or by the operator itself, such that the boundary terms in (17.15) vanish, then the operator is said to be *Hermitian* over the interval $a \le x \le b$. Thus, in this case,

$$\int_a^b f^*(x)\,[\mathcal{L}g(x)]\,dx = \int_a^b [\mathcal{L}f(x)]^* g(x)\,dx. \tag{17.16}$$

A little careful study will reveal the similarity between the definition of an Hermitian operator and the definition of an Hermitian matrix given in chapter 8.

►*Show that the linear operator $\mathcal{L} = d^2/dt^2$ is self-adjoint, and determine the required boundary conditions for the operator to be Hermitian over the interval t_0 to $t_0 + T$.*

Substituting into the LHS of the definition of the adjoint operator (17.15) and integrating by parts gives

$$\int_{t_0}^{t_0+T} f^* \frac{d^2 g}{dt^2}\,dt = \left[f^* \frac{dg}{dt} \right]_{t_0}^{t_0+T} - \int_{t_0}^{t_0+T} \frac{df^*}{dt}\frac{dg}{dt}\,dt.$$

Integrating the second term on the RHS by parts once more yields

$$\int_{t_0}^{t_0+T} f^* \frac{d^2 g}{dt^2}\,dt = \left[f^* \frac{dg}{dt} \right]_{t_0}^{t_0+T} + \left[-\frac{df^*}{dt} g \right]_{t_0}^{t_0+T} + \int_{t_0}^{t_0+T} g \frac{d^2 f^*}{dt^2}\,dt,$$

which, by comparison with (17.15), proves that $\mathcal{L}$ is a self-adjoint operator. Moreover, from (17.16), we see that $\mathcal{L}$ is an Hermitian operator over the required interval provided

$$\left[f^* \frac{dg}{dt} \right]_{t_0}^{t_0+T} = \left[\frac{df^*}{dt} g \right]_{t_0}^{t_0+T}. \blacktriangleleft$$

We showed in chapter 8 that the eigenvalues of Hermitian matrices are real and that their eigenvectors can be chosen to be orthogonal. Similarly, the eigenvalues of Hermitian operators are real and their eigenfunctions can be chosen to be orthogonal (we will prove these properties in the following section). Hermitian operators (or matrices) are often used in the formulation of quantum mechanics. The eigenvalues then give the possible measured values of an observable quantity such as energy or angular momentum, and the physical requirement that such quantities must be real is ensured by the reality of these eigenvalues. Furthermore, the infinite set of eigenfunctions of an Hermitian operator form a complete basis set over the relevant interval, so that it is possible to expand any function $y(x)$ obeying the appropriate conditions in an eigenfunction series over this interval:

$$y(x) = \sum_{n=0}^{\infty} c_n y_n(x), \tag{17.17}$$

where the choice of suitable values for the c_n will make the sum arbitrarily close to $y(x)$.[§] These useful properties provide the motivation for a detailed study of Hermitian operators.

[§] The proof of the completeness of the eigenfunctions of an Hermitian operator is beyond the scope of this book. The reader should refer, for example, to R. Courant and D. Hilbert, *Methods of Mathematical Physics* (New York: Interscience, 1953).

17.3 Properties of Hermitian operators

We now provide proofs of some of the useful properties of Hermitian operators. Again much of the analysis is similar to that for Hermitian matrices in chapter 8, although the present section stands alone. (Here, and throughout the remainder of this chapter, we will write out inner products in full. We note, however, that the inner product notation often provides a neat form in which to express results.)

17.3.1 Reality of the eigenvalues

Consider an Hermitian operator for which (17.5) is satisfied by at least two eigenfunctions $y_i(x)$ and $y_j(x)$, which have corresponding eigenvalues λ_i and λ_j, so that

$$\mathcal{L}y_i = \lambda_i \rho(x) y_i, \tag{17.18}$$

$$\mathcal{L}y_j = \lambda_j \rho(x) y_j, \tag{17.19}$$

where we have allowed for the presence of a weight function $\rho(x)$. Multiplying (17.18) by y_j^* and (17.19) by y_i^* and then integrating gives

$$\int_a^b y_j^* \mathcal{L} y_i \, dx = \lambda_i \int_a^b y_j^* y_i \rho \, dx, \tag{17.20}$$

$$\int_a^b y_i^* \mathcal{L} y_j \, dx = \lambda_j \int_a^b y_i^* y_j \rho \, dx. \tag{17.21}$$

Remembering that we have required $\rho(x)$ to be real, the complex conjugate of (17.20) becomes

$$\int_a^b y_j (\mathcal{L} y_i)^* \, dx = \lambda_i^* \int_a^b y_i^* y_j \rho \, dx, \tag{17.22}$$

and using the definition of an Hermitian operator (17.16) it follows that the LHS of (17.22) is equal to the LHS of (17.21). Thus

$$(\lambda_i^* - \lambda_j) \int_a^b y_i^* y_j \rho \, dx = 0. \tag{17.23}$$

If $i = j$ then $\lambda_i = \lambda_i^*$ (since $\int_a^b y_i^* y_i \rho \, dx \neq 0$), which is a statement that the eigenvalue λ_i is real.

17.3.2 Orthogonality and normalisation of the eigenfunctions

From (17.23), it is immediately apparent that two eigenfunctions y_i and y_j that correspond to different eigenvalues, i.e. such that $\lambda_i \neq \lambda_j$, satisfy

$$\int_a^b y_i^* y_j \rho \, dx = 0, \tag{17.24}$$

which is a statement of the orthogonality of y_i and y_j.

If one (or more) of the eigenvalues is degenerate, however, we have different eigenfunctions corresponding to the same eigenvalue, and the proof of orthogonality is not so straightforward. Nevertheless, an orthogonal set of eigenfunctions may be constructed using the *Gram–Schmidt orthogonalisation* method mentioned earlier in this chapter and used in chapter 8 to construct a set of orthogonal eigenvectors of an Hermitian matrix. We repeat the analysis here for completeness.

Suppose, for the sake of our proof, that λ_0 is k-fold degenerate, i.e.

$$\mathcal{L}y_i = \lambda_0 \rho y_i \qquad \text{for } i = 0, 1, \ldots, k-1, \tag{17.25}$$

but that λ_0 is different from any of λ_k, λ_{k+1}, etc. Then any linear combination of these y_i is also an eigenfunction with eigenvalue λ_0 since

$$\mathcal{L}z \equiv \mathcal{L}\sum_{i=0}^{k-1} c_i y_i = \sum_{i=0}^{k-1} c_i \mathcal{L}y_i = \sum_{i=0}^{k-1} c_i \lambda_0 \rho y_i = \lambda_0 \rho z. \tag{17.26}$$

If the y_i defined in (17.25) are not already mutually orthogonal then consider the new eigenfunctions z_i constructed by the following procedure, in which each of the new functions z_i is to be normalised, to give $\hat{z}_i$, before proceeding to the construction of the next one (the normalisation can be carried out by dividing the eigenfunction z_i by $(\int_a^b z_i^* z_i \rho \, dx)^{1/2}$):

$$\begin{aligned}
z_0 &= y_0, \\
z_1 &= y_1 - \left(\hat{z}_0 \int_a^b \hat{z}_0^* y_1 \rho \, dx\right), \\
z_2 &= y_2 - \left(\hat{z}_1 \int_a^b \hat{z}_1^* y_2 \rho \, dx\right) - \left(\hat{z}_0 \int_a^b \hat{z}_0^* y_2 \rho \, dx\right), \\
&\ \vdots \\
z_{k-1} &= y_{k-1} - \left(\hat{z}_{k-2} \int_a^b \hat{z}_{k-2}^* y_{k-1} \rho \, dx\right) - \cdots - \left(\hat{z}_0 \int_a^b \hat{z}_0^* y_{k-1} \rho \, dx\right).
\end{aligned}$$

Each of the integrals is just a number and thus each new function z_i is, as can be shown from (17.26), an eigenvector of $\mathcal{L}$ with eigenvalue λ_0. It is straightforward to check that each z_i is orthogonal to all its predecessors. Thus, by this explicit construction we have shown that an orthogonal set of eigenfunctions of an Hermitian operator $\mathcal{L}$ can be obtained. Clearly the orthogonal set obtained, z_i, is not unique.

In general, since $\mathcal{L}$ is linear, the normalisation of its eigenfunctions $y_i(x)$ is arbitrary. It is often convenient, however, to work in terms of the normalised eigenfunctions $\hat{y}_i(x)$, so that $\int_a^b \hat{y}_i^* \hat{y}_i \rho \, dx = 1$. These therefore form an orthonormal set and we can write

$$\int_a^b \hat{y}_i^* \hat{y}_j \rho \, dx = \delta_{ij}, \tag{17.27}$$

which is valid for all pairs of values i, j.

17.3.3 Completeness of the eigenfunctions

As noted earlier, the eigenfunctions of an Hermitian operator may be shown to form a complete basis set over the relevant interval. One may thus expand any (reasonable) function $y(x)$ obeying appropriate boundary conditions in an eigenfunction series over the interval, as in (17.17). Working in terms of the normalised eigenfunctions $\hat{y}_n(x)$, we may thus write

$$\begin{aligned} f(x) &= \sum_n \hat{y}_n(x) \int_a^b \hat{y}_n^*(z) f(z) \rho(z) \, dz \\ &= \int_a^b f(z) \rho(z) \sum_n \hat{y}_n(x) \hat{y}_n^*(z) \, dz. \end{aligned}$$

Since this is true for any $f(x)$, we must have that

$$\rho(z) \sum_n \hat{y}_n(x) \hat{y}_n^*(z) = \delta(x - z). \tag{17.28}$$

This is called the *completeness* or *closure* property of the eigenfunctions. It defines a complete set. If the spectrum of eigenvalues of $\mathcal{L}$ is anywhere continuous then the eigenfunction $y_n(x)$ must be treated as $y(n, x)$ and an integration carried out over n.

We also note that the RHS of (17.28) is a δ-function and so is only non-zero when $z = x$; thus $\rho(z)$ on the LHS can be replaced by $\rho(x)$ if required, i.e.

$$\rho(z) \sum_n \hat{y}_n(x) \hat{y}_n^*(z) = \rho(x) \sum_n \hat{y}_n(x) \hat{y}_n^*(z). \tag{17.29}$$

17.3.4 Construction of real eigenfunctions

Recall that the eigenfunction y_i satisfies

$$\mathcal{L} y_i = \lambda_i \rho y_i \tag{17.30}$$

and that the complex conjugate of this gives

$$\mathcal{L} y_i^* = \lambda_i^* \rho y_i^* = \lambda_i \rho y_i^*, \tag{17.31}$$

where the last equality follows because the eigenvalues are real, i.e. $\lambda_i = \lambda_i^*$. Thus, y_i and y_i^* are eigenfunctions corresponding to the same eigenvalue and hence, because of the linearity of $\mathcal{L}$, at least one of $y_i^* + y_i$ and $i(y_i^* - y_i)$, which are both real, is a non-zero eigenfunction corresponding to that eigenvalue. It follows that the eigenfunctions can always be made real by taking suitable linear combinations, though taking such linear combinations will only be necessary in cases where a particular λ is degenerate, i.e. corresponds to more than one linearly independent eigenfunction.

17.4 Sturm–Liouville equations

One of the most important applications of our discussion of Hermitian operators is to the study of *Sturm–Liouville equations*, which take the general form

$$p(x)\frac{d^2y}{dx^2} + r(x)\frac{dy}{dx} + q(x)y + \lambda\rho(x)y = 0, \qquad \text{where } r(x) = \frac{dp(x)}{dx} \tag{17.32}$$

and p, q and r are real functions of x.[§] A variational approach to the Sturm–Liouville equation, which is useful in estimating the eigenvalues λ for a given set of boundary conditions on y, is discussed in chapter 22. For now, however, we concentrate on demonstrating that solutions of the Sturm–Liouville equation that satisfy appropriate boundary conditions are the eigenfunctions of an Hermitian operator.

It is clear that (17.32) can be written

$$\mathcal{L}y = \lambda\rho(x)y, \qquad \text{where } \mathcal{L} \equiv -\left[p(x)\frac{d^2}{dx^2} + r(x)\frac{d}{dx} + q(x)\right]. \tag{17.33}$$

Using the condition that $r(x) = p'(x)$, it will be seen that the general Sturm–Liouville equation (17.32) can also be rewritten as

$$(py')' + qy + \lambda\rho y = 0, \tag{17.34}$$

where primes denote differentiation with respect to x. Using (17.33) this may also be written $\mathcal{L}y \equiv -(py')' - qy = \lambda\rho y$, which defines a more useful form for the Sturm–Liouville linear operator, namely

$$\mathcal{L} \equiv -\left[\frac{d}{dx}\left(p(x)\frac{d}{dx}\right) + q(x)\right]. \tag{17.35}$$

17.4.1 Hermitian nature of the Sturm–Liouville operator

As we now show, the linear operator of the Sturm–Liouville equation (17.35) is self-adjoint. Moreover, the operator is Hermitian over the range $[a, b]$ provided

[§] We note that sign conventions vary in this expression for the general Sturm–Liouville equation; some authors use $-\lambda\rho(x)y$ on the LHS of (17.32).

certain boundary conditions are met, namely that any two eigenfunctions y_i and y_j of (17.33) must satisfy

$$\left[y_i^* p y_j'\right]_{x=a} = \left[y_i^* p y_j'\right]_{x=b} \qquad \text{for all } i, j. \tag{17.36}$$

Rearranging (17.36), we can write

$$\left[y_i^* p y_j'\right]_{x=a}^{x=b} = 0 \tag{17.37}$$

as an equivalent statement of the required boundary conditions. These boundary conditions are in fact not too restrictive and are met, for instance, by the sets $y(a) = y(b) = 0$; $y(a) = y'(b) = 0$; $p(a) = p(b) = 0$ and by many other sets. It is important to note that in order to satisfy (17.36) and (17.37) one boundary condition must be specified at each end of the range.

►*Prove that the Sturm–Liouville operator is Hermitian over the range $[a, b]$ and under the boundary conditions (17.37).*

Putting the Sturm–Liouville form $\mathcal{L}y = -(py')' - qy$ into the definition (17.16) of an Hermitian operator, the LHS may be written as a sum of two terms, i.e.

$$-\int_a^b \left[y_i^*(py_j')' + y_i^* q y_j\right]\, dx = -\int_a^b y_i^*(py_j')'\, dx - \int_a^b y_i^* q y_j\, dx.$$

The first term may be integrated by parts to give

$$-\left[y_i^* p y_j'\right]_a^b + \int_a^b (y_i^*)' p y_j'\, dx.$$

The boundary-value term in this is zero because of the boundary conditions, and so integrating by parts again yields

$$\left[(y_i^*)' p y_j\right]_a^b - \int_a^b ((y_i^*)' p)' y_j\, dx.$$

Again, the boundary-value term is zero, leaving us with

$$-\int_a^b \left[y_i^*(py_j')' + y_i^* q y_j\right]\, dx = -\int_a^b \left[y_j(p(y_i^*)')' + y_j q y_i^*\right]\, dx,$$

which proves that the Sturm–Liouville operator is Hermitian over the prescribed interval. ◄

It is also worth noting that, since $p(a) = p(b) = 0$ is a valid set of boundary conditions, many Sturm–Liouville equations possess a ‘natural’ interval $[a, b]$ over which the corresponding differential operator $\mathcal{L}$ is Hermitian *irrespective* of the boundary conditions satisfied by its eigenfunctions at $x = a$ and $x = b$ (the only requirement being that they are regular at these end-points).

17.4.2 Transforming an equation into Sturm–Liouville form

Many of the second-order differential equations encountered in physical problems are examples of the Sturm–Liouville equation (17.34). Moreover, *any* second-order

Equation	$p(x)$	$q(x)$	λ	$\rho(x)$
Hypergeometric	$x^c(1-x)^{a+b-c+1}$	0	$-ab$	$x^{c-1}(1-x)^{a+b-c}$
Legendre	$1-x^2$	0	$\ell(\ell+1)$	1
Associated Legendre	$1-x^2$	$-m^2/(1-x^2)$	$\ell(\ell+1)$	1
Chebyshev	$(1-x^2)^{1/2}$	0	ν^2	$(1-x^2)^{-1/2}$
Confluent hypergeometric	$x^c e^{-x}$	0	$-a$	$x^{c-1}e^{-x}$
Bessel*	x	$-\nu^2/x$	α^2	x
Laguerre	xe^{-x}	0	ν	e^{-x}
Associated Laguerre	$x^{m+1}e^{-x}$	0	ν	$x^m e^{-x}$
Hermite	e^{-x^2}	0	2ν	e^{-x^2}
Simple harmonic	1	0	ω^2	1

Table 17.1 The Sturm–Liouville form (17.34) for important ODEs in the physical sciences and engineering. The asterisk denotes that, for Bessel's equation, a change of variable $x \to x/a$ is required to give the conventional normalisation used here, but is not needed for the transformation into Sturm–Liouville form.

differential equation of the form

$$p(x)y'' + r(x)y' + q(x)y + \lambda\rho(x)y = 0 \tag{17.38}$$

can be converted into Sturm–Liouville form by multiplying through by a suitable integrating factor, which is given by

$$F(x) = \exp\left\{\int^x \frac{r(u) - p'(u)}{p(u)}\,du\right\}. \tag{17.39}$$

It is easily verified that (17.38) then takes the Sturm–Liouville form,

$$[F(x)p(x)y']' + F(x)q(x)y + \lambda F(x)\rho(x)y = 0, \tag{17.40}$$

with a different, but still non-negative, weight function $F(x)\rho(x)$. Table 17.1 summarises the Sturm–Liouville form (17.34) for several of the equations listed in table 16.1. These forms can be determined using (17.39), as illustrated in the following example.

►*Put the following equations into Sturm–Liouville (SL) form:*

(i) $(1-x^2)y'' - xy' + \nu^2 y = 0$ *(Chebyshev equation)*;

(ii) $xy'' + (1-x)y' + \nu y = 0$ *(Laguerre equation)*;

(iii) $y'' - 2xy' + 2\nu y = 0$ *(Hermite equation)*.

