can define (for any vector $\mathbf{A}$):

(i) addition of two linear transformations:

$$(\mathbf{S} + \mathbf{T})\mathbf{A} \equiv \mathbf{S}\mathbf{A} + \mathbf{T}\mathbf{A} \tag{D.49}$$

(ii) multiplication of a linear transformation by a scalar:

$$(a\mathbf{T})\mathbf{A} \equiv a(\mathbf{T}\mathbf{A}) \tag{D.50}$$

(iii) multiplication (or **composition**) of two linear transformations:

$$(\mathbf{S}\mathbf{T})\mathbf{A} \equiv \mathbf{S}(\mathbf{T}\mathbf{A}) \tag{D.51}$$

With the first two operations, (D.49) and (D.50), the space of linear transformations has the structure of an $n^2$-dimensional vector space over the complex numbers (Exercise D.8). Note that multiplication of linear transformations is not commutative, however, since

$$\mathbf{S}\mathbf{T} \neq \mathbf{T}\mathbf{S} , \tag{D.52}$$

in general. (Think of rotations in 3-dimensions; see, e.g., Fig. 6.5.) We will return to the multiplicative structure of linear transformations later in this section, after we develop the connection between linear transformations and matrices.

**Exercise D.8** Show that with the above definitions of addition of linear transformations and multiplication of a linear transformation by a scalar, the set of linear transformations has the structure of a vector space over the complex numbers. Note that you will need to verify that these operations satisfy the properties given in (D.1)–(D.5) and (D.6)–(D.10).

### *D.4.1 Component Form of a Linear Transformation*

The real beauty of the linearity property (D.47) is that once you know what a linear transformation $\mathbf{T}$ does to a set of basis vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, you can easily determine what it does to *any* vector $\mathbf{A}$. To see that this is the case, let's begin by writing

$$\mathbf{T}\mathbf{e}_1 = T_{11}\mathbf{e}_1 + T_{21}\mathbf{e}_2 + \cdots + T_{n1}\mathbf{e}_n = \sum_i T_{i1}\mathbf{e}_i\,, \tag{D.53}$$

which follows from the fact that any vector (in this case $\mathbf{T}\mathbf{e}_1$) can be written as a linear combination of the basis vectors). Similarly,

$$\begin{gathered}\mathbf{T}\mathbf{e}_2 = T_{12}\mathbf{e}_1 + T_{22}\mathbf{e}_2 + \cdots + T_{n2}\mathbf{e}_n = \sum_i T_{i2}\mathbf{e}_i\,, \\ \vdots \\ \mathbf{T}\mathbf{e}_n = T_{1n}\mathbf{e}_1 + T_{2n}\mathbf{e}_2 + \cdots + T_{nn}\mathbf{e}_n = \sum_i T_{in}\mathbf{e}_i\,.\end{gathered} \tag{D.54}$$

Thus, we see that the action of $\mathbf{T}$ on the $n$ basis vectors is completely captured by the $n \times n$ numbers $T_{ij}$, where

$$\mathbf{T}\mathbf{e}_j = \sum_i T_{ij}\mathbf{e}_i\,, \qquad j = 1, 2, \ldots, n\,. \tag{D.55}$$

We will write these components as an $n \times n$ matrix:

$$\mathbf{T} \quad \leftrightarrow \quad \mathsf{T} = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & T_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{n1} & T_{n2} & \cdots & T_{nn} \end{bmatrix}, \tag{D.56}$$

where the doubleheaded arrow $\leftrightarrow$ reminds us that $\mathsf{T}$ are the components of $\mathbf{T}$ with respect to a particular basis. (With respect to a different basis, the matrix components will change in a manner that we will investigate shortly.) If the basis is orthonormal, then we can write

$$T_{ij} = \mathbf{e}_i \cdot (\mathbf{T}\mathbf{e}_j) \qquad \text{(orthonomal basis)}\,, \tag{D.57}$$

which follows immediately from (D.55).

Returning now to our claim that the action of **T** on any vector **A** is completely determined once we know the action of **T** on the basis vectors, we can write

$$\mathbf{TA} = \mathbf{T}(\sum_j A_j \mathbf{e}_j) = \sum_j A_j (\mathbf{T}\mathbf{e}_j) = \sum_j A_j \sum_i T_{ij}\mathbf{e}_i = \sum_i \left( \sum_j T_{ij} A_j \right) \mathbf{e}_i\,, \tag{D.58}$$

where we used the linearity property (D.47) to get the second equality and the action of **T** on the basis vectors (D.55) to get the third. Thus, knowing the action of **T** on the basis vectors (i.e., the components $T_{ij}$), we can determine the action of **T** on any vector **A**. If we denote **TA** as $\mathbf{A}'$ and the components of $\mathbf{A}'$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ as $A'_i$, then (D.58) becomes

$$\mathbf{A}' = \mathbf{TA} \quad \Leftrightarrow \quad A'_i = \sum_j T_{ij} A_j\,, \tag{D.59}$$

which is equivalent to the matrix equation $\mathsf{A}' = \mathsf{TA}$:

$$\begin{bmatrix} A'_1 \\ A'_2 \\ \vdots \\ A'_n \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & T_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{n1} & T_{n2} & \cdots & T_{nn} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix}. \tag{D.60}$$

Thus, the mathematical structure of linear transformations on an $n$-dimensional vector space is equivalent to that of $n \times n$ matrices.

**Exercise D.9** Show that addition of linear transformations, multiplication of a linear transformation by a scalar, and multiplication (composition) of two linear transformations, defined by (D.49), (D.50), and (D.51), become simply:

$$S_{ij} + T_{ij}\,, \qquad aT_{ij}\,, \qquad \sum_k S_{ik} T_{kj}\,, \tag{D.61}$$

in terms of the corresponding components. Note that the last expression is just the component form of ordinary matrix multiplication of two matrices.

### D.4.2 Change of Basis

Before discussing properties of matrices in general, let's first determine how the components of a vector **A** and the components (i.e., matrix elements) of a linear transformation **T** transform under a change of basis. The classic example of a change of basis is given by a *rotation* of the coordinate basis vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ to a new set of basis vectors $\hat{\mathbf{x}}'$, $\hat{\mathbf{y}}'$, $\hat{\mathbf{z}}'$.

So let's denote the two sets of basis vectors by

$$\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}\,, \qquad \{\mathbf{e}_{1'}, \mathbf{e}_{2'}, \ldots, \mathbf{e}_{n'}\}\,, \tag{D.62}$$

where we use primed indices to distinguish between the two bases. We will assume, for now, that these are arbitrary bases—i.e., we do not require that they be orthonormal. Since any vector can be expanded in terms of either set of basis vectors, we can write

$$\begin{aligned}
\mathbf{e}_1 &= S_{1'1}\mathbf{e}_{1'} + S_{2'1}\mathbf{e}_{2'} + \cdots + S_{n'1}\mathbf{e}_{n'} = \sum_{j'} S_{j'1}\mathbf{e}_{j'}\,,\\
\mathbf{e}_2 &= S_{1'2}\mathbf{e}_{1'} + S_{2'2}\mathbf{e}_{2'} + \cdots + S_{n'2}\mathbf{e}_{n'} = \sum_{j'} S_{j'2}\mathbf{e}_{j'}\,,\\
&\;\;\vdots\\
\mathbf{e}_n &= S_{1'n}\mathbf{e}_{1'} + S_{2'n}\mathbf{e}_{2'} + \cdots + S_{n'2}\mathbf{e}_{n'} = \sum_{j'} S_{j'n}\mathbf{e}_{j'}\,,
\end{aligned} \tag{D.63}$$

for some set of components $S_{j'i}$. In compact form

$$\mathbf{e}_i = \sum_{j'} S_{j'i}\mathbf{e}_{j'}\,, \qquad i = 1, 2, \ldots, n\,. \tag{D.64}$$

The components $S_{j'i}$ define an $n \times n$ matrix, which is necessarily invertible (otherwise $\{\mathbf{e}_1, \mathbf{e}_2, \cdots\}$ wouldn't form a basis).

Now consider a single vector **A**, which has components $A_i$ and $A_{i'}$, respectively, with respect to the unprimed and primed bases:

$$\mathbf{A} = \sum_i A_i\mathbf{e}_i = \sum_{i'} A_{i'}\mathbf{e}_{i'}\,. \tag{D.65}$$

Then using (D.64), it follows that

$$\sum_i A_i\mathbf{e}_i = \sum_i A_i \sum_{j'} S_{j'i}\mathbf{e}_{j'} = \sum_{j'}\left(\sum_i S_{j'i}A_i\right)\mathbf{e}_{j'}\,. \tag{D.66}$$

Comparing with (D.65) we see that

$$A_{j'} = \sum_i S_{j'i} A_i \,, \qquad j' = 1', 2', \ldots, n' \,. \tag{D.67}$$

The inverse of the above equation can be written as

$$A_i = \sum_{j'} (S^{-1})_{ij'} A_{j'} \,, \tag{D.68}$$

where $(S^{-1})_{ij'}$ are the components of the inverse matrix to $S_{j'i}$:

$$\sum_{j'} (S^{-1})_{ij'} S_{j'k} = \delta_{ik} \,, \qquad \sum_i S_{j'i} (S^{-1})_{ik'} = \delta_{j'k'} \,. \tag{D.69}$$

Now take a linear transformation $\mathbf{T}$, which maps $\mathbf{A}$ to $\mathbf{B} \equiv \mathbf{TA}$. Using (D.67), (D.59) and (D.68), it follows that

$$\begin{aligned} B_{i'} &= \sum_k S_{i'k} B_k = \sum_k S_{i'k} \sum_l T_{kl} A_l = \sum_k S_{i'k} \sum_l T_{kl} \sum_{j'} (S^{-1})_{lj'} A_{j'} \\ &= \sum_{j'} \left( \sum_{k,l} S_{i'k} T_{kl} (S^{-1})_{lj'} \right) A_{j'} = \sum_{j'} T_{i'j'} A_{j'} \end{aligned} \tag{D.70}$$

where

$$T_{i'j'} \equiv \sum_{k,l} S_{i'k} T_{kl} (S^{-1})_{lj'} \,. \tag{D.71}$$

Noting that the products and summations on the right-hand side are exactly those for a product of matrices, we have

$$\mathsf{T}' = \mathsf{STS}^{-1} \,, \tag{D.72}$$

where $\mathsf{T}'$ is the $n \times n$ matrix of components $T_{i'j'}$. Such a transformation of matrices is called a **similarity transformation**.

### *D.4.3 Matrix Definitions and Operations*

As illustrated by the calculations in the last two subsections, matrices play a key role in linear algebra. In this section, we summarize some important definitions and operations involving matrices, which we will refer to repeatedly in the main text.

Most of the discussion will be restricted to $n \times n$ (i.e., square) matrices, although the transpose and conjugate operations (complex conjugate and Hermitian conjugate) can be defined for arbitrary $n \times m$ (i.e., rectangular) matrices.

#### D.4.3.1 Transpose and Conjugate Matrices

The **transpose**, **conjugate**, and **Hermitian conjugate** of a matrix $\mathsf{T}$ are defined by:

$$(\mathsf{T}^T)_{ij} = T_{ji}\,, \quad (\mathsf{T}^*)_{ij} = T^*_{ij}\,, \quad (\mathsf{T}^\dagger)_{ij} = T^*_{ji}\,. \tag{D.73}$$

A matrix is said to be **symmetric** if and only if

$$\mathsf{T} = \mathsf{T}^T \quad \leftrightarrow \quad T_{ij} = T_{ji}\,, \tag{D.74}$$

and **Hermitian** if and only if

$$\mathsf{T} = \mathsf{T}^\dagger \quad \leftrightarrow \quad T_{ij} = T^*_{ji}\,. \tag{D.75}$$

**Anti-symmetric** and **Anti-hermitian** matrices are defined with minus signs in the last two equations.

**Exercise D.10** Show that the transpose of a product of matrices equals the product of transposes in the *opposite* order:

$$(\mathsf{ST})^T = \mathsf{T}^T\,\mathsf{S}^T\,, \tag{D.76}$$

and similarly for the Hermitian conjugate:

$$(\mathsf{ST})^\dagger = \mathsf{T}^\dagger\,\mathsf{S}^\dagger\,. \tag{D.77}$$

Note that these relations hold in general for the product of an $m \times n$ matrix and an $n \times p$ matrix.

**Exercise D.11** Show that the inner product (D.40) of two vectors **A** and **B** can be written in terms of row and column matrices as

$$\mathbf{A} \cdot \mathbf{B} = \mathsf{A}^\dagger \mathsf{B}\,. \tag{D.78}$$

#### D.4.3.2 Determinants

The **determinant** of a $2 \times 2$ matrix is defined by

$$\mathsf{T} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad \det \mathsf{T} \equiv \begin{vmatrix} a & b \\ c & d \end{vmatrix} \equiv ad - bc\,. \tag{D.79}$$

For a higher-order $n \times n$ matrix $\mathsf{T}$, we define its determinant in terms of the determinants of $(n-1) \times (n-1)$ sub-matrices of $\mathsf{T}$. (This procedure is called **Laplace development** of the determinant.) Explicitly, if we expand off of the $i$th row of $\mathsf{T}$ then

$$\det \mathsf{T} = \sum_j T_{ij} (-1)^{i+j} M_{ij} = \sum_j T_{ij} C_{ij}\,, \tag{D.80}$$

where $M_{ij}$ is the **minor** of $T_{ij}$ and $C_{ij} \equiv (-1)^{i+j} M_{ij}$ is the corresponding **cofactor**. The minor $M_{ij}$ is calculated by taking the determinant of the $(n-1) \times (n-1)$ matrix obtained from $\mathsf{T}$ by removing its $i$th row and $j$th column. Note that you get the same answer for the determinant of $\mathsf{T}$ regardless of which row you expand off of, or if you expand off of a *column* instead of a row. In addition, as you will show in part (c) of Exercise D.12 below, adding a multiple of one row (or column) of a square matrix to another row (or column) does not change the value of its determinant. Thus, a judicious choice of such elementary row (or column) operations can simplify the calculation of the determinant.