(i) From (17.39), the required integrating factor is

$$F(x) = \exp\left(\int^x \frac{u}{1-u^2}\,du\right) = \exp\left[-\tfrac{1}{2}\ln(1-x^2)\right] = (1-x^2)^{-1/2}.$$

Thus, the Chebyshev equation becomes

$$(1-x^2)^{1/2}y'' - x(1-x^2)^{-1/2}y' + \nu^2(1-x^2)^{-1/2}y = \left[(1-x^2)^{1/2}y'\right]' + \nu^2(1-x^2)^{-1/2}y = 0,$$

which is in SL form with $p(x) = (1-x^2)^{1/2}$, $q(x) = 0$, $\rho(x) = (1-x^2)^{-1/2}$ and $\lambda = \nu^2$.

(ii) From (17.39), the required integrating factor is

$$F(x) = \exp\left(\int^x -1\,du\right) = \exp(-x).$$

Thus, the Laguerre equation becomes

$$xe^{-x}y'' + (1-x)e^{-x}y' + \nu e^{-x}y = (xe^{-x}y')' + \nu e^{-x}y = 0,$$

which is in SL form with $p(x) = xe^{-x}$, $q(x) = 0$, $\rho(x) = e^{-x}$ and $\lambda = \nu$.

(iii) From (17.39), the required integrating factor is

$$F(x) = \exp\left(\int^x -2u\,du\right) = \exp(-x^2).$$

Thus, the Hermite equation becomes

$$e^{-x^2}y'' - 2xe^{-x^2}y' + 2\nu e^{-x^2}y = (e^{-x^2}y')' + 2\nu e^{-x^2}y = 0,$$

which is in SL form with $p(x) = e^{-x^2}$, $q(x) = 0$, $\rho(x) = e^{-x^2}$ and $\lambda = 2\nu$. ◄

From the $p(x)$ entries in table 17.1, we may read off the natural interval over which the corresponding Sturm–Liouville operator (17.35) is Hermitian; in each case this is given by $[a, b]$, where $p(a) = p(b) = 0$. Thus, the natural interval for the Legendre equation, the associated Legendre equation and the Chebyshev equation is $[-1, 1]$; for the Laguerre and associated Laguerre equations the interval is $[0, \infty]$; and for the Hermite equation it is $[-\infty, \infty]$. In addition, from (17.37), one sees that for the simple harmonic equation one requires only that $[a, b] = [x_0, x_0 + 2\pi]$. We also note that, as required, the weight function in each case is finite and non-negative over the natural interval. Occasionally, a little more care is required when determining the conditions for a Sturm–Liouville operator of the form (17.35) to be Hermitian over some natural interval, as is illustrated in the following example.

►*Express the hypergeometric equation,*

$$x(1-x)y'' + [\,c - (a+b+1)x\,]y' - aby = 0,$$

in Sturm–Liouville form. Hence determine the natural interval over which the resulting Sturm–Liouville operator is Hermitian and the corresponding conditions that one must impose on the parameters a, b and c.

As usual for an equation not already in SL form, we first determine the appropriate integrating factor. This is given, as in equation (17.39), by

$$\begin{aligned}
F(x) &= \exp\left[\int^x \frac{c-(a+b+1)u-1+2u}{u(1-u)}\,du\right] \\
&= \exp\left[\int^x \frac{c-1-(a+b-1)u}{u(1-u)}\,du\right] \\
&= \exp\left[\int^x \left(\frac{c-1}{1-u}+\frac{c-1}{u}-\frac{a+b-1}{1-u}\,du\right)\right] \\
&= \exp\left[(a+b-c)\ln(1-x)+(c-1)\ln x\right] \\
&= x^{c-1}(1-x)^{a+b-c}.
\end{aligned}$$

When the equation is multiplied through by $F(x)$ it takes the form

$$\left[x^c(1-x)^{a+b-c+1}y'\right]' - abx^{c-1}(1-x)^{a+b-c}y = 0.$$

Now, for the corresponding Sturm–Liouville operator to be Hermitian, the conditions to be imposed are as follows.

(i) The boundary condition (17.37); if $c > 0$ and $a+b-c+1 > 0$, this is satisfied automatically for $0 \le x \le 1$, which is thus the natural interval in this case.
(ii) The weight function $x^{c-1}(1-x)^{a+b-c}$ must be finite and not change sign in the interval $0 \le x \le 1$. This means that both exponents in it must be positive, i.e. $c-1 > 0$ and $a+b-c > 0$.

Putting together the conditions on the parameters gives the double inequality $a+b > c > 1$. ◄

Finally, we consider Bessel's equation,

$$x^2y'' + xy' + (x^2 - \nu^2)y = 0,$$

which may be converted into Sturm–Liouville form, but only in a somewhat unorthodox fashion. It is conventional first to divide the Bessel equation by x and then to change variables to $\bar{x} = x/\alpha$. In this case, it becomes

$$\bar{x}y''(\alpha\bar{x}) + y'(\alpha\bar{x}) - \frac{\nu^2}{\bar{x}}y(\alpha\bar{x}) + \alpha^2\bar{x}y(\alpha\bar{x}) = 0, \qquad (17.41)$$

where a prime now indicates differentiation with respect to $\bar{x}$. Dropping the bars on the independent variable, we thus have

$$[xy'(\alpha x)]' - \frac{\nu^2}{x}y(\alpha x) + \alpha^2 xy(\alpha x) = 0, \qquad (17.42)$$

which is in SL form with $p(x) = x$, $q(x) = -\nu^2/x$, $\rho(x) = x$ and $\lambda = \alpha^2$. It should be noted, however, that in this case the eigenvalue (actually its square root) appears in the argument of the dependent variable.

17.5 Superposition of eigenfunctions: Green's functions

We have already seen that if

$$\mathcal{L}y_n(x) = \lambda_n \rho(x) y_n(x), \tag{17.43}$$

where $\mathcal{L}$ is an Hermitian operator, then the eigenvalues λ_n are real and the eigenfunctions $y_n(x)$ are orthogonal (or can be made so). Let us assume that we know the eigenfunctions $y_n(x)$ of $\mathcal{L}$ that individually satisfy (17.43) and some imposed boundary conditions (for which $\mathcal{L}$ is Hermitian).

Now let us suppose we wish to solve the inhomogeneous differential equation

$$\mathcal{L}y(x) = f(x), \tag{17.44}$$

subject to the same boundary conditions. Since the eigenfunctions of $\mathcal{L}$ form a complete set, the full solution, $y(x)$, to (17.44) may be written as a superposition of eigenfunctions, i.e.

$$y(x) = \sum_{n=0}^{\infty} c_n y_n(x), \tag{17.45}$$

for some choice of the constants c_n. Making full use of the linearity of $\mathcal{L}$, we have

$$f(x) = \mathcal{L}y(x) = \mathcal{L}\left(\sum_{n=0}^{\infty} c_n y_n(x)\right) = \sum_{n=0}^{\infty} c_n \mathcal{L}y_n(x) = \sum_{n=0}^{\infty} c_n \lambda_n \rho(x) y_n(x). \tag{17.46}$$

Multiplying the first and last terms of (17.46) by y_j^* and integrating, we obtain

$$\int_a^b y_j^*(z) f(z)\,dz = \sum_{n=0}^{\infty} \int_a^b c_n \lambda_n y_j^*(z) y_n(z) \rho(z)\,dz, \tag{17.47}$$

where we have used z as the integration variable for later convenience. Finally, using the orthogonality condition (17.27), we see that the integrals on the RHS are zero unless $n = j$, and so obtain

$$c_n = \frac{1}{\lambda_n} \frac{\int_a^b y_n^*(z) f(z)\,dz}{\int_a^b y_n^*(z) y_n(z) \rho(z)\,dz}. \tag{17.48}$$

Thus, if we can find all the eigenfunctions of a differential operator then (17.48) can be used to find the weighting coefficients for the superposition, to give as the full solution

$$y(x) = \sum_{n=0}^{\infty} \frac{1}{\lambda_n} \frac{\int_a^b y_n^*(z) f(z)\,dz}{\int_a^b y_n^*(z) y_n(z) \rho(z)\,dz}\, y_n(x). \tag{17.49}$$

If we work with normalised eigenfunctions $\hat{y}_n(x)$, so that

$$\int_a^b \hat{y}_n^*(z) \hat{y}_n(z) \rho(z)\,dz = 1 \qquad \text{for all } n,$$

and we assume that we may interchange the order of summation and integration, then (17.49) can be written as

$$y(x) = \int_a^b \left\{ \sum_{n=0}^{\infty} \left[\frac{1}{\lambda_n} \hat{y}_n(x) \hat{y}_n^*(z) \right] \right\} f(z)\, dz.$$

The quantity in braces, which is a function of x and z only, is usually written $G(x,z)$, and is the *Green's function* for the problem. With this notation,

$$y(x) = \int_a^b G(x,z) f(z)\, dz, \tag{17.50}$$

where

$$G(x,z) = \sum_{n=0}^{\infty} \frac{1}{\lambda_n} \hat{y}_n(x) \hat{y}_n^*(z). \tag{17.51}$$

We note that $G(x,z)$ is determined entirely by the boundary conditions and the eigenfunctions $\hat{y}_n$, and hence by $\mathcal{L}$ itself, and that $f(z)$ depends purely on the RHS of the inhomogeneous equation (17.44). Thus, for a given $\mathcal{L}$ and boundary conditions we can establish, once and for all, a function $G(x,z)$ that will enable us to solve the inhomogeneous equation for *any* RHS. From (17.51) we also note that

$$G(x,z) = G^*(z,x). \tag{17.52}$$

We have already met the Green's function in the solution of second-order differential equations in chapter 15, as the function that satisfies the equation $\mathcal{L}[G(x,z)] = \delta(x-z)$ (and the boundary conditions). The formulation given above is an alternative, though equivalent, one.

►*Find an appropriate Green's function for the equation*

$$y'' + \tfrac{1}{4} y = f(x),$$

with boundary conditions $y(0) = y(\pi) = 0$. *Hence, solve for* (i) $f(x) = \sin 2x$ *and* (ii) $f(x) = x/2$.

One approach to solving this problem is to use the methods of chapter 15 and find a complementary function and particular integral. However, in order to illustrate the techniques developed in the present chapter we will use the superposition of eigenfunctions, which, as may easily be checked, produces the same solution.

The operator on the LHS of this equation is already Hermitian under the given boundary conditions, and so we seek its eigenfunctions. These satisfy the equation

$$y'' + \tfrac{1}{4} y = \lambda y.$$

This equation has the familiar solution

$$y(x) = A \sin\left(\sqrt{\tfrac{1}{4} - \lambda}\right) x + B \cos\left(\sqrt{\tfrac{1}{4} - \lambda}\right) x.$$

Now, the boundary conditions require that $B = 0$ and $\sin\left(\sqrt{\frac{1}{4}-\lambda}\right)\pi = 0$, and so

$$\sqrt{\tfrac{1}{4}-\lambda} = n, \qquad \text{where } n = 0, \pm 1, \pm 2, \ldots.$$

Therefore, the independent eigenfunctions that satisfy the boundary conditions are

$$y_n(x) = A_n \sin nx,$$

where n is any non-negative integer, and the corresponding eigenvalues are $\lambda_n = \frac{1}{4} - n^2$. The normalisation condition further requires

$$\int_0^\pi A_n^2 \sin^2 nx \, dx = 1 \qquad \Rightarrow \qquad A_n = \left(\frac{2}{\pi}\right)^{1/2}.$$

Comparison with (17.51) shows that the appropriate Green's function is therefore given by

$$G(x,z) = \frac{2}{\pi}\sum_{n=0}^{\infty} \frac{\sin nx \sin nz}{\frac{1}{4}-n^2}.$$

Case (i). Using (17.50), the solution with $f(x) = \sin 2x$ is given by

$$y(x) = \frac{2}{\pi}\int_0^\pi \left(\sum_{n=0}^{\infty} \frac{\sin nx \sin nz}{\frac{1}{4}-n^2}\right) \sin 2z \, dz = \frac{2}{\pi}\sum_{n=0}^{\infty} \frac{\sin nx}{\frac{1}{4}-n^2}\int_0^\pi \sin nz \sin 2z \, dz.$$

Now the integral is zero unless $n = 2$, in which case it is

$$\int_0^\pi \sin^2 2z \, dz = \frac{\pi}{2}.$$

Thus

$$y(x) = -\frac{2}{\pi}\frac{\sin 2x}{15/4}\frac{\pi}{2} = -\frac{4}{15}\sin 2x$$

is the full solution for $f(x) = \sin 2x$. This is, of course, exactly the solution found by using the methods of chapter 15.

Case (ii). The solution with $f(x) = x/2$ is given by

$$y(x) = \int_0^\pi \left(\frac{2}{\pi}\sum_{n=0}^{\infty} \frac{\sin nx \sin nz}{\frac{1}{4}-n^2}\right) \frac{z}{2}\, dz = \frac{1}{\pi}\sum_{n=0}^{\infty} \frac{\sin nx}{\frac{1}{4}-n^2}\int_0^\pi z \sin nz \, dz.$$

The integral may be evaluated by integrating by parts. For $n \neq 0$,

$$\begin{aligned}\int_0^\pi z \sin nz \, dz &= \left[-\frac{z\cos nz}{n}\right]_0^\pi + \int_0^\pi \frac{\cos nz}{n}\, dz \\ &= \frac{-\pi\cos n\pi}{n} + \left[\frac{\sin nz}{n^2}\right]_0^\pi \\ &= -\frac{\pi(-1)^n}{n}.\end{aligned}$$

For $n = 0$ the integral is zero, and thus

$$y(x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\sin nx}{n\left(\frac{1}{4}-n^2\right)},$$

is the full solution for $f(x) = x/2$. Using the methods of subsection 15.1.2, the solution is found to be $y(x) = 2x - 2\pi\sin(x/2)$, which may be shown to be equal to the above solution by expanding $2x - 2\pi\sin(x/2)$ as a Fourier sine series. ◄

17.6 A useful generalisation

Sometimes we encounter inhomogeneous equations of a form slightly more general than (17.1), given by

$$\mathcal{L}y(x) - \mu\rho(x)y(x) = f(x) \tag{17.53}$$

for some Hermitian operator $\mathcal{L}$, with y subject to the appropriate boundary conditions and μ a given (i.e. *fixed*) constant. To solve this equation we expand $y(x)$ and $f(x)$ in terms of the eigenfunctions $y_n(x)$ of the operator $\mathcal{L}$, which satisfy

$$\mathcal{L}y_n(x) = \lambda_n\rho(x)y_n(x).$$

Working in terms of the normalised eigenfunctions $\hat{y}_n(x)$, we first expand $f(x)$ as follows:

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} \hat{y}_n(x) \int_a^b \hat{y}_n^*(z)f(z)\rho(z)\,dz \\ &= \int_a^b \rho(z) \sum_{n=0}^{\infty} \hat{y}_n(x)\hat{y}_n^*(z)f(z)\,dz. \end{aligned} \tag{17.54}$$

Using (17.29) this becomes

$$\begin{aligned} f(x) &= \int_a^b \rho(x) \sum_{n=0}^{\infty} \hat{y}_n(x)\hat{y}_n^*(z)f(z)\,dz \\ &= \rho(x) \sum_{n=0}^{\infty} \hat{y}_n(x) \int_a^b \hat{y}_n^*(z)f(z)\,dz. \end{aligned} \tag{17.55}$$

Next, we expand $y(x)$ as $y = \sum_{n=0}^{\infty} c_n\hat{y}_n(x)$ and seek the coefficients c_n. Substituting this and (17.55) into (17.53) we have

$$\rho(x)\sum_{n=0}^{\infty}(\lambda_n - \mu)c_n\hat{y}_n(x) = \rho(x)\sum_{n=0}^{\infty}\hat{y}_n(x)\int_a^b \hat{y}_n^*(z)f(z)\,dz,$$

from which we find that

$$c_n = \sum_{n=0}^{\infty} \frac{\int_a^b \hat{y}_n^*(z)f(z)\,dz}{\lambda_n - \mu}.$$

Hence the solution of (17.53) is given by

$$y = \sum_{n=0}^{\infty} c_n\hat{y}_n(x) = \sum_{n=0}^{\infty} \frac{\hat{y}_n(x)}{\lambda_n - \mu}\int_a^b \hat{y}_n^*(z)f(z)\,dz = \int_a^b \sum_{n=0}^{\infty} \frac{\hat{y}_n(x)\hat{y}_n^*(z)}{\lambda_n - \mu}f(z)\,dz.$$

From this we may identify the Green's function

$$G(x,z) = \sum_{n=0}^{\infty} \frac{\hat{y}_n(x)\hat{y}_n^*(z)}{\lambda_n - \mu}.$$

We note that if $\mu = \lambda_n$, i.e. if μ equals one of the eigenvalues of $\mathcal{L}$, then $G(x,z)$ becomes infinite and this method runs into difficulty. No solution then exists unless the RHS of (17.53) satisfies the relation

$$\int_a^b \hat{y}_n^*(x) f(x)\, dx = 0.$$

If the spectrum of eigenvalues of the operator $\mathcal{L}$ is anywhere continuous, the orthonormality and closure relationships of the normalised eigenfunctions become

$$\int_a^b \hat{y}_n^*(x)\hat{y}_m(x)\rho(x)\, dx = \delta(n-m),$$
$$\int_0^\infty \hat{y}_n^*(z)\hat{y}_n(x)\rho(x)\, dn = \delta(x-z).$$

Repeating the above analysis we then find that the Green's function is given by

$$G(x,z) = \int_0^\infty \frac{\hat{y}_n(x)\hat{y}_n^*(z)}{\lambda_n - \mu}\, dn.$$

17.7 Exercises

17.1 By considering $\langle h|h\rangle$, where $h = f + \lambda g$ with λ real, prove that, for two functions f and g,

$$\langle f|f\rangle\langle g|g\rangle \geq \tfrac{1}{4}[\langle f|g\rangle + \langle g|f\rangle]^2.$$

The function $y(x)$ is real and positive for all x. Its Fourier cosine transform $\tilde{y}_c(k)$ is defined by

$$\tilde{y}_c(k) = \int_{-\infty}^{\infty} y(x)\cos(kx)\, dx,$$

and it is given that $\tilde{y}_c(0) = 1$. Prove that

$$\tilde{y}_c(2k) \geq 2[\tilde{y}_c(k)]^2 - 1.$$

17.2 Write the homogeneous Sturm-Liouville eigenvalue equation for which $y(a) = y(b) = 0$ as

$$\mathscr{L}(y;\lambda) \equiv (py')' + qy + \lambda\rho y = 0,$$

where $p(x), q(x)$ and $\rho(x)$ are continuously differentiable functions. Show that if $z(x)$ and $F(x)$ satisfy $\mathcal{L}(z;\lambda) = F(x)$, with $z(a) = z(b) = 0$, then

$$\int_a^b y(x)F(x)\, dx = 0.$$

Demonstrate the validity of this general result by direct calculation for the specific case in which $p(x) = \rho(x) = 1$, $q(x) = 0$, $a = -1$, $b = 1$ and $z(x) = 1 - x^2$.