**Exercise D.12** The determinant of an $n \times n$ matrix can also be defined by

$$\det \mathsf{T} = \sum_{i_1, i_2, \cdots i_n} \varepsilon_{i_1 i_2 \cdots i_n} T_{1 i_1} T_{2 i_2} \cdots T_{n i_n}\,, \tag{D.81}$$

where $\varepsilon_{i_1 i_2 \cdots i_n}$ is the $n$-dimensional Levi-Civita symbol:

$$\varepsilon_{i_1 i_2 \cdots i_n} \equiv \begin{cases} 1 & \text{if } i_1 i_2 \cdots i_n \text{ is an even permutation of } 12 \cdots n \\ -1 & \text{if } i_1 i_2 \cdots i_n \text{ is an odd permutation of } 12 \cdots n \\ 0 & \text{otherwise} \end{cases} \tag{D.82}$$

See (A.7) for the 3-dimensional version of the Levi-Civita symbol, which enters the expression for the vector product of two 3-dimensional vectors.

(a) Work out the explicit expression for the determinant of a $3 \times 3$ matrix using the definition given in (D.81).

(b) Do the same using the earlier definition (D.80), and confirm that the two expressions you obtain agree with one another.
(c) Using the above definition (D.81), show that the determinant of an $n \times n$ matrix $\mathsf{T}$ is unchanged if you add a multiple of one row (or column) of $\mathsf{T}$ to another row (or column) before taking its determinant.

#### D.4.3.3 Unit Matrix and Inverses

The **unit** (or **identity**) **matrix** $\mathsf{1}$ has components given by the Kronecker delta $\delta_{ij}$:

$$\mathsf{1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} . \tag{D.83}$$

A matrix $\mathsf{T}$ is said to be *invertible* if and only if there exist another matrix $\mathsf{T}^{-1}$, called the **inverse matrix** of $\mathsf{T}$, such that

$$\mathsf{T}\mathsf{T}^{-1} = \mathsf{T}^{-1}\mathsf{T} = \mathsf{1} , \tag{D.84}$$

or, equivalently,

$$\sum_k T_{ik}(T^{-1})_{kj} = \sum_k (T^{-1})_{ik} T_{kj} = \delta_{ij} . \tag{D.85}$$

It turns out that a matrix is invertible if and only if its determinant is non-zero. An explicit expression for the inverse matrix is

$$\mathsf{T}^{-1} = \frac{1}{\det \mathsf{T}} \mathsf{C}^T , \tag{D.86}$$

where $\mathsf{C}$ is the matrix of cofactors. The inverse of a product of two invertible matrices $\mathsf{S}$ and $\mathsf{T}$ is the product of the inverse matrices in $\mathsf{S}^{-1}$ and $\mathsf{T}^{-1}$ in the opposite order,

$$(\mathsf{S}\mathsf{T})^{-1} = \mathsf{T}^{-1}\mathsf{S}^{-1} . \tag{D.87}$$

**Exercise D.13** Calculate the inverse matrices for the general $2 \times 2$ and $3 \times 3$ matrices

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} , \qquad \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} , \tag{D.88}$$

assuming that the determinants are non-zero for both.

#### D.4.3.4 Orthogonal and Unitary Matrices

A matrix is said to be **orthogonal** if and only if

$$\mathsf{T}^T = \mathsf{T}^{-1} \quad \leftrightarrow \quad \sum_k T_{ik} T_{jk} = \delta_{ij} \, , \quad \sum_k T_{ki} T_{kj} = \delta_{ij} \, . \tag{D.89}$$

A matrix is said to be **unitary** if and only if

$$\mathsf{T}^\dagger = \mathsf{T}^{-1} \quad \leftrightarrow \quad \sum_k T_{ik} T^*_{jk} = \delta_{ij} \, , \quad \sum_k T^*_{ki} T_{kj} = \delta_{ij} \, . \tag{D.90}$$

**Exercise D.14** Show that an ordinary rotation in 3-dimensions, e.g.,

$$\mathsf{R}_z(\phi) = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} , \tag{D.91}$$

is an example of an orthogonal matrix.

#### D.4.3.5 Useful Properties of Determinants

The determinant satisfies several useful properties:

$$\det \mathsf{1} = 1 \, , \qquad \det(\mathsf{T}^T) = \det \mathsf{T} \, , \qquad \det(\mathsf{S}\mathsf{T}) = \det \mathsf{S} \, \det \mathsf{T} \, . \tag{D.92}$$

Applying the above results to (D.84), it follows that

$$\det(\mathsf{T}^{-1}) = \frac{1}{\det \mathsf{T}} = (\det \mathsf{T})^{-1} \, . \tag{D.93}$$

In Appendix D.4.2, we derived how the components of a linear transformation $\mathsf{T}$ change under a change of basis, (D.72). Using the above properties, it follows that

$$\det(\mathsf{STS}^{-1}) = \det \mathsf{S} \, \det \mathsf{T} \, \det(\mathsf{S}^{-1}) = \det \mathsf{S} \, \det \mathsf{T} \, (\det \mathsf{S})^{-1} = \det \mathsf{T} \,. \quad \text{(D.94)}$$

Thus, the determinant of a matrix is *invariant* under a similarity transformation. In other words, the value of the determinant doesn't depend on what basis we use to convert a linear transformation $\mathbf{T}$ to a matrix of components $T_{ij}$.

#### D.4.3.6 Trace

There is another operation on matrices that is invariant under a similarity transformation. It is the **trace**, which is defined as the sum of the diagonal elements of the matrix,

$$\mathrm{Tr}(\mathsf{T}) \equiv \sum_i T_{ii} \,. \quad \text{(D.95)}$$

Since one can show that (Exercise D.15)

$$\mathrm{Tr}(\mathsf{ST}) = \mathrm{Tr}(\mathsf{TS}) \,, \quad \text{(D.96)}$$

it follows trivially that

$$\mathrm{Tr}(\mathsf{STS}^{-1}) = \mathrm{Tr}(\mathsf{S}^{-1}\mathsf{ST}) = \mathrm{Tr}(\mathsf{T}) \,. \quad \text{(D.97)}$$

Thus, the trace of a matrix, like the determinant, is also invariant under a similarity transformation, (D.72).

**Exercise D.15** Prove property (D.96) for the trace operation.

## D.5 Eigenvectors and Eigenvalues

The last topic that we will discuss in our review of linear algebra involves **eigenvectors** and **eigenvalues** of a linear transformation $\mathbf{T}$. Eigenvectors of $\mathbf{T}$ are special vectors, which are *effectively unchanged* by the action of $\mathbf{T}$. By "effectively unchanged" we mean that the eigenvector need only be mapped to itself *up to an overall proportionality factor*, which is called the **eigenvalue** of the eigenvector. If we denote an eigenvector of $\mathbf{T}$ by $\mathbf{v}$ and its eigenvalue by $\lambda$, then

$$\mathbf{T}\mathbf{v} = \lambda \mathbf{v} \,. \tag{D.98}$$

Note that the magnitude of the eigenvector $\mathbf{v}$ is not fixed by the above equation as $\mathbf{v}' \equiv a\mathbf{v}$ is also an eigenvector of $\mathbf{T}$ with the same eigenvalue $\lambda$.

***Example D.1*** As a simple example of an eigenvector, consider the space of ordinary 3-dimensional vectors and let's take as our linear transformation a counter-clockwise rotation about some axis $\hat{\mathbf{n}}$ through the angle $\Psi$, which we will denote as $\mathbf{R}_{\hat{\mathbf{n}}}(\Psi)$. Then $\hat{\mathbf{n}}$ is trivially an eigenvector of $\mathbf{R}_{\hat{\mathbf{n}}}(\Psi)$ with eigenvalue 1, since all points on the axis of rotation are left invariant by the transformation. □

**Exercise D.16** Suppose we restrict attention to ordinary *real-valued* vectors and rotations in 2-dimensions. Do any non-zero real-valued eigenvectors exist for such transformations? If so, what rotation angles do they correspond to?

## *D.5.1 Characteristic Equation*

If we introduce a basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, then we can recast (D.98) as a matrix equation

$$\mathsf{T}\mathsf{v} = \lambda \mathsf{v} \,, \tag{D.99}$$

where $\mathsf{v}$ and $\mathsf{T}$ are the matrix representations of $\mathbf{v}$ and $\mathbf{T}$ with respect to the basis. This last equation is equivalent to

$$(\mathsf{T} - \lambda \mathsf{1})\mathsf{v} = \mathsf{0} \,, \tag{D.100}$$

where the right-hand side is the zero-vector $\mathsf{0} \equiv [0, 0, \ldots, 0]^T$. Since this is a homogeneous equation, $\mathsf{v} = \mathsf{0}$ is a (trivial) solution, and it is the only solution if $(\mathsf{T} - \lambda \mathsf{1})$ is invertible. Hence, a non-zero solution to this equation requires that the matrix $(\mathsf{T} - \lambda \mathsf{1})$ *not* be invertible or, equivalently, that

$$\det(\mathsf{T} - \lambda \mathsf{1}) = 0 \,. \tag{D.101}$$

Expanding the determinant yields an $n$th-order polynomial equation for $\lambda$, which is called the **characteristic equation**

$$\det(\mathsf{T} - \lambda \mathbb{1}) = c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_n\lambda^n = 0\,, \tag{D.102}$$

where the coefficients $c_i$ are algebraic expressions involving the matrix elements $T_{ij}$. By the **fundamental theorem of algebra**, this equations admits $n$ *complex* roots $\lambda_i$, which might be zero or repeated multiple times,

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda) = 0\,. \tag{D.103}$$

The $n$ roots are the eigenvalues of (D.99).

Given the eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_n$, we can now substitute them back into (D.100), one at a time, and solve for the elements of the corresponding eigenvectors $\mathsf{v}_1, \mathsf{v}_2, \ldots, \mathsf{v}_n$. Since $\det(\mathsf{T} - \lambda\mathbb{1}) = 0$, not all of the components of the individual eigenvectors $\mathsf{v}_i$ will be uniquely determined. As mentioned earlier, there is always the freedom of an overall normalization factor, which we will usually chose to make each eigenvector have unit norm.

***Example D.2*** Find the eigenvectors and eigenvalues of the matrix

$$\mathsf{T} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{D.104}$$

We start by writing down the characteristic equation

$$\det(\mathsf{T} - \lambda\mathbb{1}) = \begin{vmatrix} -\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - 1 = 0\,. \tag{D.105}$$

This has two real solutions

$$\lambda_+ = 1\,, \qquad \lambda_- = -1\,. \tag{D.106}$$

Substituting the solution $\lambda_+ = 1$ back into the eigenvector-eigenvalue equation (D.100), we have

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{D.107}$$

This yields two equations

$$\begin{aligned} -v_1 + v_2 &= 0\,, \\ v_1 - v_2 &= 0\,, \end{aligned} \tag{D.108}$$

which (as expected) are linearly dependent on one another. The solution to these equations is

$$v_1 = v_2\,. \tag{D.109}$$

Using our freedom in choice of an overall multiplicative factor, we can choose $v_1 = v_2 = 1/\sqrt{2}$ for which

$$\mathsf{v}_+ = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{D.110}$$

Repeating this procedure for $\lambda_- = -1$, we find

$$\mathsf{v}_- = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{D.111}$$

Note that these two eigenvectors have unit norm and are orthogonal to one another,

$$\mathsf{v}_+^\dagger \mathsf{v}_- = \frac{1}{2}\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{2}(1-1) = 0\,. \tag{D.112}$$

Thus, the corresponding vectors $\mathbf{v}_+, \mathbf{v}_-$ form an orthonormal basis for the (real-valued) 2-dimensional vector space. But as we shall explain in the next subsection, it is not always the case that the eigenvectors of an arbitrary matrix form a basis for the vector space. □

**Exercise D.17** Find the eigenvectors and eigenvalues of the 2-dimensional rotation matrix

$$\mathsf{R}(\phi) = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}. \tag{D.113}$$

Note that you will need to allow complex-valued eigenvectors in general.

## *D.5.2 Diagonalizing a Matrix*

If the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of a linear transformation $\mathbf{T}$ span the $n$-dimensional vector space, then they can be used as a *new* set of basis vectors

$$\mathbf{e}_{1'} \equiv \mathbf{v}_1\,, \qquad \mathbf{e}_{2'} \equiv \mathbf{v}_2\,, \qquad \cdots\,, \qquad \mathbf{e}_{n'} \equiv \mathbf{v}_n\,, \tag{D.114}$$

in place of the original basis vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. Since

$$\mathbf{T}\mathbf{e}_{i'} = \lambda_i \mathbf{e}_{i'}\,, \qquad i' = 1', 2', \ldots, n'\,, \tag{D.115}$$

it follows that the components $T_{i'j'}$ of $\mathbf{T}$ in this new basis are given by $T_{i'j'} = \lambda_i \delta_{i'j'}$, which in matrix form is

$$\mathsf{T}' = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}. \tag{D.116}$$

Thus, the matrix $\mathsf{T}'$ is *diagonal*.

It's not too hard to show that the matrix $\mathsf{S}$, which transforms the components $T_{ij}$ to the components $T_{i'j'}$ via

$$\mathsf{T}' = \mathsf{STS}^{-1}, \tag{D.117}$$

has

$$\mathsf{S}^{-1} = \left[\, \mathsf{e}_{1'} \; \mathsf{e}_{2'} \cdots \mathsf{e}_{n'} \,\right] = \left[\, \mathsf{v}_1 \; \mathsf{v}_2 \cdots \mathsf{v}_n \,\right], \tag{D.118}$$

or, equivalently,

$$(S^{-1})_{ij'} = (\mathsf{e}_{j'})_i\,. \tag{D.119}$$

In other words, the columns of $\mathsf{S}^{-1}$ are just the eigenvectors of $\mathbf{T}$ in the original basis.