17.3 Consider the real eigenfunctions $y_n(x)$ of a Sturm–Liouville equation,

$$(py')' + qy + \lambda\rho y = 0, \qquad a \leq x \leq b,$$

in which $p(x)$, $q(x)$ and $\rho(x)$ are continuously differentiable real functions and $p(x)$ does not change sign in $a \leq x \leq b$. Take $p(x)$ as positive throughout the interval, if necessary by changing the signs of all eigenvalues. For $a \leq x_1 \leq x_2 \leq b$, establish the identity

$$(\lambda_n - \lambda_m) \int_{x_1}^{x_2} \rho y_n y_m \, dx = \left[y_n \, p \, y_m' - y_m \, p \, y_n' \right]_{x_1}^{x_2}.$$

Deduce that if $\lambda_n > \lambda_m$ then $y_n(x)$ must change sign between two successive zeros of $y_m(x)$.
[The reader may find it helpful to illustrate this result by sketching the first few eigenfunctions of the system $y'' + \lambda y = 0$, with $y(0) = y(\pi) = 0$, and the Legendre polynomials $P_n(z)$ for $n = 2, 3, 4, 5$.]

17.4 Show that the equation

$$y'' + a\delta(x)y + \lambda y = 0,$$

with $y(\pm\pi) = 0$ and a real, has a set of eigenvalues λ satisfying

$$\tan(\pi\sqrt{\lambda}) = \frac{2\sqrt{\lambda}}{a}.$$

Investigate the conditions under which negative eigenvalues, $\lambda = -\mu^2$, with μ real, are possible.

17.5 Use the properties of Legendre polynomials to carry out the following exercises.

(a) Find the solution of $(1 - x^2)y'' - 2xy' + by = f(x)$, valid in the range $-1 \leq x \leq 1$ and finite at $x = 0$, in terms of Legendre polynomials.
(b) If $b = 14$ and $f(x) = 5x^3$, find the explicit solution and verify it by direct substitution.

[The first six Legendre polynomials are listed in Subsection 18.1.1.]

17.6 Starting from the linearly independent functions $1, x, x^2, x^3, \ldots,$ in the range $0 \leq x < \infty$, find the first three orthogonal functions ϕ_0, ϕ_1 and ϕ_2, with respect to the weight function $\rho(x) = e^{-x}$. By comparing your answers with the Laguerre polynomials generated by the recurrence relation (18.115), deduce the form of $\phi_3(x)$.

17.7 Consider the set of functions, $\{f(x)\}$, of the real variable x, defined in the interval $-\infty < x < \infty$, that $\to 0$ at least as quickly as x^{-1} as $x \to \pm\infty$. For unit weight function, determine whether each of the following linear operators is Hermitian when acting upon $\{f(x)\}$:

$$\text{(a)}\ \frac{d}{dx} + x; \quad \text{(b)}\ -i\frac{d}{dx} + x^2; \quad \text{(c)}\ ix\frac{d}{dx}; \quad \text{(d)}\ i\frac{d^3}{dx^3}.$$

17.8 A particle moves in a parabolic potential in which its natural angular frequency of oscillation is $\frac{1}{2}$. At time $t = 0$ it passes through the origin with velocity v. It is then suddenly subjected to an additional acceleration, of $+1$ for $0 \leq t \leq \pi/2$, followed by -1 for $\pi/2 < t \leq \pi$. At the end of this period it is again at the origin. Apply the results of the worked example in section 17.5 to show that

$$v = -\frac{8}{\pi} \sum_{m=0}^{\infty} \frac{1}{(4m+2)^2 - \frac{1}{4}} \approx -0.81.$$

17.9 Find an eigenfunction expansion for the solution, with boundary conditions $y(0) = y(\pi) = 0$, of the inhomogeneous equation

$$\frac{d^2y}{dx^2} + \kappa y = f(x),$$

where κ is a constant and

$$f(x) = \begin{cases} x & 0 \le x \le \pi/2, \\ \pi - x & \pi/2 < x \le \pi. \end{cases}$$

17.10 Consider the following two approaches to constructing a Green's function.

(a) Find those eigenfunctions $y_n(x)$ of the self-adjoint linear differential operator d^2/dx^2 that satisfy the boundary conditions $y_n(0) = y_n(\pi) = 0$, and hence construct its Green's function $G(x, z)$.
(b) Construct the same Green's function using a method based on the complementary function of the appropriate differential equation and the boundary conditions to be satisfied at the position of the δ-function, showing that it is

$$G(x, z) = \begin{cases} x(z - \pi)/\pi & 0 \le x \le z, \\ z(x - \pi)/\pi & z \le x \le \pi. \end{cases}$$

(c) By expanding the function given in (b) in terms of the eigenfunctions $y_n(x)$, verify that it is the same function as that derived in (a).

17.11 The differential operator $\mathcal{L}$ is defined by

$$\mathcal{L}y = -\frac{d}{dx}\left(e^x \frac{dy}{dx}\right) - \tfrac{1}{4}e^x y.$$

Determine the eigenvalues λ_n of the problem

$$\mathcal{L}y_n = \lambda_n e^x y_n \qquad 0 < x < 1,$$

with boundary conditions

$$y(0) = 0, \qquad \frac{dy}{dx} + \tfrac{1}{2}y = 0 \quad \text{at} \quad x = 1.$$

(a) Find the corresponding unnormalised y_n, and also a weight function $\rho(x)$ with respect to which the y_n are orthogonal. Hence, select a suitable normalisation for the y_n.
(b) By making an eigenfunction expansion, solve the equation

$$\mathcal{L}y = -e^{x/2}, \qquad 0 < x < 1,$$

subject to the same boundary conditions as previously.

17.12 Show that the linear operator

$$\mathcal{L} \equiv \tfrac{1}{4}(1 + x^2)^2 \frac{d^2}{dx^2} + \tfrac{1}{2}x(1 + x^2)\frac{d}{dx} + a,$$

acting upon functions defined in $-1 \le x \le 1$ and vanishing at the end-points of the interval, is Hermitian with respect to the weight function $(1 + x^2)^{-1}$.

By making the change of variable $x = \tan(\theta/2)$, find two even eigenfunctions, $f_1(x)$ and $f_2(x)$, of the differential equation

$$\mathcal{L}u = \lambda u.$$

17.13 By substituting $x = \exp t$, find the normalised eigenfunctions $y_n(x)$ and the eigenvalues λ_n of the operator $\mathcal{L}$ defined by

$$\mathcal{L}y = x^2 y'' + 2xy' + \tfrac{1}{4}y, \qquad 1 \le x \le e,$$

with $y(1) = y(e) = 0$. Find, as a series $\sum a_n y_n(x)$, the solution of $\mathcal{L}y = x^{-1/2}$.

17.14 Express the solution of Poisson's equation in electrostatics,

$$\nabla^2\phi(\mathbf{r}) = -\rho(\mathbf{r})/\epsilon_0,$$

where ρ is the non-zero charge density over a finite part of space, in the form of an integral and hence identify the Green's function for the ∇^2 operator.

17.15 In the quantum-mechanical study of the scattering of a particle by a potential, a Born-approximation solution can be obtained in terms of a function $y(\mathbf{r})$ that satisfies an equation of the form

$$(-\nabla^2 - K^2)y(\mathbf{r}) = F(\mathbf{r}).$$

Assuming that $y_{\mathbf{k}}(\mathbf{r}) = (2\pi)^{-3/2}\exp(i\mathbf{k}\cdot\mathbf{r})$ is a suitably normalised eigenfunction of $-\nabla^2$ corresponding to eigenvalue k^2, find a suitable Green's function $G_K(\mathbf{r},\mathbf{r}')$. By taking the direction of the vector $\mathbf{r}-\mathbf{r}'$ as the polar axis for a $\mathbf{k}$-space integration, show that $G_K(\mathbf{r},\mathbf{r}')$ can be reduced to

$$\frac{1}{4\pi^2|\mathbf{r}-\mathbf{r}'|}\int_{-\infty}^{\infty}\frac{w\sin w}{w^2-w_0^2}\,dw,$$

where $w_0 = K|\mathbf{r}-\mathbf{r}'|$.

[This integral can be evaluated using a contour integration (chapter 24) to give $(4\pi|\mathbf{r}-\mathbf{r}'|)^{-1}\exp(iK|\mathbf{r}-\mathbf{r}'|)$.]

17.8 Hints and answers

17.1 Express the condition $\langle h|h\rangle \geq 0$ as a quadratic equation in λ and then apply the condition for no real roots, noting that $\langle f|g\rangle + \langle g|f\rangle$ is real. To put a limit on $\int y\cos^2 kx\,dx$, set $f = y^{1/2}\cos kx$ and $g = y^{1/2}$ in the inequality.

17.3 Follow an argument similar to that used for proving the reality of the eigenvalues, but integrate from x_1 to x_2, rather than from a to b. Take x_1 and x_2 as two successive zeros of $y_m(x)$ and note that, if the sign of y_m is α then the sign of $y_m'(x_1)$ is α whilst that of $y_m'(x_2)$ is $-\alpha$. Now assume that $y_n(x)$ does not change sign in the interval and has a constant sign β; show that this leads to a contradiction between the signs of the two sides of the identity.

17.5 (a) $y = \sum a_n P_n(x)$ with

$$a_n = \frac{n+1/2}{b-n(n+1)}\int_{-1}^{1} f(z)P_n(z)\,dz;$$

(b) $5x^3 = 2P_3(x)+3P_1(x)$, giving $a_1 = 1/4$ and $a_3 = 1$, leading to $y = 5(2x^3-x)/4$.

17.7 (a) No, $\int gf^{*\prime}dx \neq 0$; (b) yes; (c) no, $i\int f^*g dx \neq 0$; (d) yes.

17.9 The normalised eigenfunctions are $(2/\pi)^{1/2}\sin nx$, with n an integer.
$y(x) = (4/\pi)\sum_{n\text{ odd}}[(-1)^{(n-1)/2}\sin nx]/[n^2(\kappa-n^2)]$.

17.11 $\lambda_n = (n+1/2)^2\pi^2$, $n = 0,1,2,\ldots$.
(a) Since $y_n(1)y_m'(1) \neq 0$, the Sturm–Liouville boundary conditions are not satisfied and the appropriate weight function has to be justified by inspection. The normalised eigenfunctions are $\sqrt{2}e^{-x/2}\sin[(n+1/2)\pi x]$, with $\rho(x) = e^x$.
(b) $y(x) = (-2/\pi^3)\sum_{n=0}^{\infty} e^{-x/2}\sin[(n+1/2)\pi x]/(n+1/2)^3$.

17.13 $y_n(x) = \sqrt{2}x^{-1/2}\sin(n\pi\ln x)$ with $\lambda_n = -n^2\pi^2$;

$$a_n = \begin{cases} -(n\pi)^{-2}\int_1^e \sqrt{2}x^{-1}\sin(n\pi\ln x)\,dx = -\sqrt{8}(n\pi)^{-3} & \text{for } n \text{ odd},\\ 0 & \text{for } n \text{ even}.\end{cases}$$

17.15 Use the form of Green's function that is the integral over all eigenvalues of the 'outer product' of two eigenfunctions corresponding to the same eigenvalue, but with arguments $\mathbf{r}$ and $\mathbf{r}'$.

18

Special functions

In the previous two chapters, we introduced the most important second-order linear ODEs in physics and engineering, listing their regular and irregular singular points in table 16.1 and their Sturm–Liouville forms in table 17.1. These equations occur with such frequency that solutions to them, which obey particular commonly occurring boundary conditions, have been extensively studied and given special names. In this chapter, we discuss these so-called 'special functions' and their properties. In addition, we also discuss some special functions that are not derived from solutions of important second-order ODEs, namely the gamma function and related functions. These convenient functions appear in a number of contexts, and so in section 18.12 we gather together some of their properties, with a minimum of formal proofs.

18.1 Legendre functions

Legendre's differential equation has the form

$$(1-x^2)y'' - 2xy' + \ell(\ell+1)y = 0, \tag{18.1}$$

and has three regular singular points, at $x = -1, 1, \infty$. It occurs in numerous physical applications and particularly in problems with axial symmetry that involve the ∇^2 operator, when they are expressed in spherical polar coordinates. In normal usage the variable x in Legendre's equation is the cosine of the polar angle in spherical polars, and thus $-1 \leq x \leq 1$. The parameter ℓ is a given real number, and any solution of (18.1) is called a *Legendre function*.

In subsection 16.1.1, we showed that $x = 0$ is an ordinary point of (18.1), and so we expect to find two linearly independent solutions of the form $y = \sum_{n=0}^{\infty} a_n x^n$. Substituting, we find

$$\sum_{n=0}^{\infty} \left[n(n-1)a_n x^{n-2} - n(n-1)a_n x^n - 2na_n x^n + \ell(\ell+1)a_n x^n \right] = 0,$$

which on collecting terms gives

$$\sum_{n=0}^{\infty} \{(n+2)(n+1)a_{n+2} - [n(n+1) - \ell(\ell+1)]a_n\} x^n = 0.$$

The recurrence relation is therefore

$$a_{n+2} = \frac{[n(n+1) - \ell(\ell+1)]}{(n+1)(n+2)} a_n, \tag{18.2}$$

for $n = 0, 1, 2, \ldots$. If we choose $a_0 = 1$ and $a_1 = 0$ then we obtain the solution

$$y_1(x) = 1 - \ell(\ell+1)\frac{x^2}{2!} + (\ell-2)\ell(\ell+1)(\ell+3)\frac{x^4}{4!} - \cdots, \tag{18.3}$$

whereas on choosing $a_0 = 0$ and $a_1 = 1$ we find a second solution

$$y_2(x) = x - (\ell-1)(\ell+2)\frac{x^3}{3!} + (\ell-3)(\ell-1)(\ell+2)(\ell+4)\frac{x^5}{5!} - \cdots. \tag{18.4}$$

By applying the ratio test to these series (see subsection 4.3.2), we find that both series converge for $|x| < 1$, and so their radius of convergence is unity, which (as expected) is the distance to the nearest singular point of the equation. Since (18.3) contains only even powers of x and (18.4) contains only odd powers, these two solutions cannot be proportional to one another, and are therefore linearly independent. Hence, the general solution to (18.1) for $|x| < 1$ is

$$y(x) = c_1 y_1(x) + c_2 y_2(x).$$

18.1.1 Legendre functions for integer ℓ

In many physical applications the parameter ℓ in Legendre's equation (18.1) is an integer, i.e. $\ell = 0, 1, 2, \ldots$. In this case, the recurrence relation (18.2) gives

$$a_{\ell+2} = \frac{[\ell(\ell+1) - \ell(\ell+1)]}{(\ell+1)(\ell+2)} a_\ell = 0,$$

i.e. the series terminates and we obtain a polynomial solution of order ℓ. In particular, if ℓ is even, then $y_1(x)$ in (18.3) reduces to a polynomial, whereas if ℓ is odd the same is true of $y_2(x)$ in (18.4). These solutions (suitably normalised) are called the *Legendre polynomials* of order ℓ; they are written $P_\ell(x)$ and are valid for all finite x. It is conventional to normalise $P_\ell(x)$ in such a way that $P_\ell(1) = 1$, and as a consequence $P_\ell(-1) = (-1)^\ell$. The first few Legendre polynomials are easily constructed and are given by

$$\begin{aligned} P_0(x) &= 1, & P_1(x) &= x, \\ P_2(x) &= \tfrac{1}{2}(3x^2 - 1), & P_3(x) &= \tfrac{1}{2}(5x^3 - 3x), \\ P_4(x) &= \tfrac{1}{8}(35x^4 - 30x^2 + 3), & P_5(x) &= \tfrac{1}{8}(63x^5 - 70x^3 + 15x). \end{aligned}$$

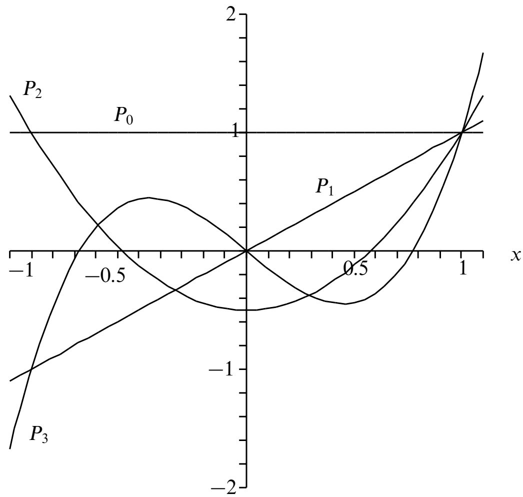


Figure 18.1 The first four Legendre polynomials.

The first four Legendre polynomials are plotted in figure 18.1.