*Proof*

$$\begin{aligned}(\mathsf{STS}^{-1})_{i'j'} &= \sum_k \sum_l S_{i'k} T_{kl} (S^{-1})_{lj'} = \sum_k \sum_l S_{i'k} T_{kl} (\mathsf{e}_{j'})_l \\ &= \sum_k S_{i'k} \lambda_j (\mathsf{e}_{j'})_k = \lambda_j \sum_k S_{i'k} (S^{-1})_{kj'} \\ &= \lambda_j \delta_{i'j'} = T_{i'j'}\,,\end{aligned} \tag{D.120}$$

where we used (D.119) twice and also $\mathsf{Te}_{j'} = \lambda_j \mathsf{e}_{j'}$. □

If, in addition to spanning the vector space, the eigenvectors of $\mathbf{T}$ are *orthonormal*, then the matrix $\mathsf{S}$ also has a simple form,

$$\mathsf{S} = \begin{bmatrix} \mathsf{v}_1^\dagger \\ \mathsf{v}_2^\dagger \\ \vdots \\ \mathsf{v}_n^\dagger \end{bmatrix} = \left[\, \mathsf{v}_1 \; \mathsf{v}_2 \cdots \mathsf{v}_n \,\right]^\dagger. \tag{D.121}$$

Comparing with (D.118) we see that

$$\mathsf{S}^\dagger = \mathsf{S}^{-1}, \tag{D.122}$$

so the similarity matrix $\mathsf{S}$ is *unitary*, cf. (D.90).

**Exercise D.18** Using the results of Example D.2, show explicitly that

$$\mathsf{S}^{-1} = [\mathsf{v}_+ \ \mathsf{v}_-] = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{D.123}$$

diagonalizes

$$\mathsf{T} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{D.124}$$

In so doing, calculate $\mathsf{S}$ and show that it is unitary (actually *orthogonal* in this case, since the eigenvectors are all real-valued).

Although every linear transformation (or $n \times n$ matrix) admits (complex-valued) eigenvectors, not all $n \times n$ matrices can be diagonalized. The eigenvectors would need to span the vector space, but this is not the case in general. As a simple example, consider the $2 \times 2$ matrix

$$\mathsf{T} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{D.125}$$

Its two eigenvalues $\lambda_1, \lambda_2$ are both equal to 0, and the corresponding eigenvectors $\mathsf{v}_1$, $\mathsf{v}_2$ are both proportional to $[1, 0]^T$. Hence the eigenvectors span only a 1-dimensional subspace of the 2-dimensional vector space, and the similarity transformation $\mathsf{S}$ needed to map $\mathsf{T}$ to the diagonal matix

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{D.126}$$

does not exist. Thus, the matrix given by (D.125) cannot be diagonalized.

Fortunately, there is a certain class of matrices that are *guaranteed* to be diagonalizable. These are *Hermitian* matrices, for which $T_{ij} = T_{ji}^*$. (For a real-valued vector space, these matrices are *symmetric*, i.e., $T_{ij} = T_{ji}$.) Not only do the eigenvectors of a Hermitian matrix span the space, but the eigenvalues are *real*, and the eigenvectors corresponding to distinct eigenvalues are *orthogonal* to one another. These results are especially relevant for quantum mechanics, where the *observables* of the theory are represented by Hermitian transformations. (For proofs of these statements regarding Hermitian matrices, and for an excellent introduction to quantum theory, see Griffiths 2005.)

**Exercise D.19** Diagonalize the Hermitian matrix

$$\mathsf{T} = \begin{bmatrix} 1 & \mathrm{i} \\ -\mathrm{i} & 1 \end{bmatrix} \tag{D.127}$$

by finding its eigenvalues and eigenvectors, etc. Verify that the similarity transformation that diagonalizes $\mathsf{T}$ is unitary.

### *D.5.3 Determinant and Trace in Terms of Eigenvalues*

We end this section by showing that for *any* matrix $\mathsf{T}$ (diagonalizable or *not*), the determinant and trace of $\mathsf{T}$ can be written very simply in terms of its eigenvalues:

$$\det \mathsf{T} = \prod_i \lambda_i\,, \qquad \mathrm{Tr}(\mathsf{T}) = \sum_i \lambda_i\,. \tag{D.128}$$

For a diagonalizable matrix, the above two results follow immediately from (D.116) for $\mathsf{T}'$ and the fact that the determinant and trace of a matrix are invariant under a similarity transformation, (D.94) and (D.97). For a non-diagonalizable matrix, we proceed by first equating the expansion of the characteristic equation (D.102) in terms of powers of $\lambda$ and its factorization (D.103) in terms of its eigenvalues:

$$c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_n\lambda^n = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)\,. \tag{D.129}$$

From this equality we can see that the constant term $c_0$ is given by

$$c_0 = \prod_i \lambda_i\,, \tag{D.130}$$

while the factor multiplying $\lambda^{n-1}$ is given by

$$c_{n-1} = (-1)^{n-1}\sum_i \lambda_i\,. \tag{D.131}$$

Now return to (D.102),

$$\det(\mathsf{T} - \lambda\mathsf{1}) = c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_n\lambda^n\,. \tag{D.132}$$

Setting $\lambda = 0$ in this equation gives

$$c_0 = \det \mathsf{T}\,, \tag{D.133}$$

while expanding the determinant using (D.81),

$$\det(\mathsf{T} - \lambda \mathsf{1}) = \sum_{i_1,i_2,\ldots,i_n} \varepsilon_{i_1 i_2 \cdots i_n} (T_{1i_1} - \lambda\delta_{1i_1})(T_{2i_2} - \lambda\delta_{2i_2}) \cdots (T_{ni_n} - \lambda\delta_{ni_n})\,, \tag{D.134}$$

gives

$$c_n = (-1)^n\,, \qquad c_{n-1} = (-1)^{n-1}\mathrm{Tr}(\mathsf{T})\,. \tag{D.135}$$

(To see this, note that the terms proportional to $\lambda^n$ and $\lambda^{n-1}$ in (D.134) must come from the product

$$(T_{11} - \lambda)(T_{22} - \lambda) \cdots (T_{nn} - \lambda)\,, \tag{D.136}$$

which leads to (D.135).) Then by comparing (D.130) and (D.131) with (D.133) and (D.135), we get (D.128).

## Suggested References

*Full references are given in the bibliography at the end of the book.*

Boas (2006): Chapter 3 is devoted to linear algebra. The treatment is especially suited for undergraduates, with many examples and problems.

Dennery and Kryzwicki (1967): A mathematical methods book suited for advanced undergraduates and graduate students. Chapter 2 discusses finite-dimensional vector spaces; Chap. 3 extends the formalism to (infinite-dimensional) function spaces.

Griffiths (2005): Appendix A provides a review of linear algebra, especially relevant for calculations that arise in quantum mechanics. Our presentation follows that of Griffiths.

Halmos (1958): A classic text on vector spaces and linear algebra, written primarily for undergraduate students majoring in mathematics. As such, the mathematical rigor is higher than that in most mathematical methods books for scientists and engineers.

# Appendix E
# Special Functions

Special functions play an important role in the physical sciences. They often arise as power series solutions of ordinary differential equations, which in turn come from a separation-of-variables decomposition of common partial differential equations (e.g., Laplace's equation, Helmholtz's equation, the wave equation, the diffusion equation, $\cdots$). Special functions behave like vectors in an infinite-dimensional vector space, sharing many of the properties of vectors described in Appendix D. Each set of special functions is *orthogonal* with respect to an inner product defined as an appropriate integral of a product of two such functions. Special functions also form a set of *basis* functions in terms of which one can expand the general solution of the original partial differential equation.

In this appendix, we review the key properties of several special functions, with particular emphasis on those functions that appear often in classical mechanics applications. We will assume that the reader is already familiar with the general (Frobenius) method of power series solutions, which we describe very briefly in Appendix E.1. As such we will omit detailed derivations of the recursion relations for the coefficients of the various power series solutions. For those details, you should consult, e.g., Chap. 12 in Boas (2006). The definitive source for anything related to special functions is Abramowitz and Stegun (1972).

## E.1 Series Solutions of Ordinary Differential Equations

The general form of a homogeneous, linear, 2nd-order ordinary differential equation is

$$y''(x) + p(x)\,y'(x) + q(x)\,y(x) = 0\,, \tag{E.1}$$

where $p(x)$ and $q(x)$ are arbitrary functions of $x$. We are interested in **power series solutions** of the form



M.J. Benacquista and J.D. Romano, *Classical Mechanics*, Undergraduate Lecture Notes in Physics, https://doi.org/10.1007/978-3-319-68780-3

$$y(x) = \sum_{n=0}^{\infty} a_n x^n \quad \text{or} \quad y(x) = x^\sigma \sum_{n=0}^{\infty} a_n x^n \tag{E.2}$$

for some value of $\sigma$. We need to consider the second (more general) power series expansion (i.e., with $\sigma \neq 0$), called a **Frobenius series**, if $x = 0$ is a **regular singular point** of the differential equation—that is, if $p(x)$ or $q(x)$ is singular (i.e., infinite) at $x = 0$, but $xp(x)$ and $x^2q(x)$ are *finite* at $x = 0$. If $x = 0$ is a regular point of the differential equation, then one can simply set $\sigma = 0$ and use the first expansion.

The basic procedure for finding a power series solution to (E.1) is to differentiate the power series expansion for $y(x)$ term by term, and then substitute the expansion into the differential equation for $y(x)$. Since the resulting sum must vanish for all values of $x$, the coefficients of $x^n$ must all equal zero, leading to a **recursion relation**, which relates $a_n$ to some subset of the previous $a_r$ $(r < n)$, and a quadratic equation for $\sigma$, called the **indicial equation**. The following theorerm, called *Fuch's theorem*, tells us how to obtain the general solution of the differential equation from two Frobenius series solutions.

**Theorem E.1 Fuch's theorem**: *The general solution of the differential equation (E.1) with a regular singular point at $x = 0$ consists of of either:*

*(i) a sum of two Frobenius series $S_1(x)$ and $S_2(x)$, or*
*(ii) the sum of one Frobenius series $S_1(x)$, and a second solution of the form $S_1(x) \ln x + S_2(x)$, where $S_2(x)$ is another Frobenius series.*

*Case (ii) occurs only if the roots of the indicial equation for $\sigma$ are equal to one another or differ by an integer.*

If $x = 0$ is a regular point of the differential equation, then the general solution is simply the sum of two ordinary series solutions.

## E.2 Trigonometric and Hyperbolic Functions

Trigonometric and hyperbolic functions (e.g., $\cos\theta$, $\sin\theta$, $\cosh\chi$, $\sinh\chi$, etc.) can be defined geometrically in terms of circles and hyperbolae. For example, $\cos\theta$ is the projection onto the $x$-axis of a point $P$ on the unit circle making an angle $\theta$ with respect to the $x$-axis. Here, instead, we define these functions in terms of power series solutions to differential equations.

### *E.2.1 Trig Functions*

**Trigonometric functions** are solutions to the differential equation

$$y'' + k^2 y = 0 . \tag{E.3}$$

A power series expansion leads to the two-term recursion relation

$$a_{n+2} = \frac{-k^2}{(n+1)\,(n+2)} a_n . \tag{E.4}$$

Thus, there are two independent solutions, one starting with $a_0$ and the other starting with $a_1$. If $a_0 = A$ and $a_1 = B$, then the general solution to this equation is a linear superposition of sine and cosine functions,

$$y(x) = A\ \cos kx + B\ \sin kx , \tag{E.5}$$

where

$$\begin{aligned} \cos kx &\equiv 1 - \frac{(kx)^2}{2!} + \frac{(kx)^4}{4!} - \cdots , \\ \sin kx &\equiv kx - \frac{(kx)^3}{3!} + \frac{(kx)^5}{5!} - \cdots . \end{aligned} \tag{E.6}$$

These functions can be written in terms of complex exponentials using Euler's identity

$$e^{i\theta} = \cos\theta + i\sin\theta , \tag{E.7}$$

which can be inverted to yield explicit expressions for the cosine and sine functions:

$$\cos\theta = \frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right) , \qquad \sin\theta = \frac{1}{2i}\left(e^{i\theta} - e^{-i\theta}\right) . \tag{E.8}$$

The trig functions are periodic with period $2\pi$, and form an orthogonal set of functions on the interval $[-\pi, \pi]$:

$$\begin{aligned} \int_{-\pi}^{\pi} dx\ \sin(nx)\sin(mx) &= \pi\delta_{nm} , \\ \int_{-\pi}^{\pi} dx\ \cos(nx)\cos(mx) &= \pi\delta_{nm} , \\ \int_{-\pi}^{\pi} dx\ \sin(nx)\cos(mx) &= 0 . \end{aligned} \tag{E.9}$$

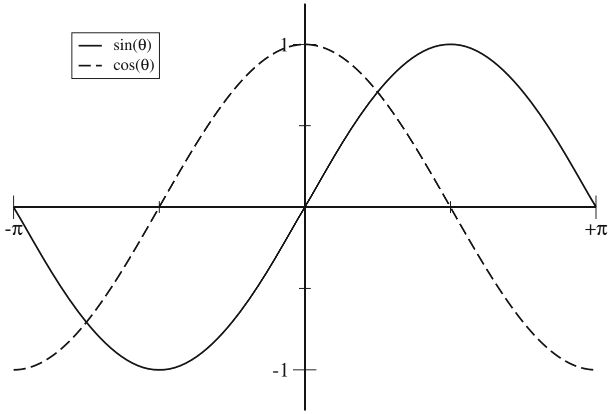



**Fig. E.1** The functions $\sin\theta$ and $\cos\theta$ plotted over the interval $-\pi$ to $\pi$

This is a key property of trig functions used in Fourier expansions of periodic functions. Plots of $\sin\theta$ and $\cos\theta$ are given in Fig. E.1. Finally, from sine and cosine we can define other trig fucntions:

$$\tan x \equiv \frac{\sin x}{\cos x} \equiv \frac{1}{\cot x}\,, \qquad \sec x \equiv \frac{1}{\cos x}\,, \qquad \csc x \equiv \frac{1}{\sin x}\,. \tag{E.10}$$

**Exercise E.1** Verify the recursion relation given in (E.4).

**Exercise E.2** Verify the orthogonality property of the sine and cosine functions, (E.9).

### *E.2.2 Hyperbolic Functions*

**Hyperbolic functions** are solutions to the differential equation

$$y'' - k^2 y = 0\,. \tag{E.11}$$

The recursion relation for this case is

$$a_{n+2} = \frac{k^2}{(n+1)\,(n+2)} a_n\,, \tag{E.12}$$

and the general solution to this equation is a linear combination of sinh and cosh functions:

$$y(x) = A \cosh kx + B \sinh kx \,, \tag{E.13}$$

where

$$\begin{aligned} \cosh kx &\equiv 1 + \frac{(kx)^2}{2!} + \frac{(kx)^4}{4!} + \cdots \,, \\ \sinh kx &\equiv kx + \frac{(kx)^3}{3!} + \frac{(kx)^5}{5!} + \cdots \,. \end{aligned} \tag{E.14}$$

Hyperbolic functions can be also be written in terms of ordinary exponentials,

$$\cosh x = \frac{1}{2}\left(\mathrm{e}^x + \mathrm{e}^{-x}\right) \,, \qquad \sinh x = \frac{1}{2}\left(\mathrm{e}^x - \mathrm{e}^{-x}\right) \,, \tag{E.15}$$

and trig functions,

$$\cosh x = \cos(\mathrm{i}x) \,, \qquad \sinh x = -\mathrm{i} \sin(\mathrm{i}x) \,. \tag{E.16}$$

From sinh and cosh we can define other hyperbolic functions, analogous to (E.10):

$$\tanh x \equiv \frac{\sinh x}{\cosh x} \equiv \frac{1}{\coth x} \,, \qquad \operatorname{sech} x \equiv \frac{1}{\cosh x} \,, \qquad \operatorname{csch} x \equiv \frac{1}{\sinh x} \,. \tag{E.17}$$

Plots of $\sinh x$, $\cosh x$, and $\tanh x$ are given in Fig. E.2.

**Exercise E.3** Verify (E.15) and (E.16).

## E.3 Legendre Polynomials and Associated Legendre Functions

Legendre's equation for $y(x)$ is

$$(1 - x^2)\, y'' - 2x\, y' + l(l+1)\, y = 0 \,, \tag{E.18}$$

where $l$ is a constant. This ordinary differential equation arises when one uses separation of variables for Laplace's equation $\nabla^2 \Phi = 0$ in spherical coordinates (here

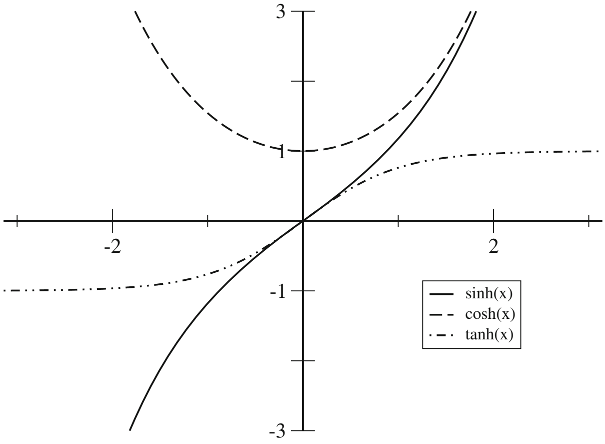



**Fig. E.2** The hyperbolic functions $\sinh x$, $\cosh x$, and $\tanh x$

$x \equiv \cos\theta$). One can show that $y(x)$ admits a regular power series solution with recursion relation

$$a_{n+2} = \frac{n(n+1) - l(l+1)}{(n+1)(n+2)}\, a_n\,, \qquad n = 0, 1, \cdots \tag{E.19}$$

Using the ratio test, it follows that the power series solution converges for $|x| < 1$. But one can also show (Exercise E.4, part (c)) that the power series solution diverges at $x = \pm 1$ (corresponding to the North and South poles of the sphere) unless the series terminates after some finite value of $n$.

**Exercise E.4** (a) Verify the recursion relation (E.19). (b) Show that for $l = 0$, the power series solution obtained by taking $a_0 = 0$ and $a_1 = 1$ is

$$y(x) = x + \frac{1}{3}x^3 + \frac{1}{5}x^5 + \cdots\,. \tag{E.20}$$

(c) Using the integral test, show that this solution diverges at $x = 1$ or $x = -1$.

### *E.3.1 Legendre Polynomials*

From the recursion relation (E.19), we see that if $l$ is a non-negative integer ($l = 0, 1, \cdots$), one of the power series solutions terminates (the even solution if $l$ is even, and the odd solution if $l$ is odd). The other solution can be set to zero (by hand) by choosing $a_1 = 0$ or $a_0 = 0$. The finite solutions thus obtained are *polynomials*

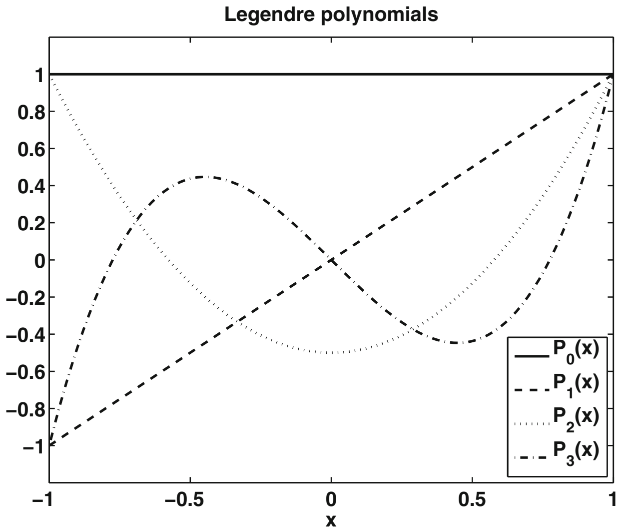



**Fig. E.3** First few Legendre polynomials $P_l(x)$ plotted as functions of $x \in [-1, 1]$

of order $l$. When appropriately normalized, they are called **Legendre polynomials**, denoted $P_l(x)$. By convention, the normalization condition is $P_l(1) = 1$. The first four Legendre polynomials are

$$
\begin{aligned}
P_0(x) &= 1\,,\\
P_1(x) &= x\,,\\
P_2(x) &= \frac{1}{2}(3x^2 - 1)\,,\\
P_3(x) &= \frac{1}{2}(5x^3 - 3x)\,.
\end{aligned}
\tag{E.21}
$$

Figures E.3 and E.4 give two different graphical representations of the first few Legendre polynomials. Note that $P_l(-x) = (-1)^l P_l(x)$.

**Exercise E.5** Verify (E.21).

**Exercise E.6** (a) Show that one also obtains a polynomial solution if $l$ is a *negative integer* ($l = -1, -2, \cdots$). (b) Verify that these solutions are the *same* as those for non-negative $l$ (e.g., $l = -1$ yields the same solution as $l = 0$, and $l = -2$ yields the same solution as $l = 1$, etc.). Thus, there is no loss of generality in restricting attention to $l = 0, 1, \cdots$.

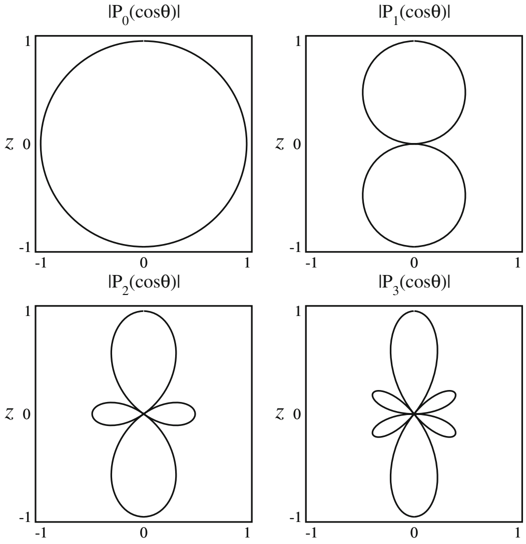



**Fig. E.4** The *magnitude* $|P_l(\cos\theta)|$ of the first few Legendre polynomials plotted as functions of $\cos\theta$ in the $xz$ plane (or $yz$) plane. The angle $\theta$ is measured with respect to the positive $z$-axis. Note that by plotting the magnitude, information about the *sign* (i.e., $\pm$) of the Legendre polynomials $P_l(\cos\theta)$ is lost in this graphical representation

## *E.3.2 Some Properties of Legendre Polynomials*

### E.3.2.1 Rodrigues' Formula

The Legendre polynomials can be generated using **Rodrigues' formula**:

$$P_l(x) = \frac{1}{2^l l!}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^l (x^2-1)^l\,. \tag{E.22}$$

### E.3.2.2 Orthogonality

The Legendre polynomials for different values of $l$ are orthogonal to one another,

$$\int_{-1}^{1} \mathrm{d}x\; P_l(x)P_{l'}(x) = \frac{2}{2l+1}\delta_{ll'}\,. \tag{E.23}$$

**Exercise E.7** Prove (E.23). (*Hint*: The proof of orthogonality is simple if you write down Legendre's equation for both $P_l(x)$ and $P_{l'}(x)$; multiply these equations by $P_{l'}(x)$ and $P_l(x)$; and then subtract and integrate the result between $-1$ and 1. The derivation of the normalization constant is harder, but can be proved using mathematical induction and Rodrigues' formula for $P_l(x)$.)

#### E.3.2.3 Completeness

The Legendre polynomials are *complete* in the sense that any square-integrable function $f(x)$ defined on the interval $x \in [-1, 1]$ can be expanded in terms of Legendre polynomials:

$$f(x) = \sum_{l=0}^{\infty} A_l\, P_l(x)\,, \quad \text{where} \quad A_l = \frac{2l+1}{2} \int_{-1}^{1} \mathrm{d}x\, f(x)\, P_l(x)\,. \tag{E.24}$$

**Exercise E.8** Show that the function

$$f(x) = \begin{cases} -1\,, & -1 \le x < 0 \\ +1\,, & 0 < x \le 1 \end{cases} \tag{E.25}$$

can be expanded in terms of Legendre polynomials as

$$f(x) = \frac{3}{2}\, P_1(x) - \frac{7}{8}\, P_3(x) + \frac{11}{16}\, P_5(x) + \cdots \tag{E.26}$$

#### E.3.2.4 Generating Function

The Legendre polynomials can also be obtained as the coefficients of a power series expansion in $t$ of a so-called **generating function**

$$\frac{1}{\sqrt{1 - 2xt + t^2}} = \sum_{n=0}^{\infty} P_n(x)\, t^n\,. \tag{E.27}$$

With this result, one can rather easily express $1/r$ potentials using a series of Legendre polynomials

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \sum_{l=0}^{\infty} \frac{r_<^l}{r_>^{l+1}} P_l(\cos\gamma)\,, \tag{E.28}$$

where $r_<$ ($r_>$) is the smaller (larger) of $r$ and $r'$, and $\gamma$ is the angle between $\mathbf{r}$ and $\mathbf{r}'$,

$$\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}' \equiv \cos\gamma = \cos\theta\cos\theta' + \sin\theta\sin\theta'\,\cos(\phi - \phi')\,. \tag{E.29}$$

#### E.3.2.5 Recurrence Relations

Using the generating function, one can derive the following relations, called **recurrence relations**,[1] which relate Legendre polynomials $P_n(x)$ and their derivatives $P_n'(x)$ to neighboring Legendre polynomials:

$$(n+1)\,P_{n+1} = (2n+1)x\,P_n - n\,P_{n-1}\,, \tag{E.30a}$$
$$P_n = P_{n+1}' - 2x\,P_n' + P_{n-1}'\,, \tag{E.30b}$$
$$n\,P_n = x\,P_n' - P_{n-1}'\,, \tag{E.30c}$$
$$(n+1)\,P_n = P_{n+1}' - x\,P_n'\,, \tag{E.30d}$$
$$(2n+1)\,P_n = P_{n+1}' - P_{n-1}'\,, \tag{E.30e}$$
$$(1-x^2)\,P_n' = n(P_{n-1} - x\,P_n)\,. \tag{E.30f}$$

Note that Legendre's equation

$$(1-x^2)\,P_n'' - 2x\,P_n' + n(n+1)\,P_n = 0 \tag{E.31}$$

can be obtained by differentiating (E.30f) with respect to $x$ and then using (E.30c). In addition, the normalization $P_n(1) = 1$ also follows simply from the generating function.

**Exercise E.9** Prove the above recurrence relations by differentiating the generating function with respect to $t$ and $x$ separately, and then combining the various expressions.

[1]Most authors use either "recursion relation" or "recurrence relation" exclusively, and apply it to any relation between indexed objects of different order. Here, we have decided to use "recurrence relation" when describing relationships between special functions of different order, while using "recursion relation" when describing relationships between the coefficients of the power series.

### *E.3.3 Associated Legendre Functions*

The associated Legendre equation is given by

$$(1-x^2)\,y'' - 2x\,y' + \left[l(l+1) - \frac{m^2}{(1-x^2)}\right] y = 0\,. \tag{E.32}$$

It differs from the ordinary Legendre equation, (E.18), by the extra term proportional to $m^2$. It turns out that power series solutions of this differential equation also diverge at the poles ($x = \pm 1$) unless $l = 0, 1, \cdots$ (as before) and $m = -l, -l+1, \ldots, l$. The finite solutions are called **associated Legendre functions**, $P_l^m(x)$, and are given by derivatives of the Legendre polynomials,

$$\begin{aligned} P_l^m(x) &= (-1)^m (1-x^2)^{m/2}\,\frac{\mathrm{d}^m}{\mathrm{d}x^m} P_l(x)\,, \quad \text{for} \quad m \geq 0\,, \\ P_l^{-m}(x) &= (-1)^m \frac{(l-m)!}{(l+m)!}\, P_l^m(x)\,, \quad \text{for} \quad m < 0\,. \end{aligned} \tag{E.33}$$

**Exercise E.10** Prove by direct substitution that the above expression for $P_l^m(x)$ satisfies the associated Legendre equation (E.32).