Although, according to whether ℓ is an even or odd integer, respectively, either $y_1(x)$ in (18.3) or $y_2(x)$ in (18.4) terminates to give a multiple of the corresponding Legendre polynomial $P_\ell(x)$, the other series in each case does not terminate and therefore converges only for $|x| < 1$. According to whether ℓ is even or odd, we define *Legendre functions of the second kind* as $Q_\ell(x) = \alpha_\ell y_2(x)$ or $Q_\ell(x) = \beta_\ell y_1(x)$, respectively, where the constants α_ℓ and β_ℓ are conventionally taken to have the values

$$\alpha_\ell = \frac{(-1)^{\ell/2} 2^\ell [(\ell/2)!]^2}{\ell!} \qquad \text{for } \ell \text{ even,} \tag{18.5}$$

$$\beta_\ell = \frac{(-1)^{(\ell+1)/2} 2^{\ell-1} \{[(\ell-1)/2]!\}^2}{\ell!} \qquad \text{for } \ell \text{ odd.} \tag{18.6}$$

These normalisation factors are chosen so that the $Q_\ell(x)$ obey the same recurrence relations as the $P_\ell(x)$ (see subsection 18.1.2).

The general solution of Legendre's equation for *integer* ℓ is therefore

$$y(x) = c_1 P_\ell(x) + c_2 Q_\ell(x), \tag{18.7}$$

where $P_\ell(x)$ is a polynomial of order ℓ, and so converges for all x, and $Q_\ell(x)$ is an infinite series that converges only for $|x| < 1$.§

By using the Wronskian method, section 16.4, we may obtain closed forms for the $Q_\ell(x)$.

►*Use the Wronskian method to find a closed-form expression for $Q_0(x)$.*

From (16.25) a second solution to Legendre's equation (18.1), with $\ell = 0$, is

$$\begin{aligned} y_2(x) &= P_0(x) \int^x \frac{1}{[P_0(u)]^2} \exp\left(\int^u \frac{2v}{1-v^2}\, dv \right) du \\ &= \int^x \exp\left[-\ln(1-u^2) \right] du \\ &= \int^x \frac{du}{(1-u^2)} = \tfrac{1}{2} \ln\left(\frac{1+x}{1-x} \right), \end{aligned} \tag{18.8}$$

where in the second line we have used the fact that $P_0(x) = 1$.

All that remains is to adjust the normalisation of this solution so that it agrees with (18.5). Expanding the logarithm in (18.8) as a Maclaurin series we obtain

$$y_2(x) = x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots .$$

Comparing this with the expression for $Q_0(x)$, using (18.4) with $\ell = 0$ and the normalisation (18.5), we find that $y_2(x)$ is already correctly normalised, and so

$$Q_0(x) = \tfrac{1}{2} \ln\left(\frac{1+x}{1-x} \right).$$

Of course, we might have recognised the series (18.4) for $\ell = 0$, but to do so for larger ℓ would prove progressively more difficult. ◄

Using the above method for $\ell = 1$, we find

$$Q_1(x) = \tfrac{1}{2} x \ln\left(\frac{1+x}{1-x} \right) - 1.$$

Closed forms for higher-order $Q_\ell(x)$ may now be found using the recurrence relation (18.27) derived in the next subsection. The first few Legendre functions of the second kind are plotted in figure 18.2.

18.1.2 Properties of Legendre polynomials

As stated earlier, when encountered in physical problems the variable x in Legendre's equation is usually the cosine of the polar angle θ in spherical polar coordinates, and we then require the solution $y(x)$ to be regular at $x = \pm 1$, which corresponds to $\theta = 0$ or $\theta = \pi$. For this to occur we require the equation to have a polynomial solution, and so ℓ must be an integer. Furthermore, we also require

§ It is possible, in fact, to find a second solution in terms of an infinite series of *negative* powers of x that is finite for $|x| > 1$ (see exercise 16.16).

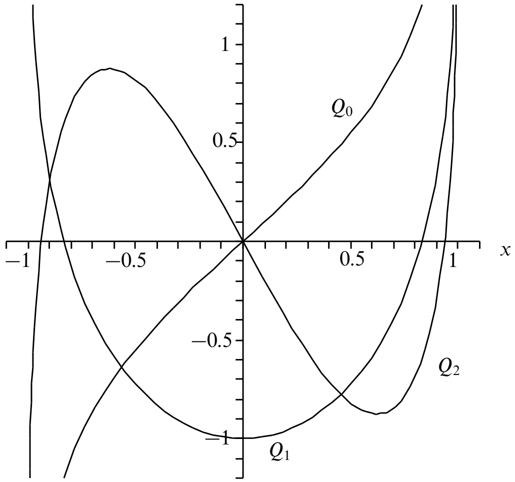


Figure 18.2 The first three Legendre functions of the second kind.

the coefficient c_2 of the function $Q_\ell(x)$ in (18.7) to be zero, since $Q_\ell(x)$ is singular at $x = \pm 1$, with the result that the general solution is simply some multiple of the relevant Legendre polynomial $P_\ell(x)$. In this section we will study the properties of the Legendre polynomials $P_\ell(x)$ in some detail.

Rodrigues' formula

As an aid to establishing further properties of the Legendre polynomials we now develop Rodrigues' representation of these functions. Rodrigues' formula for the $P_\ell(x)$ is

$$P_\ell(x) = \frac{1}{2^\ell \ell!} \frac{d^\ell}{dx^\ell}(x^2 - 1)^\ell. \tag{18.9}$$

To prove that this is a representation we let $u = (x^2 - 1)^\ell$, so that $u' = 2\ell x(x^2 - 1)^{\ell - 1}$ and

$$(x^2 - 1)u' - 2\ell x u = 0.$$

If we differentiate this expression $\ell + 1$ times using Leibnitz' theorem, we obtain

$$\left[(x^2 - 1)u^{(\ell+2)} + 2x(\ell + 1)u^{(\ell+1)} + \ell(\ell + 1)u^{(\ell)}\right] - 2\ell\left[xu^{(\ell+1)} + (\ell + 1)u^{(\ell)}\right] = 0,$$

which reduces to

$$(x^2-1)u^{(\ell+2)}+2xu^{(\ell+1)}-\ell(\ell+1)u^{(\ell)}=0.$$

Changing the sign all through, we recover Legendre's equation (18.1) with $u^{(\ell)}$ as the dependent variable. Since, from (18.9), ℓ is an integer and $u^{(\ell)}$ is regular at $x=\pm 1$, we may make the identification

$$u^{(\ell)}(x)=c_\ell P_\ell(x), \tag{18.10}$$

for some constant c_ℓ that depends on ℓ. To establish the value of c_ℓ we note that the only term in the expression for the ℓth derivative of $(x^2-1)^\ell$ that does not contain a factor x^2-1, and therefore does not vanish at $x=1$, is $(2x)^\ell \ell!(x^2-1)^0$. Putting $x=1$ in (18.10) and recalling that $P_\ell(1)=1$, therefore shows that $c_\ell=2^\ell \ell!$, thus completing the proof of Rodrigues' formula (18.9).

►*Use Rodrigues' formula to show that*

$$I_\ell=\int_{-1}^{1}P_\ell(x)P_\ell(x)\,dx=\frac{2}{2\ell+1}. \tag{18.11}$$

The result is trivially obvious for $\ell=0$ and so we assume $\ell\geq 1$. Then, by Rodrigues' formula,

$$I_\ell=\frac{1}{2^{2\ell}(\ell!)^2}\int_{-1}^{1}\left[\frac{d^\ell(x^2-1)^\ell}{dx^\ell}\right]\left[\frac{d^\ell(x^2-1)^\ell}{dx^\ell}\right]dx.$$

Repeated integration by parts, with all boundary terms vanishing, reduces this to

$$\begin{aligned}I_\ell&=\frac{(-1)^\ell}{2^{2\ell}(\ell!)^2}\int_{-1}^{1}(x^2-1)^\ell\frac{d^{2\ell}}{dx^{2\ell}}(x^2-1)^\ell\,dx\\&=\frac{(2\ell)!}{2^{2\ell}(\ell!)^2}\int_{-1}^{1}(1-x^2)^\ell\,dx.\end{aligned}$$

If we write

$$K_\ell=\int_{-1}^{1}(1-x^2)^\ell\,dx,$$

then integration by parts (taking a factor 1 as the second part) gives

$$K_\ell=\int_{-1}^{1}2\ell x^2(1-x^2)^{\ell-1}\,dx.$$

Writing $2\ell x^2$ as $2\ell-2\ell(1-x^2)$ we obtain

$$\begin{aligned}K_\ell&=2\ell\int_{-1}^{1}(1-x^2)^{\ell-1}\,dx-2\ell\int_{-1}^{1}(1-x^2)^\ell\,dx\\&=2\ell K_{\ell-1}-2\ell K_\ell\end{aligned}$$

and hence the recurrence relation $(2\ell+1)K_\ell=2\ell K_{\ell-1}$. We therefore find

$$K_\ell=\frac{2\ell}{2\ell+1}\,\frac{2\ell-2}{2\ell-1}\cdots\frac{2}{3}K_0=2^\ell \ell!\,\frac{2^\ell \ell!}{(2\ell+1)!}2=\frac{2^{2\ell+1}(\ell!)^2}{(2\ell+1)!},$$

which, when substituted into the expression for I_ℓ, establishes the required result. ◄

Mutual orthogonality

In section 17.4, we noted that Legendre's equation was of Sturm–Liouville form with $p = 1 - x^2$, $q = 0$, $\lambda = \ell(\ell+1)$ and $\rho = 1$, and that its natural interval was $[-1, 1]$. Since the Legendre polynomials $P_\ell(x)$ are regular at the end-points $x = \pm 1$, they must be mutually orthogonal over this interval, i.e.

$$\int_{-1}^{1} P_\ell(x) P_k(x)\, dx = 0 \qquad \text{if } \ell \neq k. \tag{18.12}$$

Although this result follows from the general considerations of the previous chapter, it may also be proved directly, as shown in the following example.

►*Prove directly that the Legendre polynomials $P_\ell(x)$ are mutually orthogonal over the interval $-1 < x < 1$.*

Since the $P_\ell(x)$ satisfy Legendre's equation we may write

$$\left[(1 - x^2) P_\ell'\right]' + \ell(\ell+1) P_\ell = 0,$$

where $P_\ell' = dP_\ell/dx$. Multiplying through by P_k and integrating from $x = -1$ to $x = 1$, we obtain

$$\int_{-1}^{1} P_k \left[(1 - x^2) P_\ell'\right]' dx + \int_{-1}^{1} P_k \ell(\ell+1) P_\ell\, dx = 0.$$

Integrating the first term by parts and noting that the boundary contribution vanishes at both limits because of the factor $1 - x^2$, we find

$$-\int_{-1}^{1} P_k'(1 - x^2) P_\ell'\, dx + \int_{-1}^{1} P_k \ell(\ell+1) P_\ell\, dx = 0.$$

Now, if we reverse the roles of ℓ and k and subtract one expression from the other, we conclude that

$$[k(k+1) - \ell(\ell+1)] \int_{-1}^{1} P_k P_\ell\, dx = 0,$$

and therefore, since $k \neq \ell$, we must have the result (18.12). As a particular case, we note that if we put $k = 0$ we obtain

$$\int_{-1}^{1} P_\ell(x)\, dx = 0 \qquad \text{for } \ell \neq 0. \blacktriangleleft$$

As we discussed in the previous chapter, the mutual orthogonality (and completeness) of the $P_\ell(x)$ means that any reasonable function $f(x)$ (i.e. one obeying the Dirichlet conditions discussed at the start of chapter 12) can be expressed in the interval $|x| < 1$ as an infinite sum of Legendre polynomials,

$$f(x) = \sum_{\ell=0}^{\infty} a_\ell P_\ell(x), \tag{18.13}$$

where the coefficients a_ℓ are given by

$$a_\ell = \frac{2\ell + 1}{2} \int_{-1}^{1} f(x) P_\ell(x)\, dx. \tag{18.14}$$

►*Prove the expression (18.14) for the coefficients in the Legendre polynomial expansion of a function $f(x)$.*

If we multiply (18.13) by $P_k(x)$ and integrate from $x = -1$ to $x = 1$ then we obtain

$$\int_{-1}^{1} P_k(x) f(x)\, dx = \sum_{\ell=0}^{\infty} a_\ell \int_{-1}^{1} P_k(x) P_\ell(x)\, dx$$
$$= a_k \int_{-1}^{1} P_k(x) P_k(x)\, dx = \frac{2a_k}{2k+1},$$

where we have used the orthogonality property (18.12) and the normalisation property (18.11). ◄

Generating function

A useful device for manipulating and studying sequences of functions or quantities labelled by an integer variable (here, the Legendre polynomials $P_\ell(x)$ labelled by ℓ) is a *generating function*. The generating function has perhaps its greatest utility in the area of probability theory (see chapter 30). However, it is also a great convenience in our present study.

The generating function for, say, a series of functions $f_n(x)$ for $n = 0, 1, 2, \ldots$ is a function $G(x, h)$ containing, as well as x, a dummy variable h such that

$$G(x, h) = \sum_{n=0}^{\infty} f_n(x) h^n,$$

i.e. $f_n(x)$ is the coefficient of h^n in the expansion of G in powers of h. The utility of the device lies in the fact that sometimes it is possible to find a closed form for $G(x, h)$.

For our study of Legendre polynomials let us consider the functions $P_n(x)$ defined by the equation

$$G(x, h) = (1 - 2xh + h^2)^{-1/2} = \sum_{n=0}^{\infty} P_n(x) h^n. \tag{18.15}$$

As we show below, the functions so defined are identical to the Legendre polynomials and the function $(1 - 2xh + h^2)^{-1/2}$ is in fact the generating function for them. In the process we will also deduce several useful relationships between the various polynomials and their derivatives.

►*Show that the functions $P_n(x)$ defined by (18.15) satisfy Legendre's equation*

In the following $dP_n(x)/dx$ will be denoted by P_n'. Firstly, we differentiate the defining equation (18.15) with respect to x and get

$$h(1 - 2xh + h^2)^{-3/2} = \sum P_n' h^n. \tag{18.16}$$

Also, we differentiate (18.15) with respect to h to yield

$$(x - h)(1 - 2xh + h^2)^{-3/2} = \sum n P_n h^{n-1}. \tag{18.17}$$

Equation (18.16) can then be written, using (18.15), as

$$h \sum P_n h^n = (1 - 2xh + h^2) \sum P_n' h^n,$$

and equating the coefficients of h^{n+1} we obtain the recurrence relation

$$P_n = P_{n+1}' - 2xP_n' + P_{n-1}'. \tag{18.18}$$

Equations (18.16) and (18.17) can be combined as

$$(x - h) \sum P_n' h^n = h \sum n P_n h^{n-1},$$

from which the coefficent of h^n yields a second recurrence relation,

$$xP_n' - P_{n-1}' = nP_n; \tag{18.19}$$

eliminating P_{n-1}' between (18.18) and (18.19) then gives the further result

$$(n+1)P_n = P_{n+1}' - xP_n'. \tag{18.20}$$

If we now take the result (18.20) with n replaced by $n-1$ and add x times (18.19) to it we obtain

$$(1 - x^2)P_n' = n(P_{n-1} - xP_n). \tag{18.21}$$

Finally, differentiating both sides with respect to x and using (18.19) again, we find

$$\begin{aligned}(1 - x^2)P_n'' - 2xP_n' &= n[(P_{n-1}' - xP_n') - P_n] \\ &= n(-nP_n - P_n) = -n(n+1)P_n,\end{aligned}$$

and so the P_n defined by (18.15) do indeed satisfy Legendre's equation. ◄

The above example shows that the functions $P_n(x)$ defined by (18.15) satisfy Legendre's equation with $\ell = n$ (an integer) and, also from (18.15), these functions are regular at $x = \pm 1$. Thus P_n must be some multiple of the nth Legendre polynomial. It therefore remains only to verify the normalisation. This is easily done at $x = 1$, when G becomes

$$G(1, h) = [(1 - h)^2]^{-1/2} = 1 + h + h^2 + \cdots,$$

and we can see that all the P_n so defined have $P_n(1) = 1$ as required, and are thus identical to the Legendre polynomials.

A particular use of the generating function (18.15) is in representing the inverse distance between two points in three-dimensional space in terms of Legendre polynomials. If two points $\mathbf{r}$ and $\mathbf{r}'$ are at distances r and r', respectively, from the origin, with $r' < r$, then

$$\begin{aligned}\frac{1}{|\mathbf{r} - \mathbf{r}'|} &= \frac{1}{(r^2 + r'^2 - 2rr'\cos\theta)^{1/2}} \\ &= \frac{1}{r[1 - 2(r'/r)\cos\theta + (r'/r)^2]^{1/2}} \\ &= \frac{1}{r} \sum_{\ell=0}^{\infty} \left(\frac{r'}{r}\right)^{\ell} P_\ell(\cos\theta),\end{aligned} \tag{18.22}$$

where θ is the angle between the two position vectors $\mathbf{r}$ and $\mathbf{r}'$. If $r' > r$, however,

r and r' must be exchanged in (18.22) or the series would not converge. This result may be used, for example, to write down the electrostatic potential at a point $\mathbf{r}$ due to a charge q at the point $\mathbf{r}'$. Thus, in the case $r' < r$, this is given by

$$V(\mathbf{r}) = \frac{q}{4\pi\epsilon_0 r}\sum_{\ell=0}^{\infty}\left(\frac{r'}{r}\right)^{\ell} P_\ell(\cos\theta).$$

We note that in the special case where the charge is at the origin, and $r' = 0$, only the $\ell = 0$ term in the series is non-zero and the expression reduces correctly to the familiar form $V(\mathbf{r}) = q/(4\pi\epsilon_0 r)$.

Recurrence relations

In our discussion of the generating function above, we derived several useful recurrence relations satisfied by the Legendre polynomials $P_n(x)$. In particular, from (18.18), we have the four-term recurrence relation

$$P'_{n+1} + P'_{n-1} = P_n + 2xP'_n.$$

Also, from (18.19)–(18.21), we have the three-term recurrence relations

$$P'_{n+1} = (n+1)P_n + xP'_n, \tag{18.23}$$
$$P'_{n-1} = -nP_n + xP'_n, \tag{18.24}$$
$$(1-x^2)P'_n = n(P_{n-1} - xP_n), \tag{18.25}$$
$$(2n+1)P_n = P'_{n+1} - P'_{n-1}, \tag{18.26}$$

where the final relation is obtained immediately by subtracting the second from the first. Many other useful recurrence relations can be derived from those given above and from the generating function.