The associated Legendre functions are *not* polynomials in $x$ on account of the square root factor $(1-x^2)^{m/2}$ for odd $m$. But since we are often ultimately interested in the replacement $x \equiv \cos\theta$, these non-polynomial factors are just proportional to $\sin^m\theta$. Thus, the associated Legendre functions can be written as polynomials in $\cos\theta$ if $m$ is even, and polynomials in $\cos\theta$ multiplied by $\sin\theta$ if $m$ is odd. The first few associated Legendre functions are given by:

$\underline{l=0}$:

$$P_0^0(\cos\theta) = 1\,, \tag{E.34}$$

$\underline{l=1}$:

$$\begin{aligned} P_1^0(\cos\theta) &= \cos\theta\,, \\ P_1^1(\cos\theta) &= -\sin\theta\,, \end{aligned} \tag{E.35}$$

$\underline{l=2}$:

$$\begin{aligned} P_2^0(\cos\theta) &= \frac{1}{2}\left(3\cos^2\theta - 1\right)\,, \\ P_2^1(\cos\theta) &= -3\sin\theta\,\cos\theta\,, \\ P_2^2(\cos\theta) &= 3(1-\cos^2\theta)\,, \end{aligned} \tag{E.36}$$

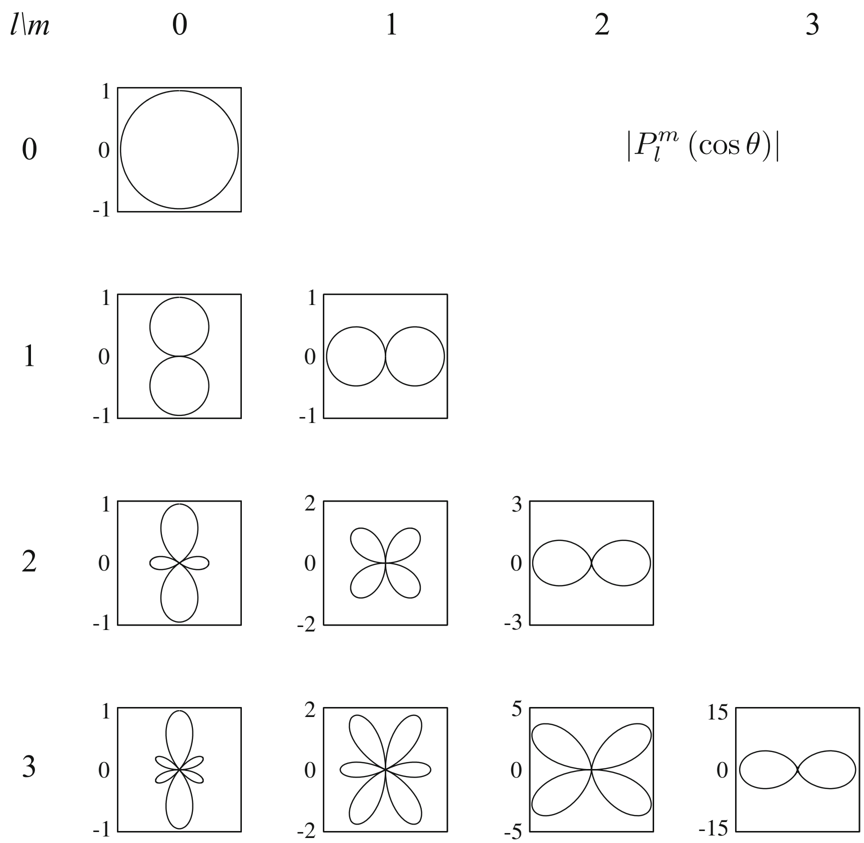



**Fig. E.5** The *magnitude* $|P_l^m(\cos\theta)|$ of the first few associated Legendre functions plotted as functions of $\cos\theta$ in the $xz$ plane (or $yz$) plane. The angle $\theta$ is measured with respect to the positive $z$-axis. Similar to the plot in Fig. E.4, the sign (i.e., $\pm$) of the associated Legendre functions $P_l^m(\cos\theta)$ is lost in this graphical representation. Note that the scale changes for larger values of $m$

$\underline{l=3}$:

$$\begin{aligned}
P_3^0(\cos\theta) &= \frac{1}{2}\left(5\cos^3\theta - 3\cos\theta\right) , \\
P_3^1(\cos\theta) &= -\frac{3}{2}\sin\theta\,\left(5\cos^2\theta - 1\right) , \\
P_3^2(\cos\theta) &= 15\left(\cos\theta - \cos^3\theta\right) , \\
P_3^3(\cos\theta) &= -15\sin\theta\,\left(1 - \cos^2\theta\right) .
\end{aligned} \tag{E.37}$$

Plots of the magnitude of the first few of these functions are given in Fig. E.5.

### E.3.4 Some Properties of Associated Legendre Functions

#### E.3.4.1 Rodrigues' Formula

Using Rodrigues' formula for Legendre polynomials (E.22), we can write down an analogous Rodrigues' formula for associated Legendre functions, valid for both positive and negative values of $m$:

$$P_l^m(x) = \frac{(-1)^m}{2^l l!}(1-x^2)^{m/2}\frac{\mathrm{d}^{l+m}}{\mathrm{d}x^{l+m}}(x^2-1)^l \,. \tag{E.38}$$

#### E.3.4.2 Orthonormality

For each $m$, the associated Legendre functions are orthogonal to one another,

$$\int_{-1}^{1} \mathrm{d}x\; P_l^m(x) P_{l'}^m(x) = \frac{2}{2l+1}\frac{(l+m)!}{(l-m)!}\,\delta_{ll'} \,. \tag{E.39}$$

#### E.3.4.3 Completeness

For each $m$, the associated Legendre functions form a complete set (in the index $l$) for square-integrable functions on $x \in [-1, 1]$:

$$f(x) = \sum_{l=0}^{\infty} A_l\, P_l^m(x)\,, \quad \text{where} \quad A_l = \frac{2l+1}{2}\frac{(l-m)!}{(l+m)!}\int_{-1}^{1} \mathrm{d}x\; f(x)\, P_l^m(x)\,. \tag{E.40}$$

## E.4 Spherical Harmonics

**Spherical harmonics** are solutions to the $(\theta, \phi)$ part of Laplace's equation $\nabla^2\Phi = 0$ in spherical coordinates. As such they are proportional to the product of associated Legendre functions $P_l^m(\cos\theta)$ and complex exponentials $\mathrm{e}^{\mathrm{i}m\phi}$:

$$Y_{lm}(\theta,\phi) \equiv N_l^m P_l^m(\cos\theta)\mathrm{e}^{\mathrm{i}m\phi}\,, \qquad N_l^m \equiv \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\,. \tag{E.41}$$

The proportionality constants have been chosen so that

$$\int_{S^2} \mathrm{d}\Omega\, Y^*_{lm}(\theta,\phi) Y_{l'm'}(\theta,\phi) = \delta_{ll'}\delta_{mm'}\,, \tag{E.42}$$

where

$$\mathrm{d}\Omega \equiv \mathrm{d}(\cos\theta)\,\mathrm{d}\phi = \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi\,. \tag{E.43}$$

This is the *orthonormality condition* for spherical harmonics. Note that for $m = 0$, spherical harmonics reduce to Legendre polynomials, up to a normalization factor:

$$Y_{l0} = \sqrt{\frac{2l+1}{4\pi}}\, P_l(\cos\theta)\,. \tag{E.44}$$

**Exercise E.11** Show that

$$Y_{l,-m}(\theta,\phi) = (-1)^m Y^*_{lm}(\theta,\phi)\,, \tag{E.45}$$

and

$$Y_{lm}(\pi-\theta,\phi+\pi) = (-1)^l Y_{lm}(\theta,\phi)\,. \tag{E.46}$$

The first equation tells you how to get $Y_{l,-m}$ from $Y_{lm}$; the second equation relates the values of the spherical harmonic $Y_{lm}$ at *antipodal* (i.e., opposite) points on the 2-sphere.

The first few spherical harmonics are given by:
$\underline{l=0}$:

$$Y_{00}(\theta,\phi) = \sqrt{\frac{1}{4\pi}}\,, \tag{E.47}$$

$\underline{l=1}$:

$$\begin{aligned} Y_{11}(\theta,\phi) &= -\sqrt{\frac{3}{8\pi}}\sin\theta\,\mathrm{e}^{\mathrm{i}\phi}\,,\\ Y_{10}(\theta,\phi) &= \sqrt{\frac{3}{4\pi}}\cos\theta\,,\\ Y_{1,-1}(\theta,\phi) &= \sqrt{\frac{3}{8\pi}}\sin\theta\,\mathrm{e}^{-\mathrm{i}\phi}\,, \end{aligned} \tag{E.48}$$

$\underline{l=2}$:

$$\begin{aligned}
Y_{22}(\theta,\phi) &= \frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\theta\, e^{2i\phi}\,, \\
Y_{21}(\theta,\phi) &= -\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta\, e^{i\phi}\,, \\
Y_{20}(\theta,\phi) &= \sqrt{\frac{5}{4\pi}}\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)\,, \\
Y_{2,-1}(\theta,\phi) &= \sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta\, e^{-i\phi}\,, \\
Y_{2,-2}(\theta,\phi) &= \frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\theta\, e^{-2i\phi}\,.
\end{aligned} \tag{E.49}$$

Since $Y_{lm}(\theta,\phi)$ differs from $P_l^m(\theta)$ by only a constant multiplicative factor and phase $e^{im\phi}$, the magnitude $|Y_{lm}(\theta,\phi)|$ has the same shape as $|P_l^m(\theta)|$ (See Fig. E.5).

## E.4.1 Some Properties of Spherical Harmonics

### E.4.1.1 Completeness

Spherical harmonics are complete in the sense that any square-integrable function $f(\theta,\phi)$ on the unit 2-sphere can be expanded in terms of spherical harmonics:

$$\begin{aligned}
f(\theta,\phi) &= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} A_{lm}\, Y_{lm}(\theta,\phi)\,, \quad \text{where} \\
A_{lm} &= \int_{S^2} d\Omega\, f(\theta,\phi)\, Y^*_{lm}(\theta,\phi)\,.
\end{aligned} \tag{E.50}$$

Equivalently, the completeness property can be written as

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y^*_{lm}(\theta',\phi') Y_{lm}(\theta,\phi) = \delta(\hat{\mathbf{n}},\hat{\mathbf{n}}')\,, \tag{E.51}$$

where $\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}'$ are the unit (radial) vectors

$$\begin{aligned}
\hat{\mathbf{n}} &= \sin\theta\cos\phi\,\hat{\mathbf{x}} + \sin\theta\sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}}\,, \\
\hat{\mathbf{n}}' &= \sin\theta'\cos\phi'\,\hat{\mathbf{x}} + \sin\theta'\sin\phi'\,\hat{\mathbf{y}} + \cos\theta'\,\hat{\mathbf{z}}\,,
\end{aligned} \tag{E.52}$$

and $\delta(\hat{\mathbf{n}},\hat{\mathbf{n}}')$ is the 2-dimensional Dirac delta function on the 2-sphere:

$$\delta(\hat{\mathbf{n}}, \hat{\mathbf{n}}') = \delta(\cos\theta - \cos\theta')\delta(\phi - \phi') = \frac{1}{\sin\theta}\delta(\theta - \theta')\delta(\phi - \phi') \,. \tag{E.53}$$

In terms of spherical harmonics, the general solution to Laplace's equation $\nabla^2\Phi = 0$ in spherical coordinates is

$$\Phi(r, \theta, \phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left[A_{lm}\, r^l + B_{lm}\, r^{-(l+1)}\right] Y_{lm}(\theta, \phi) \,, \tag{E.54}$$

where the terms in square brackets is the solution to the radial part of Laplace's equation.

#### E.4.1.2 Addition Theorem

If one sums only over $m$ in (E.51), one obtains the so-called **addition theorem** of spherical harmonics,

$$\sum_{m=-l}^{l} Y_{lm}^*(\theta', \phi')Y_{lm}(\theta, \phi) = \frac{2l+1}{4\pi}\, P_l(\cos\gamma) \,, \tag{E.55}$$

where

$$\cos\gamma \equiv \hat{\mathbf{n}} \cdot \hat{\mathbf{n}}' = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi - \phi') \,. \tag{E.56}$$

Completeness of the spherical harmonics and the addition theorem imply

$$\delta(\hat{\mathbf{n}}, \hat{\mathbf{n}}') = \sum_{l=0}^{\infty}\frac{2l+1}{4\pi}\, P_l(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}') \,, \tag{E.57}$$

which is an expansion of the Dirac delta function on the 2-sphere in terms of the Legendre polynomials.

**Exercise E.12** Using the addition theorem, show that the $1/r$ potential for a point source can be written as

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\frac{r_<^l}{r_>^{l+1}}\, Y_{lm}^*(\theta', \phi')Y_{lm}(\theta, \phi) \,, \tag{E.58}$$

where $r_<$ ($r_>$) is the smaller (larger) of $r$ and $r'$. This expression is fully-factorized into a product of functions of the unprimed and primed coordinates.