►*Prove the recurrence relation*

$$(n+1)P_{n+1} = (2n+1)xP_n - nP_{n-1}. \tag{18.27}$$

Substituting from (18.15) into (18.17), we find

$$(x-h)\sum P_n h^n = (1 - 2xh + h^2)\sum nP_n h^{n-1}.$$

Equating coefficients of h^n we obtain

$$xP_n - P_{n-1} = (n+1)P_{n+1} - 2xnP_n + (n-1)P_{n-1},$$

which on rearrangement gives the stated result. ◄

The recurrence relation derived in the above example is particularly useful in evaluating $P_n(x)$ for a given value of x. One starts with $P_0(x) = 1$ and $P_1(x) = x$ and iterates the recurrence relation until $P_n(x)$ is obtained.

18.2 Associated Legendre functions

The associated Legendre equation has the form

$$(1-x^2)y'' - 2xy' + \left[\ell(\ell+1) - \frac{m^2}{1-x^2}\right] y = 0, \tag{18.28}$$

which has three regular singular points at $x = -1, 1, \infty$ and reduces to Legendre's equation (18.1) when $m = 0$. It occurs in physical applications involving the operator ∇^2, when expressed in spherical polars. In such cases, $-\ell \le m \le \ell$ and m is restricted to integer values, which we will assume from here on. As was the case for Legendre's equation, in normal usage the variable x is the cosine of the polar angle in spherical polars, and thus $-1 \le x \le 1$. Any solution of (18.28) is called an *associated Legendre function.*

The point $x = 0$ is an ordinary point of (18.28), and one could obtain series solutions of the form $y = \sum_{n=0} a_n x^n$ in the same manner as that used for Legendre's equation. In this case, however, it is more instructive to note that if $u(x)$ is a solution of Legendre's equation (18.1), then

$$y(x) = (1-x^2)^{|m|/2} \frac{d^{|m|}u}{dx^{|m|}} \tag{18.29}$$

is a solution of the associated equation (18.28).

►*Prove that if $u(x)$ is a solution of Legendre's equation, then $y(x)$ given in (18.29) is a solution of the associated equation.*

For simplicity, let us begin by assuming that m is non-negative. Legendre's equation for u reads

$$(1-x^2)u'' - 2xu' + \ell(\ell+1)u = 0,$$

and, on differentiating this equation m times using Leibnitz' theorem, we obtain

$$(1-x^2)v'' - 2x(m+1)v' + (\ell-m)(\ell+m+1)v = 0, \tag{18.30}$$

where $v(x) = d^m u/dx^m$. On setting

$$y(x) = (1-x^2)^{m/2} v(x),$$

the derivatives v' and v'' may be written as

$$v' = (1-x^2)^{-m/2}\left(y' + \frac{mx}{1-x^2}y\right),$$
$$v'' = (1-x^2)^{-m/2}\left[y'' + \frac{2mx}{1-x^2}y' + \frac{m}{1-x^2}y + \frac{m(m+2)x^2}{(1-x^2)^2}y\right].$$

Substituting these expressions into (18.30) and simplifying, we obtain

$$(1-x^2)y'' - 2xy' + \left[\ell(\ell+1) - \frac{m^2}{1-x^2}\right] y = 0,$$

which shows that y is a solution of the associated Legendre equation (18.28). Finally, we note that if m is negative, the value of m^2 is unchanged, and so a solution for positive m is also a solution for the corresponding negative value of m. ◄

From the two linearly independent series solutions to Legendre's equation given in (18.3) and (18.4), which we now denote by $u_1(x)$ and $u_2(x)$, we may obtain two linearly-independent series solutions, $y_1(x)$ and $y_2(x)$, to the associated equation by using (18.29). From the general discussion of the convergence of power series given in section 4.5.1, we see that both $y_1(x)$ and $y_2(x)$ will also converge for $|x| < 1$. Hence the general solution to (18.28) in this range is given by

$$y(x) = c_1 y_1(x) + c_2 y_2(x).$$

18.2.1 Associated Legendre functions for integer ℓ

If ℓ and m are both integers, as is the case in many physical applications, then the general solution to (18.28) is denoted by

$$y(x) = c_1 P_\ell^m(x) + c_2 Q_\ell^m(x), \tag{18.31}$$

where $P_\ell^m(x)$ and $Q_\ell^m(x)$ are associated Legendre functions of the first and second kind, respectively. For non-negative values of m, these functions are related to the ordinary Legendre functions for integer ℓ by

$$P_\ell^m(x) = (1-x^2)^{m/2}\frac{d^m P_\ell}{dx^m}, \qquad Q_\ell^m(x) = (1-x^2)^{m/2}\frac{d^m Q_\ell}{dx^m}. \tag{18.32}$$

We see immediately that, as required, the associated Legendre functions reduce to the ordinary Legendre functions when $m = 0$. Since it is m^2 that appears in the associated Legendre equation (18.28), the associated Legendre functions for negative m values must be proportional to the corresponding function for non-negative m. The constant of proportionality is a matter of convention. For the $P_\ell^m(x)$ it is usual to regard the definition (18.32) as being valid also for negative m values. Although differentiating a negative number of times is not defined, when $P_\ell(x)$ is expressed in terms of the Rodrigues' formula (18.9), this problem does not occur for $-\ell \le m \le \ell$.[§] In this case,

$$P_\ell^{-m}(x) = (-1)^m \frac{(\ell-m)!}{(\ell+m)!} P_\ell^m(x). \tag{18.33}$$

►*Prove the result (18.33).*

From (18.32) and the Rodrigues' formula (18.9) for the Legendre polynomials, we have

$$P_\ell^m(x) = \frac{1}{2^\ell \ell!}(1-x^2)^{m/2}\frac{d^{\ell+m}}{dx^{\ell+m}}(x^2-1)^\ell,$$

and, without loss of generality, we may assume that m is non-negative. It is convenient to

[§] Some authors define $P_\ell^{-m}(x) = P_\ell^m(x)$, and similarly for the $Q_\ell^m(x)$, in which case m is replaced by $|m|$ in the definitions (18.32). It should be noted that, in this case, many of the results presented in this section also require m to be replaced by $|m|$.

write $(x^2-1)=(x+1)(x-1)$ and use Leibnitz' theorem to evaluate the derivative, which yields

$$P_\ell^m(x)=\frac{1}{2^\ell \ell!}(1-x^2)^{m/2}\sum_{r=0}^{\ell+m}\frac{(\ell+m)!}{r!(\ell+m-r)!}\frac{d^r(x+1)^\ell}{dx^r}\frac{d^{\ell+m-r}(x-1)^\ell}{dx^{\ell+m-r}}.$$

Considering the two derivative factors in a term in the summation, we note that the first is non-zero only for $r\le\ell$ and the second is non-zero for $\ell+m-r\le\ell$. Combining these conditions yields $m\le r\le\ell$. Performing the derivatives, we thus obtain

$$\begin{aligned}P_\ell^m(x)&=\frac{1}{2^\ell \ell!}(1-x^2)^{m/2}\sum_{r=m}^{\ell}\frac{(\ell+m)!}{r!(\ell+m-r)!}\frac{\ell!(x+1)^{\ell-r}}{(\ell-r)!}\frac{\ell!(x-1)^{r-m}}{(r-m)!}\\&=(-1)^{m/2}\frac{\ell!(\ell+m)!}{2^\ell}\sum_{r=m}^{\ell}\frac{(x+1)^{\ell-r+\frac{m}{2}}(x-1)^{r-\frac{m}{2}}}{r!(\ell+m-r)!(\ell-r)!(r-m)!}.\end{aligned}\tag{18.34}$$

Repeating the above calculation for $P_\ell^{-m}(x)$ and identifying once more those terms in the sum that are non-zero, we find

$$\begin{aligned}P_\ell^{-m}(x)&=(-1)^{-m/2}\frac{\ell!(\ell-m)!}{2^\ell}\sum_{r=0}^{\ell-m}\frac{(x+1)^{\ell-r-\frac{m}{2}}(x-1)^{r+\frac{m}{2}}}{r!(\ell-m-r)!(\ell-r)!(r+m)!}\\&=(-1)^{-m/2}\frac{\ell!(\ell-m)!}{2^\ell}\sum_{\bar r=m}^{\ell}\frac{(x+1)^{\ell-\bar r+\frac{m}{2}}(x-1)^{\bar r-\frac{m}{2}}}{(\bar r-m)!(\ell-\bar r)!(\ell+m-\bar r)!\bar r!},\end{aligned}\tag{18.35}$$

where, in the second equality, we have rewritten the summation in terms of the new index $\bar r=r+m$. Comparing (18.34) and (18.35), we immediately arrive at the required result (18.33). ◄

Since $P_\ell(x)$ is a polynomial of order ℓ, we have $P_\ell^m(x)=0$ for $|m|>\ell$. From its definition, it is clear that $P_\ell^m(x)$ is also a polynomial of order ℓ if m is even, but contains the factor $(1-x^2)$ to a fractional power if m is odd. In either case, $P_\ell^m(x)$ is regular at $x=\pm1$. The first few associated Legendre functions of the first kind are easily constructed and are given by (omitting the $m=0$ cases)

$$\begin{aligned}P_1^1(x)&=(1-x^2)^{1/2}, & P_2^1(x)&=3x(1-x^2)^{1/2},\\P_2^2(x)&=3(1-x^2), & P_3^1(x)&=\tfrac{3}{2}(5x^2-1)(1-x^2)^{1/2},\\P_3^2(x)&=15x(1-x^2), & P_3^3(x)&=15(1-x^2)^{3/2}.\end{aligned}$$

Finally, we note that the associated Legendre functions of the second kind $Q_\ell^m(x)$, like $Q_\ell(x)$, are singular at $x=\pm1$.

18.2.2 Properties of associated Legendre functions $P_\ell^m(x)$

When encountered in physical problems, the variable x in the associated Legendre equation (as in the ordinary Legendre equation) is usually the cosine of the polar angle θ in spherical polar coordinates, and we then require the solution $y(x)$ to be regular at $x=\pm1$ (corresponding to $\theta=0$ or $\theta=\pi$). For this to occur, we require ℓ to be an integer and the coefficient c_2 of the function $Q_\ell^m(x)$ in (18.31) to be zero, since $Q_\ell^m(x)$ is singular at $x = \pm 1$, with the result that the general solution is simply some multiple of one of the associated Legendre functions of the first kind, $P_\ell^m(x)$. We will study the further properties of these functions in the remainder of this subsection.

Mutual orthogonality

As noted in section 17.4, the associated Legendre equation is of Sturm–Liouville form $(py)' + qy + \lambda\rho y = 0$, with $p = 1 - x^2$, $q = -m^2/(1 - x^2)$, $\lambda = \ell(\ell + 1)$ and $\rho = 1$, and its natural interval is thus $[-1, 1]$. Since the associated Legendre functions $P_\ell^m(x)$ are regular at the end-points $x = \pm 1$, they must be mutually orthogonal over this interval for a fixed value of m, i.e.

$$\int_{-1}^{1} P_\ell^m(x) P_k^m(x)\, dx = 0 \qquad \text{if } \ell \neq k. \tag{18.36}$$

This result may also be proved directly in a manner similar to that used for demonstrating the orthogonality of the Legendre polynomials $P_\ell(x)$ in section 18.1.2. Note that the value of m must be the same for the two associated Legendre functions for (18.36) to hold. The normalisation condition when $\ell = k$ may be obtained using the Rodrigues' formula, as shown in the following example.

►*Show that*

$$I_{\ell m} \equiv \int_{-1}^{1} P_\ell^m(x) P_\ell^m(x)\, dx = \frac{2}{2\ell + 1} \frac{(\ell + m)!}{(\ell - m)!}. \tag{18.37}$$

From the definition (18.32) and the Rodrigues' formula (18.9) for $P_\ell(x)$, we may write

$$I_{\ell m} = \frac{1}{2^{2\ell}(\ell!)^2} \int_{-1}^{1} \left[(1 - x^2)^m \frac{d^{\ell+m}(x^2 - 1)^\ell}{dx^{\ell+m}}\right] \left[\frac{d^{\ell+m}(x^2 - 1)^\ell}{dx^{\ell+m}}\right] dx,$$

where the square brackets identify the factors to be used when integrating by parts. Performing the integration by parts $\ell + m$ times, and noting that all boundary terms vanish, we obtain

$$I_{\ell m} = \frac{(-1)^{\ell+m}}{2^{2\ell}(\ell!)^2} \int_{-1}^{1} (x^2 - 1)^\ell \frac{d^{\ell+m}}{dx^{\ell+m}} \left[(1 - x^2)^m \frac{d^{\ell+m}(x^2 - 1)^\ell}{dx^{\ell+m}}\right] dx.$$

Using Leibnitz' theorem, the second factor in the integrand may be written as

$$\frac{d^{\ell+m}}{dx^{\ell+m}} \left[(1 - x^2)^m \frac{d^{\ell+m}(x^2 - 1)^\ell}{dx^{\ell+m}}\right] = \sum_{r=0}^{\ell+m} \frac{(\ell + m)!}{r!(\ell + m - r)!} \frac{d^r(1 - x^2)^m}{dx^r} \frac{d^{2\ell+2m-r}(x^2 - 1)^\ell}{dx^{2\ell+2m-r}}.$$

Considering the two derivative factors in a term in the summation on the RHS, we see that the first is non-zero only for $r \leq 2m$, whereas the second is non-zero only for $2\ell + 2m - r \leq 2\ell$. Combining these conditions, we find that the only non-zero term in the sum is that for which $r = 2m$. Thus, we may write

$$I_{\ell m} = \frac{(-1)^{\ell+m}}{2^{2\ell}(\ell!)^2} \frac{(\ell + m)!}{(2m)!(\ell - m)!} \int_{-1}^{1} (1 - x^2)^\ell \frac{d^{2m}(1 - x^2)^m}{dx^{2m}} \frac{d^{2\ell}(1 - x^2)^\ell}{dx^{2\ell}}\, dx.$$

Since $d^{2\ell}(1-x^2)^\ell/dx^{2\ell} = (-1)^\ell(2\ell)!$, and noting that $(-1)^{2\ell+2m} = 1$, we have

$$I_{\ell m} = \frac{1}{2^{2\ell}(\ell!)^2}\frac{(2\ell)!(\ell+m)!}{(\ell-m)!}\int_{-1}^{1}(1-x^2)^\ell\,dx.$$

We have already shown in section 18.1.2 that

$$K_\ell \equiv \int_{-1}^{1}(1-x^2)^\ell\,dx = \frac{2^{2\ell+1}(\ell!)^2}{(2\ell+1)!},$$

and so we obtain the final result

$$I_{\ell m} = \frac{2}{2\ell+1}\frac{(\ell+m)!}{(\ell-m)!}. \blacktriangleleft$$

The orthogonality and normalisation conditions, (18.36) and (18.37) respectively, mean that the associated Legendre functions $P_\ell^m(x)$, with m fixed, may be used in a similar way to the Legendre polynomials to expand any reasonable function $f(x)$ on the interval $|x| < 1$ in a series of the form

$$f(x) = \sum_{k=0}^{\infty} a_{m+k}P_{m+k}^m(x), \tag{18.38}$$

where, in this case, the coefficients are given by

$$a_\ell = \frac{2\ell+1}{2}\frac{(\ell-m)!}{(\ell+m)!}\int_{-1}^{1} f(x)P_\ell^m(x)\,dx.$$

We note that the series takes the form (18.38) because $P_\ell^m(x) = 0$ for $m > \ell$.

Finally, it is worth noting that the associated Legendre functions $P_\ell^m(x)$ must also obey a second orthogonality relationship. This has to be so because one may equally well write the associated Legendre equation (18.28) in Sturm–Liouville form $(py)'+qy+\lambda\rho y = 0$, with $p = 1-x^2$, $q = \ell(\ell+1)$, $\lambda = -m^2$ and $\rho = (1-x^2)^{-1}$; once again the natural interval is $[-1, 1]$. Since the associated Legendre functions $P_\ell^m(x)$ are regular at the end-points $x = \pm 1$, they must therefore be mutually orthogonal with respect to the weight function $(1-x^2)^{-1}$ over this interval for a fixed value of ℓ, i.e.

$$\int_{-1}^{1} P_\ell^m(x)P_\ell^k(x)(1-x^2)^{-1}\,dx = 0 \qquad \text{if } |m| \neq |k|. \tag{18.39}$$

One may also show straightforwardly that the corresponding normalisation condition when $m = k$ is given by

$$\int_{-1}^{1} P_\ell^m(x)P_\ell^m(x)(1-x^2)^{-1}\,dx = \frac{(\ell+m)!}{m(\ell-m)!}.$$

In solving physical problems, however, the orthogonality condition (18.39) is not of any practical use.