#### E.4.1.3 Transformation Under a Rotation

We can imagine rotating our coordinates through the Euler angles $\alpha, \beta, \gamma$ using the $zyz$ form of the rotation matrix $\mathsf{R}(\alpha, \beta, \gamma)$ (See Sect. 6.2.3.1). This is equivalent to a rotation of the 2-sphere. In this case, the spherical harmonics transform according to

$$Y_{lm}(\theta', \phi') = \sum_{m'=-l}^{l} D_{lm,m'}(\alpha, \beta, \gamma) Y_{lm'}(\theta, \phi)\,, \tag{E.59}$$

where $(\theta', \phi')$ are the coordinates of a point $P = (\theta, \phi)$ after the rotation of the sphere. The fact that $Y_{lm}(\theta', \phi')$ can be written as a linear combination of the $Y_{lm'}(\theta, \phi)$ with the *same* $l$ is a consequence of the spherical harmonics being eigenfunctions of the (rotationally-invariant) Laplacian on the unit 2-sphere with eigenvalues depending only on $l$,

$$^{(2)}\nabla^2 Y_{lm}(\theta, \phi) = -l(l+1) Y_{lm}(\theta, \phi)\,, \tag{E.60}$$

where

$$^{(2)}\nabla^2 f(\theta, \phi) \equiv \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}\,. \tag{E.61}$$

The coefficients $D_{lm,m'}(\alpha, \beta, \gamma)$ in (E.59) are called **Wigner rotation matrices**. They arise in applications of group theory to quantum mechanics (Wigner 1931). In terms of the Euler angles, the components of the Wigner rotation matrices can be written as

$$D_{lm,m'}(\alpha, \beta, \gamma) = \mathrm{e}^{-\mathrm{i}m\alpha} d_{lm,m'}(\beta) \mathrm{e}^{-\mathrm{i}m'\gamma}\,, \tag{E.62}$$

where

$$\begin{aligned} d_{lm,m'}(\beta) \equiv & \sqrt{(l+m)!(l-m)!(l+m')!(l-m')!} \\ & \times \sum_s \left[ \frac{(-1)^{m-m'+s}}{(l+m'-s)!s!(m-m'+s)!(l-m-s)!} \right. \\ & \left. \times \left(\cos\frac{\beta}{2}\right)^{2l+m'-m-2s} \left(\sin\frac{\beta}{2}\right)^{m-m'+2s} \right], \end{aligned} \tag{E.63}$$

and where the sum over $s$ is chosen such that the factorials inside the summation always remain non-negative. The Wigner matrices also satisfy

$$\sum_{m''=-l}^{l} D_{lm,m''}(\alpha, \beta, \gamma) D^*_{lm',m''}(\alpha, \beta, \gamma) = \delta_{mm'} \tag{E.64}$$

as a consequence of

$$\int_{S^2} \mathrm{d}\Omega\, Y^*_{lm}(\theta', \phi') Y_{l'm'}(\theta', \phi') = \delta_{ll'}\delta_{mm'} \,. \tag{E.65}$$

## E.5 Bessel Functions and Spherical Bessel Functions

Separation of variables of Laplaces's equation in cylindrical coordinates $(\rho, \phi, z)$ leads to either of the following two differential equations for the radial function $R(\rho)$:

$$\begin{aligned} R''(\rho) + \frac{1}{\rho} R'(\rho) + \left(k^2 - \frac{\nu^2}{\rho^2}\right) R(\rho) = 0\,, \\ R''(\rho) + \frac{1}{\rho} R'(\rho) - \left(k^2 + \frac{\nu^2}{\rho^2}\right) R(\rho) = 0\,. \end{aligned} \tag{E.66}$$

The two equations correspond to different choices for the sign of the separation constant, $\pm k^2$. These equations can be put into more standard form by making a change of variables $x \equiv k\rho$, with $y(x)|_{x=k\rho} \equiv R(\rho)$:

$$\begin{aligned} y''(x) + \frac{1}{x}\, y'(x) + \left(1 - \frac{\nu^2}{x^2}\right) y(x) = 0\,, \\ y''(x) + \frac{1}{x}\, y'(x) - \left(1 + \frac{\nu^2}{x^2}\right) y(x) = 0\,. \end{aligned} \tag{E.67}$$

The first equation is called **Bessel's equation** of order $\nu$; the second is called the **modified Bessel's equation** of order $\nu$.

**Exercise E.13** Show that if $y(x)$ is a solution of Bessel's equation, then $\bar{y}(x) \equiv y(\mathrm{i}x)$ is a solution of the modified Bessel's equation.

### *E.5.1 Bessel Functions of the 1st Kind*

Since $x = 0$ is a *regular singular point* of Bessel's equation, the method of Frobenius (Appendix E.1) requires that we consider a power series expansion of the form

$$y(x) = x^\sigma \sum_{n=0}^{\infty} a_n x^n \,. \tag{E.68}$$

Substituting this expansion into Bessel's equation and equating coefficients multiplying like powers of $x$ leads to a quadratic equation for $\sigma$ (called the **indicial** equation) and a recursion relation relating $a_{n+2}$ to $a_n$ (and $\sigma$) for $n = 0, 1, \cdots$. Setting $a_1 = 0$ (which forces all of the higher-order odd coefficients to vanish) and choosing the normalization coefficient $a_0$ appropriately, we obtain the solution

$$J_\nu(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(n+1+\nu)} \left(\frac{x}{2}\right)^{2n+\nu} . \tag{E.69}$$

$J_\nu(x)$ is called a **Bessel function of the 1st kind** of order $\nu$. The function $\Gamma(n + 1 + \nu)$ which appears in the denominator of the expansion coefficients is the **gamma function** defined by

$$\Gamma(z) \equiv \int_0^\infty \mathrm{d}x\, x^{z-1} \mathrm{e}^{-x} , \qquad \mathrm{Re}(z) > 0 . \tag{E.70}$$

The gamma function generalizes the ordinary factorial function $n! = n(n-1)\cdots 1$ to non-integer arguments in the sense that

$$\begin{aligned} &\Gamma(n+1) = n! \quad \text{for} \quad n = 0, 1, \cdots \\ &\Gamma(z+1) = z\,\Gamma(z) \quad \text{for} \quad \mathrm{Re}(z) > 0 . \end{aligned} \tag{E.71}$$

**Exercise E.14** (a) Prove $\Gamma(z+1) = z\Gamma(z)$ for $\mathrm{Re}(z) > 0$. (*Hint*: Integrate $\Gamma(z+1)$ by parts taking $u = x^z$ and $\mathrm{d}v = \mathrm{e}^{-x}\,\mathrm{d}x$.) (b) Show by explicit calculation that $\Gamma(1) = 1$ and $\Gamma(1/2) = \sqrt{\pi}$.

### E.5.1.1 Asymptotic Form

To gain a better intuitive understanding of the Bessel function $J_\nu(x)$, it is useful to look at its *asymptotic* form—i.e., its behavior for both small and large values of $x$. Using the general definition (E.69), one can show that

$$\begin{aligned} x \ll 1 : \qquad & J_\nu(x) \to \frac{1}{\Gamma(\nu+1)} \left(\frac{x}{2}\right)^\nu , \\ x \gg 1, \nu : \qquad & J_\nu(x) \to \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right) . \end{aligned} \tag{E.72}$$

Thus, $J_0(0) = 1$ and $J_\nu(0) = 0$ for all $\nu \neq 0$; while for large $x$, $J_\nu(x)$ behaves like a *damped sinusoid*, and has infinitely many zeros $x_{\nu n}$:

$$J_\nu(x_{\nu n}) = 0 , \qquad n = 1, 2, \cdots . \tag{E.73}$$

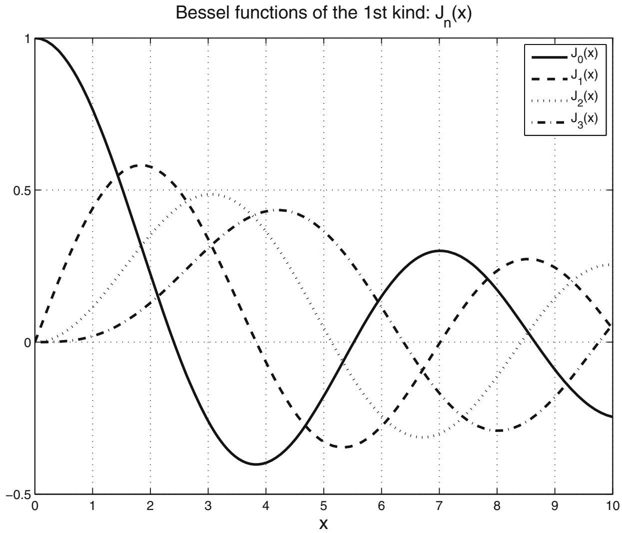



**Fig. E.6** First few Bessel functions of the 1st kind for integer $\nu$

Plots of the first few Bessel functions of the 1st kind for integer values of $\nu$ are given in Fig. E.6.

**Exercise E.15** Using (E.72), show that the zeros of $J_\nu(x)$ are given by

$$x_{\nu n} \simeq n\pi + \left(\nu - \frac{1}{2}\right)\frac{\pi}{2}\,. \tag{E.74}$$

### E.5.1.2 Integral Representation

It is also possible to write $J_n(x)$ for integer $n$ as an integral involving trig functions,

$$J_n(x) = \frac{1}{\pi}\int_0^\pi \cos(n\theta - x\sin\theta)\,\mathrm{d}\theta\,. \tag{E.75}$$

This result is useful for finding a Fourier series solution to Kepler's equation as discussed in Sect. 4.3.4.

## E.5.2 Bessel Functions of the 2nd Kind

If $\nu$ is not an integer, then $J_{-\nu}(x)$ is the second independent solution to Bessel's equation. But if $\nu = m$ is an integer, then

$$J_{-m}(x) = (-1)^m J_m(x)\,, \tag{E.76}$$

so $J_{-m}(x)$ is not an independent solution for this case. A second solution, which *is* independent of $J_\nu(x)$ for all values of $\nu$ (integer or not) is[2]

$$N_\nu(x) \equiv \frac{J_\nu(x)\,\cos(\nu\pi) - J_{-\nu}(x)}{\sin(\nu\pi)}\,. \tag{E.77}$$

$N_\nu(x)$ is called a **Neumann function** (or a **Bessel function of the 2nd kind**). In some references, $N_\nu(x)$ is denoted by $Y_\nu(x)$.

### E.5.2.1 Asymptotic Form

The asymptotic form of $N_\nu(x)$ is given by

$$\begin{aligned} x \ll 1: \quad & N_\nu(x) \to \begin{cases} \frac{2}{\pi}\left[\ln\left(\frac{x}{2}\right) + 0.5772\cdots\right], & \nu = 0 \\ -\frac{\Gamma(\nu)}{\pi}\left(\frac{2}{x}\right)^\nu, & \nu \neq 0 \end{cases} \\ x \gg 1, \nu: \quad & N_\nu(x) \to \sqrt{\frac{2}{\pi x}}\sin\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right). \end{aligned} \tag{E.78}$$

Note that for all $\nu$, $N_\nu(x) \to -\infty$ as $x \to 0$. In addition, just as we saw for $J_\nu(x)$, $N_\nu(x)$ behaves for large $x$ like a damped sinusoid, but is 90° out of phase with $J_\nu(x)$. Plots of the first few Bessel functions of the 2nd kind for integer values of $\nu$ are given in Fig. E.7.

With $J_\nu(x)$ and $N_\nu(x)$ as the two independent solutions to Bessel's equation, it follows that the most general solution to the radial part of Laplace's equation in cylindrical coordinates is

$$R(\rho) = A\,J_\nu(k\rho) + B\,N_\nu(k\rho)\,. \tag{E.79}$$

But since $N_\nu(x)$ blows up at $x = 0$, if $\rho = 0$ is in the region of interest, then all of the $B$ coefficients *must vanish* to yield a finite solution to Laplace's equation on the axis. Since both $J_\nu(x)$ and $N_\nu(x)$ go to zero as $x \to \infty$, there is no constraint on either $A$ or $B$ as $\rho \to \infty$.

[2] For $\nu = m$ an integer, one needs to use L'Hôpital's rule to show that the right-hand side of the expression defining $N_m(x)$ is well-defined.

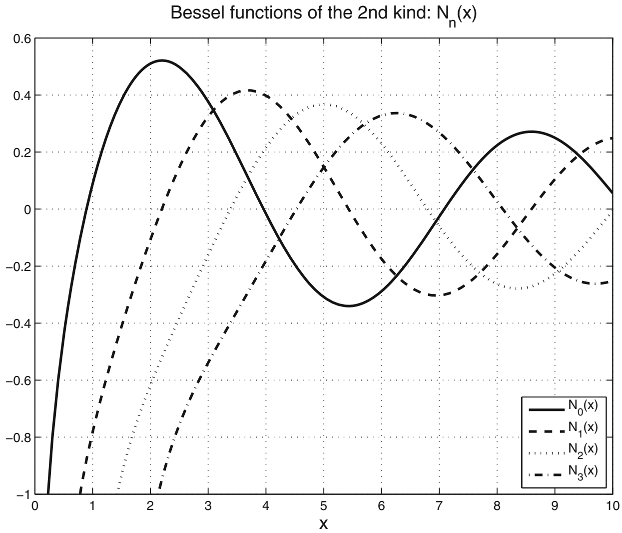



**Fig. E.7** First few Bessel functions of the 2nd kind for integer $\nu$

## *E.5.3 Some Properties of Bessel Functions*

### E.5.3.1 Recurrence Relations

The following relations hold for either $J_\nu(x)$, $N_\nu(x)$, or any linear combination of these functions with constant coefficients:

$$
\begin{aligned}
(x^\nu J_\nu(x))' &= x^\nu J_{\nu-1}(x)\,, \\
\left(x^{-\nu} J_\nu(x)\right)' &= -x^{-\nu} J_{\nu+1}(x)\,, \\
J'_\nu(x) &= -\frac{\nu}{x} J_\nu(x) + J_{\nu-1}(x)\,, \\
J'_\nu(x) &= \frac{\nu}{x} J_\nu(x) - J_{\nu+1}(x)\,, \\
2J'_\nu(x) &= J_{\nu-1}(x) - J_{\nu+1}(x)\,, \\
\frac{2\nu}{x} J_\nu(x) &= J_{\nu-1}(x) + J_{\nu+1}(x)\,.
\end{aligned}
\tag{E.80}
$$

### E.5.3.2 Orthogonality and Normalization

Bessel functions $J_\nu(x)$ satisfy the following orthogonality and normalization conditions

$$\int_0^a d\rho\, \rho J_\nu(x_{\nu n}\rho/a) J_\nu(x_{\nu n'}\rho/a) = \frac{1}{2}a^2 J^2_{\nu+1}(x_{\nu n})\,\delta_{nn'}\,, \tag{E.81}$$

where $x_{\nu n}$ and $x_{\nu n'}$ are the $n$th and $n'$th zeroes of $J_\nu(x)$. Note that the orthogonality of Bessel functions is with respect to different arguments of a *single* function $J_\nu(x)$, and not with respect to *different* functions $J_\nu(x)$ and $J_{\nu'}(x)$ of the same argument. (This latter case held for the Legendre polynomials $P_l(x)$ and $P_{l'}(x)$.) Thus, the orthogonality of Bessel functions is similar to the orthogonality of the sine functions $\sin(n2\pi x/a)$ on the interval $[0, a]$ for different values of $n$.

If the interval $[0, a]$ becomes infinite $[0, \infty)$, then the orthogonality and normalization conditions actually become simpler,

$$\int_0^\infty d\rho\, \rho J_\nu(k\rho) J_\nu(k'\rho) = \frac{1}{k}\delta(k-k')\,, \tag{E.82}$$

where $k$ now takes on a continuous range of values. This is similar to the transition from Fourier series (basis functions $\mathrm{e}^{\mathrm{i}k_n x}$ with $k_n = n2\pi/a$) to Fourier transforms (basis functions $\mathrm{e}^{\mathrm{i}kx}$ with $k$ a real variable):

$$\int_{-a/2}^{a/2} \mathrm{d}x\, \mathrm{e}^{\mathrm{i}2\pi(n-n')x/a} = a\,\delta_{nn'} \quad \rightarrow \quad \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{\mathrm{i}(k-k')x} = 2\pi\,\delta(k-k')\,. \tag{E.83}$$

**Exercise E.16** Prove the orthogonality part of (E.81). (*Hint*: Let $f(\rho) = J_\nu(x_{\nu n}\rho/a)$ and $g(\rho) = J_\nu(x_{\nu n'}\rho/a)$ with $n \neq n'$. Then write down Bessel's equation for both $f$ and $g$; multiply these equations by $g$ and $f$, respectively; then subtract and integrate.)

**Exercise E.17** Prove the normalization part of (E.81). (*Hint*: You will need to integrate by parts and then use Bessel's equation to substitute for $x^2 J_\nu(x)$ in one of the integrals.)

### *E.5.4 Modified Bessel Functions of the 1st and 2nd Kind*

As mentioned previously, the *modified* Bessel's equation of order $\nu$ is given by:

$$y''(x) + \frac{1}{x}y'(x) - \left(1 + \frac{\nu^2}{x^2}\right) y(x) = 0\,. \tag{E.84}$$

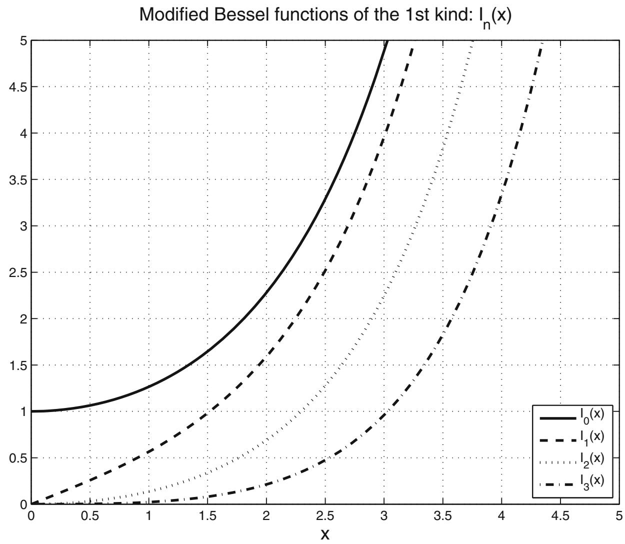



**Fig. E.8** First few modified Bessel functions of the 1st kind for integer $\nu$

It differs from the ordinary Bessel's equation only in the sign of one of the terms multiplying $y(x)$. **Modified** (or **hyperbolic**) **Bessel functions** (of the 1st and 2nd kind) are solutions to the above equation. They are defined by

$$I_\nu(x) \equiv \mathrm{i}^{-\nu} J_\nu(\mathrm{i}x)\,, \qquad K_\nu(x) \equiv \frac{\pi}{2}\mathrm{i}^{\nu+1} H_\nu^{(1)}(\mathrm{i}x)\,. \tag{E.85}$$

Note the pure imaginary arguments on the right-hand side of the above definitions, consistent with our earlier statement that if $y(x)$ is a solution of Bessel's equation then $y(\mathrm{i}x)$ is a solution of the modified Bessel's equation. Plots of the first few modified Bessel functions of the first and second kind, $I_\nu(x)$ and $K_\nu(x)$, for integer values of $\nu$ are given in Figs. E.8 and E.9.

#### E.5.4.1 Asymptotic Form

The asymptotic behavior of the modified Bessel functions $I_\nu(x)$ and $K_\nu(x)$ are given by

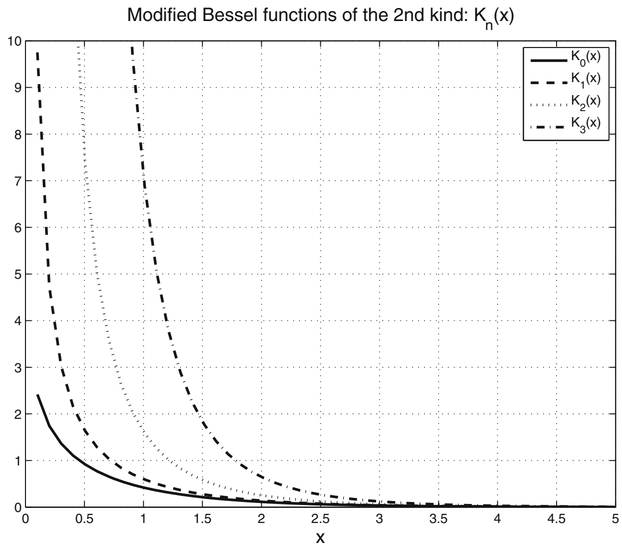



**Fig. E.9** First few modified Bessel functions of the 2nd kind for integer $\nu$

$$
\begin{aligned}
x \ll 1: \quad & I_\nu(x) \to \frac{1}{\Gamma(\nu+1)}\left(\frac{x}{2}\right)^\nu , \\
& K_\nu(x) \to \begin{cases} -\left[\ln\left(\frac{x}{2}\right) + 0.5772\cdots\right] , & \nu = 0 \\ \frac{\Gamma(\nu)}{2}\left(\frac{2}{x}\right)^\nu , & \nu \neq 0 \end{cases} \\
x \gg 1, \nu: \quad & I_\nu(x) \to \frac{1}{\sqrt{2\pi x}}\, e^x \left[1 + O\left(\frac{1}{x}\right)\right] , \\
& K_\nu(x) \to \sqrt{\frac{\pi}{2x}}\, e^{-x} \left[1 + O\left(\frac{1}{x}\right)\right] .
\end{aligned}
\tag{E.86}
$$

Thus, $I_0(0) = 1$ and $I_\nu(0) = 0$ for all $\nu \neq 0$, while $K_\nu(x) \to \infty$ as $x \to 0$ for all $\nu$. For large $x$, $I_\nu(x) \to \infty$ while $K_\nu(x) \to 0$ for all $\nu$.

Given $I_\nu(x)$ and $K_\nu(x)$, the most general solution to the radial part of Laplace's equation for the choice of negative separation constant $-k^2$ is

$$R(\rho) = A\, I_\nu(k\rho) + B\, K_\nu(k\rho) . \tag{E.87}$$

Since $K_\nu(x)$ blows up at $x = 0$, if $\rho = 0$ is in the region of interest, then all of the $B$ coefficients must vanish to yield a finite solution to Laplace's equation on the axis. Similarly, since $I_\nu(x)$ blows up as $x \to \infty$, if the solution to Laplace's equation is to vanish as $\rho \to \infty$, then all of the $A$ coefficients must vanish.

## *E.5.5 Spherical Bessel Functions*

**Spherical Bessel functions** (of the 1st and 2nd kind) are defined in terms of ordinary Bessel functions via

$$j_n(x) \equiv \sqrt{\frac{\pi}{2x}}\, J_{n+\frac{1}{2}}(x)\,, \qquad n_n(x) \equiv \sqrt{\frac{\pi}{2x}}\, N_{n+\frac{1}{2}}(x)\,, \tag{E.88}$$

where $n = 0, 1, 2, \cdots$. Given the explicit form of $J_{n+\frac{1}{2}}(x)$ one can show that

$$\begin{aligned} j_n(x) &= x^n \left(-\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}\right)^n \left(\frac{\sin x}{x}\right)\,, \\ n_n(x) &= -x^n \left(-\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}\right)^n \left(\frac{\cos x}{x}\right)\,. \end{aligned} \tag{E.89}$$

In particular, it follows that

$$j_0(x) = \frac{\sin x}{x}\,, \qquad n_0(x) = -\frac{\cos x}{x}\,. \tag{E.90}$$

Plots of the first few spherical Bessel functions are given in Figs. E.10 and E.11.

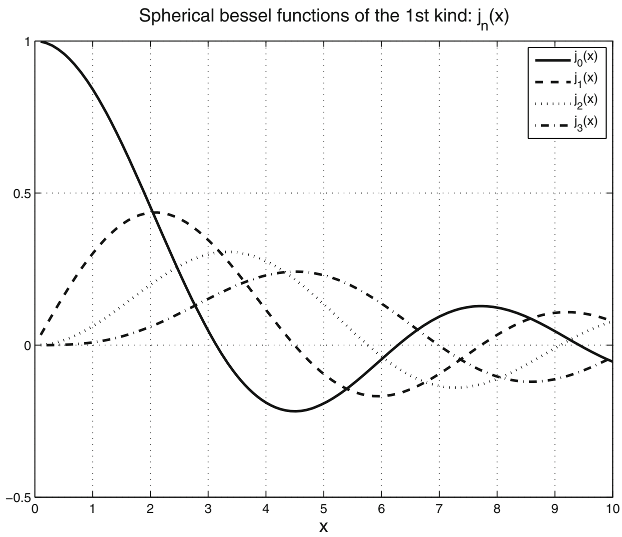



**Fig. E.10** First few spherical Bessel functions of the 1st kind

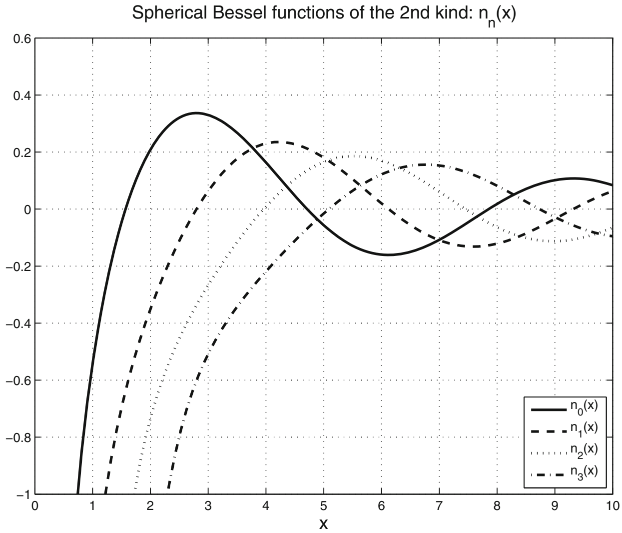



**Fig. E.11** First few spherical Bessel functions of the 2nd kind

> **Exercise E.18** Verify (E.90) for $j_0(x)$ directly from its definition in terms of the ordinary Bessel function $J_{1/2}(x)$.

### E.5.5.1 Spherical Bessel Differential Equation

Given the relationship between $j_n(x)$ and $J_{n+\frac{1}{2}}(x)$, one can show that the spherical Bessel functions satisfy the differential equation

$$j_n''(x) + \frac{2}{x} j_n'(x) + \left[1 - \frac{n(n+1)}{x^2}\right] j_n(x) = 0 \,. \tag{E.91}$$

Alternatively, one arrives at the same differential equation by using separation of variables in *spherical coordinates* to solve the **Helmholtz equation**

$$\nabla^2 \Phi(r, \theta, \phi) + k^2 \Phi(r, \theta, \phi) = 0 \,. \tag{E.92}$$

The $\phi$ equation is the standard harmonic oscillator equation with separation constant $-m^2$; the $\theta$ equation is the associated Legendre's equation with separation constants $l$ and $m$; and the radial equation is

$$R''(r) + \frac{2}{r} R'(r) + \left[k^2 - \frac{l(l+1)}{r^2}\right] R(r) = 0 \,. \tag{E.93}$$

Making the change of variables $x \equiv kr$ with $y(x)|_{x=kr} \equiv R(r)$ leads to

$$y''(x) + \frac{2}{x}y'(x) + \left[1 - \frac{l(l+1)}{x^2}\right]y(x) = 0\,, \tag{E.94}$$

which is the differential equation (E.91) we found earlier with solution $y(x) = j_l(x)$.

#### E.5.5.2 Integral Representation

Spherical Bessel functions can be written as an integral involving Legendre polynomials and complex exponentials,

$$2(-\mathrm{i})^l j_l(x) = \int_{-1}^{1} \mathrm{d}y\; P_l(y)\mathrm{e}^{-\mathrm{i}xy}\,. \tag{E.95}$$

## E.6 Elliptic Integrals and Elliptic Functions

Elliptic integrals and elliptic functions arise in some simple applications, such as finding the length of a conic section (e.g., an ellipse) and solving for the motion of a simple pendulum when one goes beyond the small-angle approximation. In the following two subsections, we briefly define elliptic integrals and elliptic functions using the notation given in Chap. 12 of Boas 2006. Other references may use slightly different notation.

### *E.6.1 Elliptic Integrals*

**Elliptic integrals** of the 1st and 2nd kind are often written in two different forms; the **Legendre forms**:

$$\begin{aligned} F(\phi, k) &\equiv \int_0^{\phi} \frac{\mathrm{d}\theta}{\sqrt{1 - k^2\sin^2\theta}}\,, \qquad 0 \le k \le 1\,,\\ E(\phi, k) &\equiv \int_0^{\phi} \sqrt{1 - k^2\sin^2\theta}\,\mathrm{d}\theta\,, \qquad 0 \le k \le 1\,, \end{aligned} \tag{E.96}$$

and the **Jacobi forms**:

$$F(\phi, k) \equiv \int_0^x \frac{\mathrm{d}t}{\sqrt{1-k^2t^2}\sqrt{1-t^2}}\,, \qquad 0 \le k \le 1\,,$$
$$E(\phi, k) \equiv \int_0^x \frac{\sqrt{1-k^2t^2}}{\sqrt{1-t^2}}\,\mathrm{d}t\,, \qquad 0 \le k \le 1\,, \tag{E.97}$$

with $x \equiv \sin\phi$. The two arguments of these functions are called the *amplitude* $\phi$ and the *modulus* $k$. Note that the Jacobi and Legendre forms of elliptic integrals are related by the change of variables $t = \sin\theta$.