Generating function

The generating function for associated Legendre functions can be easily derived by combining their definition (18.32) with the generating function for the Legendre polynomials given in (18.15). We find that

$$G(x,h) = \frac{(2m)!(1-x^2)^{m/2}}{2^m m!(1-2hx+h^2)^{m+1/2}} = \sum_{n=0}^{\infty} P^m_{n+m}(x)h^n. \tag{18.40}$$

►*Derive the expression (18.40) for the associated Legendre generating function.*

The generating function (18.15) for the Legendre polynomials reads

$$\sum_{n=0}^{\infty} P_n h^n = (1-2xh+h^2)^{-1/2}.$$

Differentiating both sides of this result m times (assumimg m to be non-negative), mutliplying through by $(1-x^2)^{m/2}$ and using the definition (18.32) of the associated Legendre functions, we obtain

$$\sum_{n=0}^{\infty} P^m_n h^n = (1-x^2)^{m/2}\frac{d^m}{dx^m}(1-2xh+h^2)^{-1/2}.$$

Performing the derivatives on the RHS gives

$$\sum_{n=0}^{\infty} P^m_n h^n = \frac{1\cdot 3\cdot 5\cdots(2m-1)(1-x^2)^{m/2}h^m}{(1-2xh+h^2)^{m+1/2}}.$$

Dividing through by h^m, re-indexing the summation on the LHS and noting that, quite generally,

$$1\cdot 3\cdot 5\cdots(2r-1) = \frac{1\cdot 2\cdot 3\cdots 2r}{2\cdot 4\cdot 6\cdots 2r} = \frac{(2r)!}{2^r r!},$$

we obtain the final result (18.40). ◄

Recurrence relations

As one might expect, the associated Legendre functions satisfy certain recurrence relations. Indeed, the presence of the two indices n and m means that a much wider range of recurrence relations may be derived. Here we shall content ourselves with quoting just four of the most useful relations:

$$P^{m+1}_n = \frac{2mx}{(1-x^2)^{1/2}}P^m_n + [m(m-1)-n(n+1)]P^{m-1}_n, \tag{18.41}$$

$$(2n+1)xP^m_n = (n+m)P^m_{n-1} + (n-m+1)P^m_{n+1}, \tag{18.42}$$

$$(2n+1)(1-x^2)^{1/2}P^m_n = P^{m+1}_{n+1} - P^{m+1}_{n-1}, \tag{18.43}$$

$$2(1-x^2)^{1/2}(P^m_n)' = P^{m+1}_n - (n+m)(n-m+1)P^{m-1}_n. \tag{18.44}$$

We note that, by virtue of our adopted definition (18.32), these recurrence relations are equally valid for negative and non-negative values of m. These relations may be derived in a number of ways, such as using the generating function (18.40) or by differentiation of the recurrence relations for the Legendre polynomials $P_\ell(x)$.

►*Use the recurrence relation* $(2n+1)P_n = P'_{n+1} - P'_{n-1}$ *for Legendre polynomials to derive the result (18.43).*

Differentiating the recurrence relation for the Legendre polynomials m times, we have

$$(2n+1)\frac{d^m P_n}{dx^m} = \frac{d^{m+1} P_{n+1}}{dx^{m+1}} - \frac{d^{m+1} P_{n-1}}{dx^{m+1}}.$$

Multiplying through by $(1-x^2)^{(m+1)/2}$ and using the definition (18.32) immediately gives the result (18.43). ◄

18.3 Spherical harmonics

The associated Legendre functions discussed in the previous section occur most commonly when obtaining solutions in spherical polar coordinates of Laplace's equation $\nabla^2 u = 0$ (see section 21.3.1). In particular, one finds that, for solutions that are finite on the polar axis, the angular part of the solution is given by

$$\Theta(\theta)\Phi(\phi) = P_\ell^m(\cos\theta)(C\cos m\phi + D\sin m\phi),$$

where ℓ and m are integers with $-\ell \le m \le \ell$. This general form is sufficiently common that particular functions of θ and ϕ called *spherical harmonics* are defined and tabulated. The spherical harmonics $Y_\ell^m(\theta,\phi)$ are defined by

$$Y_\ell^m(\theta,\phi) = (-1)^m \left[\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}\right]^{1/2} P_\ell^m(\cos\theta)\exp(im\phi). \tag{18.45}$$

Using (18.33), we note that

$$Y_\ell^{-m}(\theta,\phi) = (-1)^m \left[Y_\ell^m(\theta,\phi)\right]^*,$$

where the asterisk denotes complex conjugation. The first few spherical harmonics $Y_\ell^m(\theta,\phi) \equiv Y_\ell^m$ are as follows:

$$\begin{aligned}
Y_0^0 &= \sqrt{\tfrac{1}{4\pi}}, & Y_1^0 &= \sqrt{\tfrac{3}{4\pi}}\cos\theta,\\
Y_1^{\pm1} &= \mp\sqrt{\tfrac{3}{8\pi}}\sin\theta\exp(\pm i\phi), & Y_2^0 &= \sqrt{\tfrac{5}{16\pi}}(3\cos^2\theta-1),\\
Y_2^{\pm1} &= \mp\sqrt{\tfrac{15}{8\pi}}\sin\theta\cos\theta\exp(\pm i\phi), & Y_2^{\pm2} &= \sqrt{\tfrac{15}{32\pi}}\sin^2\theta\exp(\pm 2i\phi).
\end{aligned}$$

Since they contain as their θ-dependent part the solution P_ℓ^m to the associated Legendre equation, the Y_ℓ^m are mutually orthogonal when integrated from -1 to $+1$ over $d(\cos\theta)$. Their mutual orthogonality with respect to ϕ $(0 \le \phi \le 2\pi)$ is even more obvious. The numerical factor in (18.45) is chosen to make the Y_ℓ^m an orthonormal set, i.e.

$$\int_{-1}^{1}\int_{0}^{2\pi}\left[Y_\ell^m(\theta,\phi)\right]^* Y_{\ell'}^{m'}(\theta,\phi)\,d\phi\,d(\cos\theta) = \delta_{\ell\ell'}\delta_{mm'}. \tag{18.46}$$

In addition, the spherical harmonics form a complete set in that any reasonable function (i.e. one that is likely to be met in a physical situation) of θ and ϕ can be expanded as a sum of such functions,

$$f(\theta,\phi) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell m} Y_\ell^m(\theta,\phi), \tag{18.47}$$

the constants $a_{\ell m}$ being given by

$$a_{\ell m} = \int_{-1}^{1}\int_{0}^{2\pi}\left[Y_\ell^m(\theta,\phi)\right]^* f(\theta,\phi)\,d\phi\,d(\cos\theta). \tag{18.48}$$

This is in exact analogy with a Fourier series and is a particular example of the general property of Sturm–Liouville solutions.

Aside from the orthonormality condition (18.46), the most important relationship obeyed by the Y_ℓ^m is the *spherical harmonic addition theorem.* This reads

$$P_\ell(\cos\gamma) = \frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell} Y_\ell^m(\theta,\phi)[Y_\ell^m(\theta',\phi')]^*, \tag{18.49}$$

where (θ,ϕ) and (θ',ϕ') denote two different directions in our spherical polar coordinate system that are separated by an angle γ. In general, spherical trigonometry (or vector methods) shows that these angles obey the identity

$$\cos\gamma = \cos\theta\,\cos\theta' + \sin\theta\,\sin\theta'\,\cos(\phi-\phi'). \tag{18.50}$$

►*Prove the spherical harmonic addition theorem (18.49).*

For the sake of brevity, it will be useful to denote the directions (θ,ϕ) and (θ',ϕ') by Ω and Ω', respectively. We will also denote the element of solid angle on the sphere by $d\Omega = d\phi\,d(\cos\theta)$. We begin by deriving the form of the closure relationship obeyed by the spherical harmonics. Using (18.47) and (18.48), and reversing the order of the summation and integration, we may write

$$f(\Omega) = \int_{4\pi} d\Omega'\, f(\Omega')\sum_{\ell m} Y_\ell^{m*}(\Omega')Y_\ell^m(\Omega),$$

where $\sum_{\ell m}$ is a convenient shorthand for the double summation in (18.47). Thus we may write the closure relationship for the spherical harmonics as

$$\sum_{\ell m} Y_\ell^m(\Omega)Y_\ell^{m*}(\Omega') = \delta(\Omega-\Omega'), \tag{18.51}$$

where $\delta(\Omega-\Omega')$ is a Dirac delta function with the properties that $\delta(\Omega-\Omega') = 0$ if $\Omega \neq \Omega'$ and $\int_{4\pi}\delta(\Omega)\,d\Omega = 1$.

Since $\delta(\Omega - \Omega')$ can depend only on the angle γ between the two directions Ω and Ω', we may also expand it in terms of a series of Legendre polynomials of the form

$$\delta(\Omega - \Omega') = \sum_{\ell} b_\ell P_\ell(\cos\gamma). \tag{18.52}$$

From (18.14), the coefficients in this expansion are given by

$$\begin{aligned} b_\ell &= \frac{2\ell+1}{2}\int_{-1}^{1} \delta(\Omega - \Omega')P_\ell(\cos\gamma)\, d(\cos\gamma) \\ &= \frac{2\ell+1}{4\pi}\int_0^{2\pi}\int_{-1}^{1} \delta(\Omega - \Omega')P_\ell(\cos\gamma)\, d(\cos\gamma)\, d\psi, \end{aligned}$$

where, in the second equality, we have introduced an additional integration over an azimuthal angle ψ about the direction Ω' (and γ is now the polar angle measured from Ω' to Ω). Since the rest of the integrand does not depend upon ψ, this is equivalent to multiplying it by $2\pi/2\pi$. However, the resulting double integral now has the form of a solid-angle integration over the whole sphere. Moreover, when $\Omega = \Omega'$, the angle γ separating the two directions is zero, and so $\cos\gamma = 1$. Thus, we find

$$b_\ell = \frac{2\ell+1}{4\pi} P_\ell(1) = \frac{2\ell+1}{4\pi},$$

and combining this expression with (18.51) and (18.52) gives

$$\sum_{\ell m} Y_\ell^m(\Omega) Y_\ell^{m*}(\Omega') = \sum_{\ell} \frac{2\ell+1}{4\pi} P_\ell(\cos\gamma). \tag{18.53}$$

Comparing this result with (18.49), we see that, to complete the proof of the addition theorem, we now only need to show that the summations in ℓ on either side of (18.53) can be equated *term by term*.

That such a procedure is valid may be shown by considering an arbitrary rigid rotation of the coordinate axes, thereby defining new spherical polar coordinates $\bar{\Omega}$ on the sphere. Any given spherical harmonic $Y_\ell^m(\bar{\Omega})$ in the new coordinates can be written as a linear combination of the spherical harmonics $Y_\ell^m(\Omega)$ of the old coordinates, *all* having the *same* value of ℓ. Thus,

$$Y_\ell^m(\bar{\Omega}) = \sum_{m'=-\ell}^{\ell} D_\ell^{mm'} Y_\ell^{m'}(\Omega),$$

where the coefficients $D_\ell^{mm'}$ depend on the rotation; note that in this expression Ω and $\bar{\Omega}$ refer to the same direction, but expressed in the two different coordinate systems. If we choose the polar axis of the new coordinate system to lie along the Ω' direction, then from (18.45), with m in that equation set equal to zero, we may write

$$P_\ell(\cos\gamma) = \sqrt{\frac{4\pi}{2\ell+1}}\, Y_\ell^0(\bar{\Omega}) = \sum_{m'=-\ell}^{\ell} C_\ell^{0m'} Y_\ell^{m'}(\Omega)$$

for some set of coefficients C_ℓ^{0m} that depend on Ω'. Thus, we see that the equality (18.53) does indeed hold term by term in ℓ, thus proving the addition theorem (18.49). ◄

18.4 Chebyshev functions

Chebyshev's equation has the form

$$(1 - x^2)y'' - xy' + \nu^2 y = 0, \tag{18.54}$$

and has three regular singular points, at $x = -1, 1, \infty$. By comparing it with (18.1), we see that the Chebyshev equation is very similar in form to Legendre's equation. Despite this similarity, equation (18.54) does not occur very often in physical problems, though its solutions are of considerable importance in numerical analysis. The parameter ν is a given real number, but in nearly all practical applications it takes an integer value. From here on we thus assume that $\nu = n$, where n is a non-negative integer. As was the case for Legendre's equation, in normal usage the variable x is the cosine of an angle, and so $-1 \le x \le 1$. Any solution of (18.54) is called a *Chebyshev function.*

The point $x = 0$ is an ordinary point of (18.54), and so we expect to find two linearly independent solutions of the form $y = \sum_{m=0}^{\infty} a_m x^m$. One could find the recurrence relations for the coefficients a_m in a similar manner to that used for Legendre's equation in section 18.1 (see exercise 16.15). For Chebyshev's equation, however, it is easier and more illuminating to take a different approach. In particular, we note that, on making the substitution $x = \cos\theta$, and consequently $d/dx = (-1/\sin\theta)\, d/d\theta$, Chebyshev's equation becomes (with $\nu = n$)

$$\frac{d^2y}{d\theta^2} + n^2 y = 0,$$

which is the simple harmonic equation with solutions $\cos n\theta$ and $\sin n\theta$. The corresponding linearly independent solutions of Chebyshev's equation are thus given by

$$T_n(x) = \cos(n\cos^{-1} x) \quad \text{and} \quad V_n(x) = \sin(n\cos^{-1} x). \tag{18.55}$$

It is straightforward to show that the $T_n(x)$ are *polynomials* of order n, whereas the $V_n(x)$ are *not* polynomials

►*Find explicit forms for the series expansions of $T_n(x)$ and $V_n(x)$.*

Writing $x = \cos\theta$, it is convenient first to form the complex superposition

$$\begin{aligned} T_n(x) + iV_n(x) &= \cos n\theta + i\sin n\theta \\ &= (\cos\theta + i\sin\theta)^n \\ &= \left(x + i\sqrt{1-x^2}\right)^n \qquad \text{for } |x| \le 1. \end{aligned}$$

Then, on expanding out the last expression using the binomial theorem, we obtain

$$T_n(x) = x^n - {}^nC_2 x^{n-2}(1-x^2) + {}^nC_4 x^{n-4}(1-x^2)^2 - \cdots, \tag{18.56}$$

$$V_n(x) = \sqrt{1-x^2}\left[{}^nC_1 x^{n-1} - {}^nC_3 x^{n-3}(1-x^2) + {}^nC_5 x^{n-5}(1-x^2)^2 - \cdots\right], \tag{18.57}$$

where ${}^nC_r = n!/[r!(n-r)!]$ is a binomial coefficient. We thus see that $T_n(x)$ is a polynomial of order n, but $V_n(x)$ is not a polynomial. ◄

It is conventional to define the additional functions

$$W_n(x) = (1-x^2)^{-1/2}\, T_{n+1}(x) \quad \text{and} \quad U_n(x) = (1-x^2)^{-1/2}\, V_{n+1}(x). \tag{18.58}$$

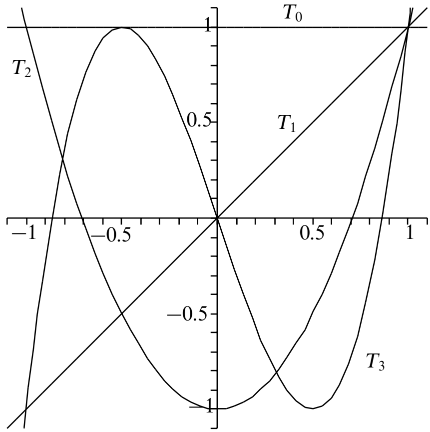


Figure 18.3 The first four Chebyshev polynomials of the first kind.

From (18.56) and (18.57), we see immediately that $U_n(x)$ is a *polynomial* of order n, but that $W_n(x)$ is *not* a polynomial. In practice, it is usual to work entirely in terms of $T_n(x)$ and $U_n(x)$, which are known, respectively, as *Chebyshev polynomials of the first and second kind*. In particular, we note that the general solution to Chebyshev's equation can be written in terms of these polynomials as

$$y(x) = \begin{cases} c_1 T_n(x) + c_2\sqrt{1-x^2}\, U_{n-1}(x) & \text{for } n = 1, 2, 3, \ldots, \\ c_1 + c_2 \sin^{-1} x & \text{for } n = 0. \end{cases}$$

The $n = 0$ solution could also be written as $d_1 + c_2 \cos^{-1} x$ with $d_1 = c_1 + \frac{1}{2}\pi c_2$.

The first few Chebyshev polynomials of the first kind are easily constructed and are given by

$$\begin{aligned} T_0(x) &= 1, & T_1(x) &= x, \\ T_2(x) &= 2x^2 - 1, & T_3(x) &= 4x^3 - 3x, \\ T_4(x) &= 8x^4 - 8x^2 + 1, & T_5(x) &= 16x^5 - 20x^3 + 5x. \end{aligned}$$

The functions $T_0(x)$, $T_1(x)$, $T_2(x)$ and $T_3(x)$ are plotted in figure 18.3. In general, the Chebyshev polynomials $T_n(x)$ satisfy $T_n(-x) = (-1)^n T_n(x)$, which is easily deduced from (18.56). Similarly, it is straightforward to deduce the following

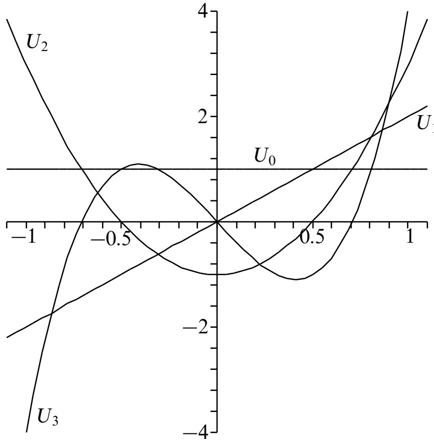


Figure 18.4 The first four Chebyshev polynomials of the second kind.

special values:

$$T_n(1) = 1, \qquad T_n(-1) = (-1)^n, \qquad T_{2n}(0) = (-1)^n, \qquad T_{2n+1}(0) = 0.$$

The first few Chebyshev polynomials of the second kind are also easily found and read

$$U_0(x) = 1, \qquad U_1(x) = 2x,$$
$$U_2(x) = 4x^2 - 1, \qquad U_3(x) = 8x^3 - 4x,$$
$$U_4(x) = 16x^4 - 12x^2 + 1, \qquad U_5(x) = 32x^5 - 32x^3 + 6x.$$

The functions $U_0(x)$, $U_1(x)$, $U_2(x)$ and $U_3(x)$ are plotted in figure 18.4. The Chebyshev polynomials $U_n(x)$ also satisfy $U_n(-x) = (-1)^n U_n(x)$, which may be deduced from (18.57) and (18.58), and have the special values:

$$U_n(1) = n + 1, \qquad U_n(-1) = (-1)^n(n + 1), \qquad U_{2n}(0) = (-1)^n, \qquad U_{2n+1}(0) = 0.$$

►*Show that the Chebyshev polynomials $U_n(x)$ satisfy the differential equation*

$$(1 - x^2)U_n''(x) - 3xU_n'(x) + n(n + 2)U_n(x) = 0. \qquad (18.59)$$

From (18.58), we have $V_{n+1} = (1 - x^2)^{1/2}U_n$ and these functions satisfy the Chebyshev equation (18.54) with $\nu = n + 1$, namely

$$(1 - x^2)V_{n+1}'' - xV_{n+1}' + (n + 1)^2 V_{n+1} = 0. \qquad (18.60)$$

Evaluating the first and second derivatives of V_{n+1}, we obtain

$$V'_{n+1} = (1-x^2)^{1/2}U'_n - x(1-x^2)^{-1/2}U_n$$
$$V''_{n+1} = (1-x^2)^{1/2}U''_n - 2x(1-x^2)^{-1/2}U'_n - (1-x^2)^{-1/2}U_n - x^2(1-x^2)^{-3/2}U_n.$$

Substituting these expressions into (18.60) and dividing through by $(1-x^2)^{1/2}$, we find

$$(1-x^2)U''_n - 3xU'_n - U_n + (n+1)^2U_n = 0,$$

which immediately simplifies to give the required result (18.59). ◄

18.4.1 Properties of Chebyshev polynomials

The Chebyshev polynomials $T_n(x)$ and $U_n(x)$ have their principal applications in numerical analysis. Their use in representing other functions over the range $|x| < 1$ plays an important role in numerical integration; Gauss–Chebyshev integration is of particular value for the accurate evaluation of integrals whose integrands contain factors $(1-x^2)^{\pm 1/2}$. It is therefore worthwhile outlining some of their main properties.