**Complete elliptic integrals** of the 1st and 2nd kind, $K(k)$ and $E(k)$, are defined by setting the amplitude $\phi = \pi/2$ (or $x = 1$) in the above expressions:

$$K(k) \equiv F(\pi/2, k)\,, \qquad E(k) \equiv E(\pi/2, k)\,. \tag{E.98}$$

**Exercise E.19** (a) Show that the arc length of an ellipse $(x/a)^2 + (y/b)^2 = 1$ from $\theta = \phi_1$ to $\theta = \phi_2$ can be written as

$$s(\phi_1, \phi_2) = a\,[E(\phi_2, e) - E(\phi_1, e)]\,, \tag{E.99}$$

where $e \equiv \sqrt{1-(b/a)^2}$ is the eccentricity of the ellipse. (Here $\theta$ is defined by $x = a\sin\theta$, $y = b\cos\theta$, and we are assuming that $a \ge b$.) (b) Using the result of part (a), show that the total arc length $s = 4a\,E(e)$. (c) Show that for nearly circular ellipses (i.e., for $e \ll 1$), $s \approx 2\pi a(1 - e^2/4)$.

**Exercise E.20** (a) Show that the period of a simple pendulum of mass $m$, length $\ell$, released from rest at $\theta = \theta_0$ is given by

$$P(\theta_0) = 4\sqrt{\frac{\ell}{g}}K\,(\sin(\theta_0/2))\,. \tag{E.100}$$

Do not assume that the small-angle approximation is valid for this part of the problem. (*Hint*: Use conservation of total mechanical energy to find an equation for $\dot\theta$ in terms of $\theta$ and $\theta_0$.) (b) Show that for $\theta_0 \ll 1$, the answer from part (a) reduces to

$$P(\theta_0) \approx 2\pi\sqrt{\frac{\ell}{g}}\left(1 + \frac{1}{16}\theta_0^2\right)\,, \tag{E.101}$$

which in the limit of very small $\theta_0$ is the small-angle approximation for the period of a simple pendulum, $P \approx 2\pi\sqrt{\ell/g}$.

### E.6.2 Elliptic Functions

The **elliptic function** sn $y$ is defined as the inverse of the elliptic integral $y = F(\phi, k)$ for a fixed value of $k$,

$$y = \int_0^x \frac{\mathrm{d}t}{\sqrt{1-k^2t^2}\sqrt{1-t^2}} \equiv \mathrm{sn}^{-1}x \quad \Leftrightarrow \quad x = \mathrm{sn}\, y\,. \qquad \text{(E.102)}$$

Since $x = \sin\phi$, we can also write $\mathrm{sn}\, y = \sin\phi$ in terms of the amplitude $\phi$. Note that the above definition of sn $y$ is very similar to the integral representation of the inverse sine function

$$y = \int_0^x \frac{\mathrm{d}t}{\sqrt{1-t^2}} = \sin^{-1}x \quad \Leftrightarrow \quad x = \sin y\,. \qquad \text{(E.103)}$$

In fact, when $k = 0$, $\mathrm{sn}\, y = \sin y$. In addition, sn $y$ is periodic with period

$$P = 4\int_0^1 \frac{\mathrm{d}t}{\sqrt{1-k^2t^2}\sqrt{1-t^2}} = 4F(\pi/2, k) = 4K(k)\,, \qquad \text{(E.104)}$$

similar to the sine function. Plots of $x = \mathrm{sn}\, y$ for $k^2 = 0$, 0.25, 0.5, and 0.75 are shown in Fig. E.12. These have periods $P = 6.28$, 6.74, 7.42, and 8.63 to three significant digits.

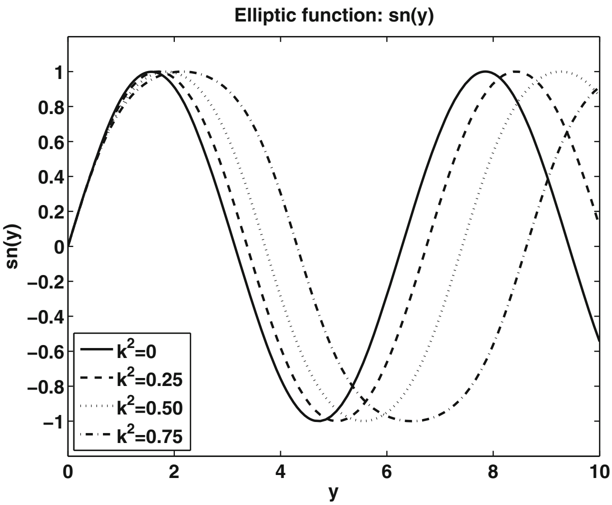



**Fig. E.12** Plots to the elliptic function sn $y$ for $k^2 = 0$, 0.25. 0.5 and 0.75. Recall that for $k^2 = 0$, $\mathrm{sn}\, y = \sin y$

Given sn $y$, one can define other elliptic functions using relations similar to those between trig functions,

$$\mathrm{cn}\, y \equiv \sqrt{1 - \mathrm{sn}^{\,2} y}\,, \qquad \mathrm{dn}\, y \equiv \sqrt{1 - k^2 \mathrm{sn}^{\,2} y}\,. \tag{E.105}$$

Using the above definitions, it is easy to show that $\mathrm{cn}\, y = \cos\phi$. In addition, using the Legendre form of the elliptic integral $F(\phi, k)$, it follows that $\mathrm{dn}\, y = \mathrm{d}\phi/\mathrm{d}y$. The proof is simply

$$\frac{\mathrm{d}\phi}{\mathrm{d}y} = \frac{1}{\mathrm{d}y/\mathrm{d}\phi} = \sqrt{1 - k^2 \sin^2\phi} = \sqrt{1 - k^2 \mathrm{sn}^{\,2} y} = \mathrm{dn}\, y\,. \tag{E.106}$$

**Exercise E.21** Show that

$$\frac{\mathrm{d}}{\mathrm{d}y}(\mathrm{sn}\, y) = \mathrm{cn}\, y\, \mathrm{dn}\, y\,. \tag{E.107}$$

## Suggested References

*Full references are given in the bibliography at the end of the book.*

Abramowitz and Stegun (1972): A must-have reference for all things related to special functions.

Boas (2006): Chapters 11, 12, and 13 discuss special functions, series solutions of differential equations, and partial differential equations, respectively, filling in most of the details omitted in this appendix. An excellent introduction to these topics, especially suited for undergraduates. There are many examples and problems to choose from.

Mathews and Walker (1970): Chapters 1, 7, and 8 discuss ordinary differential equations, special functions, and partial differential equations, respectively. The level of this text is more appropriate for graduate students or mathematically-minded undergraduates.

# References


B.P. Abbott, R. Abbott, T.D. Abbott, M.R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R.X. Adhikari et al., Observation of gravitational waves from a binary black hole merger. Phys. Rev. Lett. **116**(6), 061102 (2016). https://doi.org/10.1103/PhysRevLett.116.061102

B.P. Abbott, R. Abbott, T.D. Abbott, M.R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R.X. Adhikari et al., The basic physics of the binary black hole merger GW150914. Annalen der Physik **529**, 1600209 (2017). https://doi.org/10.1002/andp.201600209

M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions* (Dover Publications Inc, New York, 1972). ISBN 0-486-61272-4

G. Arfken, *Mathematical Methods for Physicists* (Academic Press Inc, New York, 1970)

V.I. Arnold, *Mathematical Methods of Classical Mechanics*, vol. 60, Graduate Texts in Mathematics (Springer, New York, 1978). ISBN 0-387-90314-3

M. Benacquista, *An Introduction to the Evolution of Single and Binary Stars* (Springer, New York, Heidelberg, Dordrecht, London, 2013)

R.E. Berg, D.G. Stork, *The Physics of Sound*, 3rd edn. (Pearson Prentice Hall, Englewood Cliffs, New Jersey, 2005). ISBN 978-0131457898

J. Bertrand, C.R. Acad. Sci. **77**, 849–853 (1873)

M.L. Boas, *Mathematical Methods in the Physical Sciences*, 3rd edn. (John Wiley & Sons Inc, United States of America, 2006). ISBN 0-471-19826-9

H. Bondi, *Relativity and Common Sense: A New Approach to Einstein* (Dover Publications Inc, New York, 1962). ISBN 0-486-24021-5

P. Dennery, A. Kryzwicki, *Mathematics for Physcists* (Dover Publications Inc, Mineola, New York, 1967)

S. Dutta, S. Ray, Bead on a rotating circular hoop: a simple yet feature-rich dynamical system. *ArXiv e-prints*, (December 2011)

A. Einstein, Zur Elektrodynamik bewegter Körper. Annalen der Physik **322**, 891–921 (1905). https://doi.org/10.1002/andp.19053221004

L.A. Fetter, J.D. Walecka, *Theoretical Mechanics of Particles and Continua* (McGraw-Hill Book Company, United States of America, 1980)

R.P. Feynman, *Surely You're Joking Mr. Feynman! Adventures of a Curious Character* (W.W. Norton & Company, New York, London, 1985). ISBN 0-393-31604-1

R.P. Feynman, R.B. Leighton, Matthew Sands, in *The Feyman Lectures on Physics*, vol. II (Addison-Wesley Publishing Company, Reading, Massachusetts, 1964). ISBN 0-201-02117-X-P




M.J. Benacquista and J.D. Romano, *Classical Mechanics*, Undergraduate Lecture Notes in Physics, https://doi.org/10.1007/978-3-319-68780-3

H. Flanders, *Differential Forms with Applications to the Physical Sciences* (Dover Publications Inc, New York, 1963). ISBN 0-486-66169-5

M.R. Flannery, The enigma of nonholonomic constraints. Am. J. Phys. **73**, 265–272 (2005). https://doi.org/10.1119/1.1830501

I.M. Gelfand, S.V. Fomin, *Calculus of Variations* (Dover Publications Inc, Mineola, New York, 1963). ISBN 0-486-41448-5. (Translated and Edited by Richard A. Silverman)

H. Goldstein, C. Poole, J. Safko, *Classical Mechanics*, 3rd edn. (Addison Wesley, San Francisco, CA, 2002). ISBN 0-201-65702-3

D.J. Griffiths, *Introduction to Electrodynamics*, 3rd edn. (Pearson Prentice Hall, United States of America, 1999). ISBN 0-13-805326-X

D.J. Griffiths, *Introduction to Quantum Mechanics*, 2nd edn. (Pearson Prentice Hall, United States of America, 2005). ISBN 0-13-111892-7

P.R. Halmos, *Finite-Dimensional Vector Spaces*, 2nd edn. (D. Van Nostrand Company Inc, Princeton, New Jersey, 1958)

J.B. Hartle, *Gravity: An Introduction to Einstein's General Relativity* (Benjamin Cummings, illustrate edition, January 2003). ISBN 0805386629

H. Hertz, *The Principles of Mechanics Presented in a New Form* (Dover Publications Inc, New York, 2004). ISBN 978-0486495576 (The original german edition Die Prinzipien der Mechanik in neuem zusammenhange dargestellt was published in 1894)

R.W. Hilditch, *An Introduction to Close Binary Stars* (Cambridge University Press, Cambridge, 2001)

K.V. Kuchǎr Theoretical mechanics. Unpublished lecture notes (1995)

J.B. Kuipers. *Quarternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace, and Virtual Reality* (Princeton University Press, 1999)

C. Lanczos, *The Variational Principles of Mechanics*, 4th edn. (Dover Publications Inc, New York, 1949). ISBN 0-486-65067-7

L.D. Landau, E.M Lifshitz, *Classical Theory of Fields, Course of Theoretical Physics*, 4th edn., vol. 2 (Pergamon Press, Oxford, 1975). ISBN 0-08-025072-6

L.D. Landau, E.M Lifshitz, *Mechanics, Course of Theoretical Physics*, 3rd edn., vol. 1 (Elsevier Ltd, Oxford, 1976). ISBN 978-0-7506-2896-9

J.B. Marion, S.T. Thornton, *Classical Dynamics of Particles and Systems*, 4th edn. (Saunders College Publishing, United States of America, 1995). ISBN 0-03-097302-3

J. Mathews, R.L. Walker, *Mathematical Methods of Physics* (Benjamin/Cummings, United States of America, 1970). ISBN 0-8053-7002-1

N.D. Mermin, *It's About Time: Understanding Einstein's Relativity* (Princeton University Press, Princeton, New Jersey, 2005). ISBN 0-691-12201-6

E. Noether, Invariante Variationsprobleme. *Nachr. D. König. Gesellsch. Wiss. Zu Göttingen*, 1918:235–257, 1918

E. Noether, Invariant variation problems. Trans. Theory Stat. Phys. **1**, 186–207 (1971). https://doi.org/10.1080/00411457108231446

J.D. Romano, R.H. Price, Why no shear in "Div, grad, curl, and all that"? Am. J. Phys. **80**(6), 519–524 (2012). https://doi.org/10.1119/1.3688678

T.D. Rossing, P.A. Wheeler, R.M. Taylor, *The Science of Sound*, 3rd edn. (Addison Wesley, San Francisco, 2002). ISBN 978-0805385656

F.C. Santos, V. Soares, A.C. Tort. An English translation o Bertrand's theorem. *ArXiv e-prints*, (April 2007)

H.M. Schey, *div, grad, curl and all that: An informal text on vector calculus*, 3rd edn. (W.W. Norton & Co., New York, 1996)

B. Schutz, *A First Course in General Relativity* (Cambridge University Press, May 2009). ISBN 9780521887052

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

E.F. Taylor, J.A. Wheeler, *Spacetime Physics: Introduction to special relativity* (W.H. Freeman and Company, New York, 1992)

J. Terrell, Invisibility of the Lorentz contraction. Phys. Rev. **116**, 1041–1045 (1959). https://doi.org/10.1103/PhysRev.116.1041

C.G. Torre, *Introduction to Classical Field Theory*. All Complete Monographs (2016). http://digitalcommons.usu.edu/lib_mono/3/

E.P. Wigner, *Gruppentheorie und ihre Anwendungen auf die Quantenmechanik der Atomspektren* (Vieweg Verlag, Braunschweig, Germany, 1931)

# Index





M.J. Benacquista and J.D. Romano, *Classical Mechanics*, Undergraduate Lecture Notes in Physics, https://doi.org/10.1007/978-3-319-68780-3