Rodrigues' formula

The Chebyshev polynomials $T_n(x)$ and $U_n(x)$ may be expressed in terms of a Rodrigues' formula, in a similar way to that used for the Legendre polynomials discussed in section 18.1.2. For the Chebyshev polynomials, we have

$$T_n(x) = \frac{(-1)^n\sqrt{\pi}(1-x^2)^{1/2}}{2^n(n-\frac{1}{2})!}\frac{d^n}{dx^n}(1-x^2)^{n-\frac{1}{2}},$$
$$U_n(x) = \frac{(-1)^n\sqrt{\pi}(n+1)}{2^{n+1}(n+\frac{1}{2})!(1-x^2)^{1/2}}\frac{d^n}{dx^n}(1-x^2)^{n+\frac{1}{2}}.$$

These Rodrigues' formulae may be proved in an analogous manner to that used in section 18.1.2 when establishing the corresponding expression for the Legendre polynomials.

Mutual orthogonality

In section 17.4, we noted that Chebyshev's equation could be put into Sturm–Liouville form with $p = (1-x^2)^{1/2}$, $q = 0$, $\lambda = n^2$ and $\rho = (1-x^2)^{-1/2}$, and its natural interval is thus $[-1, 1]$. Since the Chebyshev polynomials of the first kind, $T_n(x)$, are solutions of the Chebyshev equation and are regular at the end-points $x = \pm 1$, they must be mutually orthogonal over this interval with respect to the weight function $\rho = (1-x^2)^{-1/2}$, i.e.

$$\int_{-1}^{1} T_n(x)T_m(x)(1-x^2)^{-1/2}\,dx = 0 \qquad \text{if } n \neq m. \qquad (18.61)$$

The normalisation, when $m = n$, is easily found by making the substitution $x = \cos\theta$ and using (18.55). We immediately obtain

$$\int_{-1}^{1} T_n(x)T_n(x)(1-x^2)^{-1/2}\,dx = \begin{cases} \pi & \text{for } n = 0, \\ \pi/2 & \text{for } n = 1, 2, 3, \ldots . \end{cases} \tag{18.62}$$

The orthogonality and normalisation conditions mean that any (reasonable) function $f(x)$ can be expanded over the interval $|x| < 1$ in a series of the form

$$f(x) = \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n T_n(x),$$

where the coefficients in the expansion are given by

$$a_n = \frac{2}{\pi}\int_{-1}^{1} f(x)T_n(x)(1-x^2)^{-1/2}\,dx.$$

For the Chebyshev polynomials of the second kind, $U_n(x)$, we see from (18.58) that $(1-x^2)^{1/2}U_n(x) = V_{n+1}(x)$ satisfies Chebyshev's equation (18.54) with $v = n+1$. Thus, the orthogonality relation for the $U_n(x)$, obtained by replacing $T_i(x)$ by $V_{i+1}(x)$ in equation (18.61), reads

$$\int_{-1}^{1} U_n(x)U_m(x)(1-x^2)^{1/2}\,dx = 0 \qquad \text{if } n \neq m.$$

The corresponding normalisation condition, when $n = m$, can again be found by making the substitution $x = \cos\theta$, as illustrated in the following example.

►*Show that*

$$I \equiv \int_{-1}^{1} U_n(x)U_n(x)(1-x^2)^{1/2}\,dx = \frac{\pi}{2}.$$

From (18.58), we see that

$$I = \int_{-1}^{1} V_{n+1}(x)V_{n+1}(x)(1-x^2)^{-1/2}\,dx,$$

which, on substituting $x = \cos\theta$, gives

$$I = \int_{\pi}^{0} \sin(n+1)\theta\,\sin(n+1)\theta\,\frac{1}{\sin\theta}(-\sin\theta)\,d\theta = \frac{\pi}{2}. \blacktriangleleft$$

The above orthogonality and normalisation conditions allow one to expand any (reasonable) function in the interval $|x| < 1$ in a series of the form

$$f(x) = \sum_{n=0}^{\infty} a_n U_n(x),$$

in which the coefficients a_n are given by

$$a_n = \frac{2}{\pi}\int_{-1}^{1} f(x)U_n(x)(1-x^2)^{1/2}\,dx.$$

Generating functions

The generating functions for the Chebyshev polynomials of the first and second kinds are given, respectively, by

$$G_{\mathrm{I}}(x,h) = \frac{1-xh}{1-2xh+h^2} = \sum_{n=0}^{\infty} T_n(x)h^n, \tag{18.63}$$

$$G_{\mathrm{II}}(x,h) = \frac{1}{1-2xh+h^2} = \sum_{n=0}^{\infty} U_n(x)h^n. \tag{18.64}$$

These prescriptions may be proved in a manner similar to that used in section 18.1.2 for the generating function of the Legendre polynomials. For the Chebyshev polynomials, however, the generating functions are of less practical use, since most of the useful results can be obtained more easily by taking advantage of the trigonometric forms (18.55), as illustrated below.

Recurrence relations

There exist many useful recurrence relationships for the Chebyshev polynomials $T_n(x)$ and $U_n(x)$. They are most easily derived by setting $x = \cos\theta$ and using (18.55) and (18.58) to write

$$T_n(x) = T_n(\cos\theta) = \cos n\theta, \tag{18.65}$$

$$U_n(x) = U_n(\cos\theta) = \frac{\sin(n+1)\theta}{\sin\theta}. \tag{18.66}$$

One may then use standard formulae for the trigonometric functions to derive a wide variety of recurrence relations. Of particular use are the trigonometric identities

$$\cos(n\pm 1)\theta = \cos n\theta\,\cos\theta \mp \sin n\theta\,\sin\theta, \tag{18.67}$$

$$\sin(n\pm 1)\theta = \sin n\theta\,\cos\theta \pm \cos n\theta\,\sin\theta. \tag{18.68}$$

►*Show that the Chebyshev polynomials satisfy the recurrence relations*

$$T_{n+1}(x) - 2xT_n(x) + T_{n-1}(x) = 0, \tag{18.69}$$

$$U_{n+1}(x) - 2xU_n(x) + U_{n-1}(x) = 0. \tag{18.70}$$

Adding the result (18.67) with the plus sign to the corresponding result with a minus sign gives

$$\cos(n+1)\theta + \cos(n-1)\theta = 2\cos n\theta\,\cos\theta.$$

Using (18.65) and setting $x = \cos\theta$ immediately gives a rearrangement of the required result (18.69). Similarly, adding the plus and minus cases of result (18.68) gives

$$\sin(n+1)\theta + \sin(n-1)\theta = 2\sin n\theta\,\cos\theta.$$

Dividing through on both sides by $\sin\theta$ and using (18.66) yields (18.70). ◄

The recurrence relations (18.69) and (18.70) are extremely useful in the practical computation of Chebyshev polynomials. For example, given the values of $T_0(x)$ and $T_1(x)$ at some point x, the result (18.69) may be used iteratively to obtain the value of any $T_n(x)$ at that point; similarly, (18.70) may be used to calculate the value of any $U_n(x)$ at some point x, given the values of $U_0(x)$ and $U_1(x)$ at that point.

Further recurrence relations satisfied by the Chebyshev polynomials are

$$T_n(x) = U_n(x) - xU_{n-1}(x), \tag{18.71}$$
$$(1-x^2)U_n(x) = xT_{n+1}(x) - T_{n+2}(x), \tag{18.72}$$

which establish useful relationships between the two sets of polynomials $T_n(x)$ and $U_n(x)$. The relation (18.71) follows immediately from (18.68), whereas (18.72) follows from (18.67), with n replaced by $n+1$, on noting that $\sin^2\theta = 1 - x^2$. Additional useful results concerning the derivatives of Chebyshev polynomials may be obtained from (18.65) and (18.66), as illustrated in the following example.

►*Show that*

$$T_n'(x) = nU_{n-1}(x),$$
$$(1-x^2)U_n'(x) = xU_n(x) - (n+1)T_{n+1}(x).$$

These results are most easily derived from the expressions (18.65) and (18.66) by noting that $d/dx = (-1/\sin\theta)\,d/d\theta$. Thus,

$$T_n'(x) = -\frac{1}{\sin\theta}\frac{d(\cos n\theta)}{d\theta} = \frac{n\sin n\theta}{\sin\theta} = nU_{n-1}(x).$$

Similarly, we find

$$U_n'(x) = -\frac{1}{\sin\theta}\frac{d}{d\theta}\left[\frac{\sin(n+1)\theta}{\sin\theta}\right] = \frac{\sin(n+1)\theta\,\cos\theta}{\sin^3\theta} - \frac{(n+1)\cos(n+1)\theta}{\sin^2\theta}$$
$$= \frac{x\,U_n(x)}{1-x^2} - \frac{(n+1)T_{n+1}(x)}{1-x^2},$$

which rearranges immediately to yield the stated result. ◄

18.5 Bessel functions

Bessel's equation has the form

$$x^2y'' + xy' + (x^2 - \nu^2)y = 0, \tag{18.73}$$

which has a regular singular point at $x = 0$ and an essential singularity at $x = \infty$. The parameter ν is a given number, which we may take as ≥ 0 with no loss of generality. The equation arises from physical situations similar to those involving Legendre's equation but when cylindrical, rather than spherical, polar coordinates are employed. The variable x in Bessel's equation is usually a multiple of a radial distance and therefore ranges from 0 to ∞.

We shall seek solutions to Bessel's equation in the form of infinite series. Writing (18.73) in the standard form used in chapter 16, we have

$$y'' + \frac{1}{x}y' + \left(1 - \frac{v^2}{x^2}\right) y = 0. \tag{18.74}$$

By inspection, $x = 0$ is a regular singular point; hence we try a solution of the form $y = x^\sigma \sum_{n=0}^{\infty} a_n x^n$. Substituting this into (18.74) and multiplying the resulting equation by $x^{2-\sigma}$, we obtain

$$\sum_{n=0}^{\infty} \left[(\sigma+n)(\sigma+n-1) + (\sigma+n) - v^2\right] a_n x^n + \sum_{n=0}^{\infty} a_n x^{n+2} = 0,$$

which simplifies to

$$\sum_{n=0}^{\infty} \left[(\sigma+n)^2 - v^2\right] a_n x^n + \sum_{n=0}^{\infty} a_n x^{n+2} = 0.$$

Considering the coefficient of x^0, we obtain the indicial equation

$$\sigma^2 - v^2 = 0,$$

and so $\sigma = \pm v$. For coefficients of higher powers of x we find

$$\left[(\sigma+1)^2 - v^2\right] a_1 = 0, \tag{18.75}$$

$$\left[(\sigma+n)^2 - v^2\right] a_n + a_{n-2} = 0 \quad \text{for } n \geq 2. \tag{18.76}$$

Substituting $\sigma = \pm v$ into (18.75) and (18.76), we obtain the recurrence relations

$$(1 \pm 2v)a_1 = 0, \tag{18.77}$$

$$n(n \pm 2v)a_n + a_{n-2} = 0 \quad \text{for } n \geq 2. \tag{18.78}$$

We consider now the form of the general solution to Bessel's equation (18.73) for two cases: the case for which v is not an integer and that for which it is (including zero).

18.5.1 Bessel functions for non-integer v

If v is a non-integer then, in general, the two roots of the indicial equation, $\sigma_1 = v$ and $\sigma_2 = -v$, will not differ by an integer, and we may obtain two linearly independent solutions in the form of Frobenius series. Special considerations do arise, however, when $v = m/2$ for $m = 1, 3, 5, \ldots$, and $\sigma_1 - \sigma_2 = 2v = m$ is an (odd positive) integer. When this happens, we may always obtain a solution in the form of a Frobenius series corresponding to the larger root, $\sigma_1 = \nu = m/2$, as described above. However, for the smaller root, $\sigma_2 = -\nu = -m/2$, we must determine whether a second Frobenius series solution is possible by examining the recurrence relation (18.78), which reads

$$n(n-m)a_n + a_{n-2} = 0 \qquad \text{for } n \geq 2.$$

Since m is an *odd* positive integer in this case, we can use this recurrence relation (starting with $a_0 \neq 0$) to calculate $a_2, a_4, a_6, \ldots$ in the knowledge that all these terms will remain finite. It is possible in this case, therefore, to find a second solution in the form of a Frobenius series, one that corresponds to the smaller root σ_2.

Thus, in general, for non-integer ν we have from (18.77) and (18.78)

$$\begin{aligned} a_n &= -\frac{1}{n(n \pm 2\nu)} a_{n-2} && \text{for } n = 2, 4, 6, \ldots, \\ &= 0 && \text{for } n = 1, 3, 5, \ldots. \end{aligned}$$

Setting $a_0 = 1$ in each case, we obtain the two solutions

$$y_{\pm\nu}(x) = x^{\pm\nu}\left[1 - \frac{x^2}{2(2 \pm 2\nu)} + \frac{x^4}{2 \times 4(2 \pm 2\nu)(4 \pm 2\nu)} - \cdots\right].$$

It is customary, however, to set

$$a_0 = \frac{1}{2^{\pm\nu}\Gamma(1 \pm \nu)},$$

where $\Gamma(x)$ is the *gamma function*, described in subsection 18.12.1; it may be regarded as the generalisation of the factorial function to non-integer and/or negative arguments.[§] The two solutions of (18.73) are then written as $J_\nu(x)$ and $J_{-\nu}(x)$, where

$$\begin{aligned} J_\nu(x) &= \frac{1}{\Gamma(\nu+1)}\left(\frac{x}{2}\right)^\nu \left[1 - \frac{1}{\nu+1}\left(\frac{x}{2}\right)^2 + \frac{1}{(\nu+1)(\nu+2)}\frac{1}{2!}\left(\frac{x}{2}\right)^4 - \cdots\right] \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,\Gamma(\nu+n+1)}\left(\frac{x}{2}\right)^{\nu+2n}; \end{aligned} \tag{18.79}$$

replacing ν by $-\nu$ gives $J_{-\nu}(x)$. The functions $J_\nu(x)$ and $J_{-\nu}(x)$ are called *Bessel functions of the first kind, of order* ν. Since the first term of each series is a finite non-zero multiple of x^ν and $x^{-\nu}$, respectively, if ν is not an integer then $J_\nu(x)$ and $J_{-\nu}(x)$ are linearly independent. This may be confirmed by calculating the Wronskian of these two functions. Therefore, for non-integer ν the general solution of Bessel's equation (18.73) is given by

$$y(x) = c_1 J_\nu(x) + c_2 J_{-\nu}(x). \tag{18.80}$$

[§] In particular, $\Gamma(n+1) = n!$ for $n = 0, 1, 2, \ldots$, and $\Gamma(n)$ is infinite if n is any integer ≤ 0.

We note that Bessel functions of half-integer order are expressible in closed form in terms of trigonometric functions, as illustrated in the following example.

►*Find the general solution of*

$$x^2y'' + xy' + (x^2 - \tfrac{1}{4})y = 0.$$

This is Bessel's equation with $\nu = 1/2$, so from (18.80) the general solution is simply

$$y(x) = c_1 J_{1/2}(x) + c_2 J_{-1/2}(x).$$

However, Bessel functions of half-integral order can be expressed in terms of trigonometric functions. To show this, we note from (18.79) that

$$J_{\pm 1/2}(x) = x^{\pm 1/2} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{2^{2n\pm 1/2} n!\,\Gamma(1 + n \pm \frac{1}{2})}.$$

Using the fact that $\Gamma(x+1) = x\Gamma(x)$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, we find that, for $\nu = 1/2$,

$$\begin{aligned}
J_{1/2}(x) &= \frac{(\frac{1}{2}x)^{1/2}}{\Gamma(\frac{3}{2})} - \frac{(\frac{1}{2}x)^{5/2}}{1!\,\Gamma(\frac{5}{2})} + \frac{(\frac{1}{2}x)^{9/2}}{2!\,\Gamma(\frac{7}{2})} - \cdots \\
&= \frac{(\frac{1}{2}x)^{1/2}}{(\frac{1}{2})\sqrt{\pi}} - \frac{(\frac{1}{2}x)^{5/2}}{1!(\frac{3}{2})(\frac{1}{2})\sqrt{\pi}} + \frac{(\frac{1}{2}x)^{9/2}}{2!(\frac{5}{2})(\frac{3}{2})(\frac{1}{2})\sqrt{\pi}} - \cdots \\
&= \frac{(\frac{1}{2}x)^{1/2}}{(\frac{1}{2})\sqrt{\pi}} \left(1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \cdots\right) = \frac{(\frac{1}{2}x)^{1/2}}{(\frac{1}{2})\sqrt{\pi}} \frac{\sin x}{x} = \sqrt{\frac{2}{\pi x}} \sin x,
\end{aligned}$$

whereas for $\nu = -1/2$ we obtain

$$\begin{aligned}
J_{-1/2}(x) &= \frac{(\frac{1}{2}x)^{-1/2}}{\Gamma(\frac{1}{2})} - \frac{(\frac{1}{2}x)^{3/2}}{1!\,\Gamma(\frac{3}{2})} + \frac{(\frac{1}{2}x)^{7/2}}{2!\,\Gamma(\frac{5}{2})} - \cdots \\
&= \frac{(\frac{1}{2}x)^{-1/2}}{\sqrt{\pi}} \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots\right) = \sqrt{\frac{2}{\pi x}} \cos x.
\end{aligned}$$

Therefore the general solution we require is

$$y(x) = c_1 J_{1/2}(x) + c_2 J_{-1/2}(x) = c_1 \sqrt{\frac{2}{\pi x}} \sin x + c_2 \sqrt{\frac{2}{\pi x}} \cos x. \blacktriangleleft$$

18.5.2 Bessel functions for integer ν

The definition of the Bessel function $J_\nu(x)$ given in (18.79) is, of course, valid for all values of ν, but, as we shall see, in the case of integer ν the general solution of Bessel's equation cannot be written in the form (18.80). Firstly, let us consider the case $\nu = 0$, so that the two solutions to the indicial equation are equal, and we clearly obtain only one solution in the form of a Frobenius series. From (18.79),

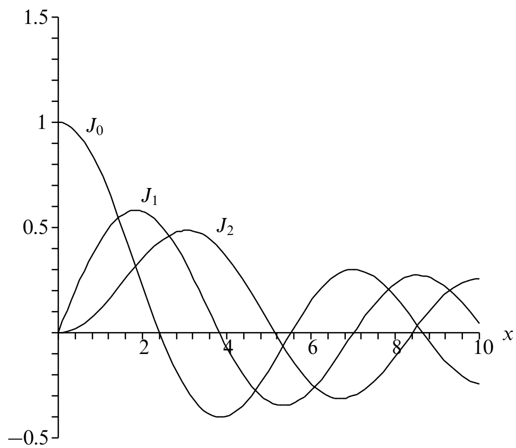


Figure 18.5 The first three integer-order Bessel functions of the first kind.

this is given by

$$J_0(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{2^{2n} n! \Gamma(1+n)}$$
$$= 1 - \frac{x^2}{2^2} + \frac{x^4}{2^2 4^2} - \frac{x^6}{2^2 4^2 6^2} + \cdots .$$

In general, however, if ν is a positive integer then the solutions of the indicial equation differ by an integer. For the larger root, $\sigma_1 = \nu$, we may find a solution $J_\nu(x)$, for $\nu = 1, 2, 3, \ldots$, in the form of the Frobenius series given by (18.79). Graphs of $J_0(x)$, $J_1(x)$ and $J_2(x)$ are plotted in figure 18.5 for real x. For the smaller root, $\sigma_2 = -\nu$, however, the recurrence relation (18.78) becomes

$$n(n-m)a_n + a_{n-2} = 0 \qquad \text{for } n \geq 2,$$

where $m = 2\nu$ is now an *even* positive integer, i.e. $m = 2, 4, 6, \ldots$. Starting with $a_0 \neq 0$ we may then calculate $a_2, a_4, a_6, \ldots$, but we see that when $n = m$ the coefficient a_n is formally infinite, and the method fails to produce a second solution in the form of a Frobenius series.

In fact, by replacing ν by $-\nu$ in the definition of $J_\nu(x)$ given in (18.79), it can be shown that, for integer ν,

$$J_{-\nu}(x) = (-1)^\nu J_\nu(x),$$

and hence that $J_\nu(x)$ and $J_{-\nu}(x)$ are linearly dependent. So, in this case, we cannot write the general solution to Bessel's equation in the form (18.80). One therefore defines the function

$$Y_\nu(x) = \frac{J_\nu(x)\cos\nu\pi - J_{-\nu}(x)}{\sin\nu\pi}, \tag{18.81}$$

which is called a Bessel function of the *second kind* of order ν (or, occasionally, a *Weber* or *Neumann* function). As Bessel's equation is linear, $Y_\nu(x)$ is clearly a solution, since it is just the weighted sum of Bessel functions of the first kind. Furthermore, for non-integer ν it is clear that $Y_\nu(x)$ is linearly independent of $J_\nu(x)$. It may also be shown that the Wronskian of $J_\nu(x)$ and $Y_\nu(x)$ is non-zero for *all* values of ν. Hence $J_\nu(x)$ and $Y_\nu(x)$ always constitute a pair of independent solutions.

►*If n is an integer, show that $Y_{n+1/2}(x) = (-1)^{n+1}J_{-n-1/2}(x)$.*

From (18.81), we have

$$Y_{n+1/2}(x) = \frac{J_{n+1/2}(x)\cos(n+\frac{1}{2})\pi - J_{-n-1/2}(x)}{\sin(n+\frac{1}{2})\pi}.$$

If n is an integer, $\cos(n+\frac{1}{2})\pi = 0$ and $\sin(n+\frac{1}{2})\pi = (-1)^n$, and so we immediately obtain $Y_{n+1/2}(x) = (-1)^{n+1}J_{-n-1/2}(x)$, as required. ◄

The expression (18.81) becomes an indeterminate form 0/0 when ν is an integer, however. This is so because for integer ν we have $\cos\nu\pi = (-1)^\nu$ and $J_{-\nu}(x) = (-1)^\nu J_\nu(x)$. Nevertheless, this indeterminate form can be evaluated using l'Hôpital's rule (see chapter 4). Therefore, for integer ν, we set

$$Y_\nu(x) = \lim_{\mu\to\nu}\left[\frac{J_\mu(x)\cos\mu\pi - J_{-\mu}(x)}{\sin\mu\pi}\right], \tag{18.82}$$

which gives a linearly independent second solution for this case. Thus, we may write the general solution of Bessel's equation, valid for *all* ν, as

$$y(x) = c_1 J_\nu(x) + c_2 Y_\nu(x). \tag{18.83}$$

The functions $Y_0(x)$, $Y_1(x)$ and $Y_2(x)$ are plotted in figure 18.6

Finally, we note that, in some applications, it is convenient to work with complex linear combinations of Bessel functions of the first and second kinds given by

$$H_\nu^{(1)}(x) = J_\nu(x) + iY_\nu(x), \qquad H_\nu^{(2)}(x) = J_\nu(x) - iY_\nu(x);$$

these are called, respectively, *Hankel functions* of the first and second kind of order ν.

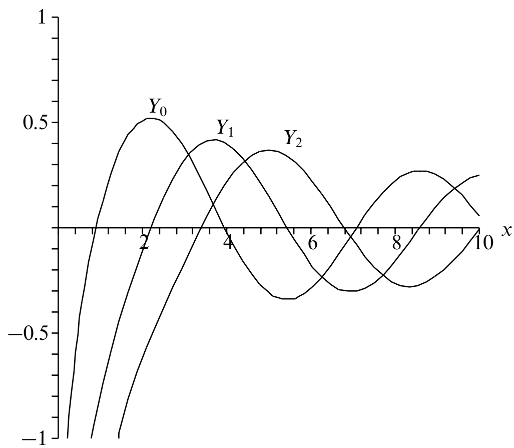


Figure 18.6 The first three integer-order Bessel functions of the second kind.

18.5.3 Properties of Bessel functions $J_\nu(x)$

In physical applications, we often require that the solution is regular at $x = 0$, but, from its definition (18.81) or (18.82), it is clear that $Y_\nu(x)$ is singular at the origin, and so in such physical situations the coefficient c_2 in (18.83) must be set to zero; the solution is then simply some multiple of $J_\nu(x)$. These Bessel functions of the first kind have various useful properties that are worthy of further discussion. Unless otherwise stated, the results presented in this section apply to Bessel functions $J_\nu(x)$ of integer and non-integer order.

Mutual orthogonality

In section 17.4, we noted that Bessel's equation (18.73) could be put into conventional Sturm–Liouville form with $p = x$, $q = -\nu^2/x$, $\lambda = \alpha^2$ and $\rho = x$, provided αx is the argument of y. From the form of p, we see that there is no natural interval over which one would expect the solutions of Bessel's equation corresponding to different eigenvalues λ (but fixed ν) to be automatically orthogonal. Nevertheless, provided the Bessel functions satisfied appropriate boundary conditions, we would expect them to obey an orthogonality relationship over some interval $[a, b]$ of the form

$$\int_a^b xJ_\nu(\alpha x)J_\nu(\beta x)\,dx = 0 \qquad \text{for } \alpha \neq \beta. \tag{18.84}$$

To determine the required boundary conditions for this result to hold, let us consider the functions $f(x) = J_\nu(\alpha x)$ and $g(x) = J_\nu(\beta x)$, which, as will be proved below, respectively satisfy the equations

$$x^2 f'' + x f' + (\alpha^2 x^2 - \nu^2) f = 0, \tag{18.85}$$

$$x^2 g'' + x g' + (\beta^2 x^2 - \nu^2) g = 0. \tag{18.86}$$

►*Show that $f(x) = J_\nu(\alpha x)$ satisfies (18.85).*

If $f(x) = J_\nu(\alpha x)$ and we write $w = \alpha x$, then

$$\frac{df}{dx} = \alpha \frac{dJ_\nu(w)}{dw} \qquad \text{and} \qquad \frac{d^2 f}{dx^2} = \alpha^2 \frac{d^2 J_\nu(w)}{dw^2}.$$

When these expressions are substituted into (18.85), its LHS becomes

$$x^2\alpha^2 \frac{d^2 J_\nu(w)}{dw^2} + x\alpha \frac{dJ_\nu(w)}{dw} + (\alpha^2 x^2 - \nu^2) J_\nu(w)$$
$$= w^2 \frac{d^2 J_\nu(w)}{dw^2} + w \frac{dJ_\nu(w)}{dw} + (w^2 - \nu^2) J_\nu(w).$$

But, from Bessel's equation itself, this final expression is equal to zero, thus verifying that $f(x)$ does satisfy (18.85). ◄

Now multiplying (18.86) by $f(x)$ and (18.85) by $g(x)$ and subtracting them gives

$$\frac{d}{dx}[x(fg' - gf')] = (\alpha^2 - \beta^2) x f g, \tag{18.87}$$

where we have used the fact that

$$\frac{d}{dx}[x(fg' - gf')] = x(fg'' - gf'') + (fg' - gf').$$

By integrating (18.87) over any given range $x = a$ to $x = b$, we obtain

$$\int_a^b x f(x) g(x)\, dx = \frac{1}{\alpha^2 - \beta^2} \Big[x f(x) g'(x) - x g(x) f'(x) \Big]_a^b,$$

which, on setting $f(x) = J_\nu(\alpha x)$ and $g(x) = J_\nu(\beta x)$, becomes

$$\int_a^b x J_\nu(\alpha x) J_\nu(\beta x)\, dx = \frac{1}{\alpha^2 - \beta^2} \Big[\beta x J_\nu(\alpha x) J_\nu'(\beta x) - \alpha x J_\nu(\beta x) J_\nu'(\alpha x) \Big]_a^b. \tag{18.88}$$

If $\alpha \neq \beta$, and the interval $[a, b]$ is such that the expression on the RHS of (18.88) equals zero, then we obtain the orthogonality condition (18.84). This happens, for example, if $J_\nu(\alpha x)$ and $J_\nu(\beta x)$ vanish at $x = a$ and $x = b$, or if $J_\nu'(\alpha x)$ and $J_\nu'(\beta x)$ vanish at $x = a$ and $x = b$, or for many more general conditions. It should be noted that the boundary term is automatically zero at the point $x = 0$, as one might expect from the fact that the Sturm–Liouville form of Bessel's equation has $p(x) = x$.

If $\alpha = \beta$, the RHS of (18.88) takes the indeterminant form 0/0. This may be evaluated using l'Hôpital's rule, or alternatively we may calculate the relevant integral directly.

►*Evaluate the integral*

$$\int_a^b J_\nu^2(\alpha x)x\,dx.$$

Ignoring the integration limits for the moment,

$$\int J_\nu^2(\alpha x)x\,dx = \frac{1}{\alpha^2}\int J_\nu^2(u)u\,du,$$

where $u = \alpha x$. Integrating by parts yields

$$I = \int J_\nu^2(u)u\,du = \tfrac{1}{2}u^2 J_\nu^2(u) - \int J_\nu(u)J_\nu'(u)u^2\,du.$$

Now Bessel's equation (18.73) can be rearranged as

$$u^2 J_\nu(u) = \nu^2 J_\nu(u) - uJ_\nu'(u) - u^2 J_\nu''(u),$$

which, on substitution into the expression for I, gives

$$\begin{aligned} I &= \tfrac{1}{2}u^2 J_\nu^2(u) - \int J_\nu'(u)[\nu^2 J_\nu(u) - uJ_\nu'(u) - u^2 J_\nu''(u)]\,du \\ &= \tfrac{1}{2}u^2 J_\nu^2(u) - \tfrac{1}{2}\nu^2 J_\nu^2(u) + \tfrac{1}{2}u^2[J_\nu'(u)]^2 + c. \end{aligned}$$

Since $u = \alpha x$, the required integral is given by

$$\int_a^b J_\nu^2(\alpha x)x\,dx = \frac{1}{2}\left[\left(x^2 - \frac{\nu^2}{\alpha^2}\right)J_\nu^2(\alpha x) + x^2[J_\nu'(\alpha x)]^2\right]_a^b, \tag{18.89}$$

which gives the normalisation condition for Bessel functions of the first kind. ◄

Since the Bessel functions $J_\nu(x)$ possess the orthogonality property (18.88), we may expand any reasonable function $f(x)$, i.e. one obeying the Dirichlet conditions discussed in chapter 12, in the interval $0 \le x \le b$ as a sum of Bessel functions of a given (non-negative) order ν,

$$f(x) = \sum_{n=0}^{\infty} c_n J_\nu(\alpha_n x), \tag{18.90}$$

provided that the α_n are chosen such that $J_\nu(\alpha_n b) = 0$. The coefficients c_n are then given by

$$c_n = \frac{2}{b^2 J_{\nu+1}^2(\alpha_n b)}\int_0^b f(x)J_\nu(\alpha_n x)x\,dx. \tag{18.91}$$

The interval is taken to be $0 \le x \le b$, as then one need only ensure that the appropriate boundary condition is satisfied at $x = b$, since the boundary condition at $x = 0$ is met automatically.

►*Prove the expression (18.91).*

If we multiply (18.90) by $xJ_\nu(\alpha_m x)$ and integrate from $x=0$ to $x=b$ then we obtain

$$\begin{aligned}\int_0^b xJ_\nu(\alpha_m x)f(x)\,dx &= \sum_{n=0}^{\infty} c_n \int_0^b xJ_\nu(\alpha_m x)J_\nu(\alpha_n x)\,dx \\ &= c_m \int_0^b J_\nu^2(\alpha_m x)x\,dx \\ &= \tfrac{1}{2}c_m b^2 J_\nu'^2(\alpha_m b) = \tfrac{1}{2}c_m b^2 J_{\nu+1}^2(\alpha_m b),\end{aligned}$$

where in the last two lines we have used (18.88) with $\alpha_m = \alpha \neq \beta = \alpha_n$, (18.89), the fact that $J_\nu(\alpha_m b) = 0$ and (18.95), which is proved below. ◄

Recurrence relations

The recurrence relations enjoyed by Bessel functions of the first kind, $J_\nu(x)$, can be derived directly from the power series definition (18.79).

►*Prove the recurrence relation*

$$\frac{d}{dx}[x^\nu J_\nu(x)] = x^\nu J_{\nu-1}(x). \tag{18.92}$$

From the power series definition (18.79) of $J_\nu(x)$ we obtain

$$\begin{aligned}\frac{d}{dx}[x^\nu J_\nu(x)] &= \frac{d}{dx}\sum_{n=0}^{\infty} \frac{(-1)^n x^{2\nu+2n}}{2^{\nu+2n} n!\Gamma(\nu+n+1)} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2\nu+2n-1}}{2^{\nu+2n-1} n!\Gamma(\nu+n)} \\ &= x^\nu \sum_{n=0}^{\infty} \frac{(-1)^n x^{(\nu-1)+2n}}{2^{(\nu-1)+2n} n!\Gamma((\nu-1)+n+1)} = x^\nu J_{\nu-1}(x). \blacktriangleleft\end{aligned}$$

It may similarly be shown that

$$\frac{d}{dx}[x^{-\nu} J_\nu(x)] = -x^{-\nu} J_{\nu+1}(x). \tag{18.93}$$

From (18.92) and (18.93) the remaining recurrence relations may be derived. Expanding out the derivative on the LHS of (18.92) and dividing through by $x^{\nu-1}$, we obtain the relation

$$xJ_\nu'(x) + \nu J_\nu(x) = xJ_{\nu-1}(x). \tag{18.94}$$

Similarly, by expanding out the derivative on the LHS of (18.93), and multiplying through by $x^{\nu+1}$, we find

$$xJ_\nu'(x) - \nu J_\nu(x) = -xJ_{\nu+1}(x). \tag{18.95}$$

Adding (18.94) and (18.95) and dividing through by x gives

$$J_{\nu-1}(x) - J_{\nu+1}(x) = 2J_\nu'(x). \tag{18.96}$$

Finally, subtracting (18.95) from (18.94) and dividing by x gives

$$J_{\nu-1}(x) + J_{\nu+1}(x) = \frac{2\nu}{x} J_\nu(x). \tag{18.97}$$

►*Given that $J_{1/2}(x) = (2/\pi x)^{1/2} \sin x$ and that $J_{-1/2}(x) = (2/\pi x)^{1/2} \cos x$, express $J_{3/2}(x)$ and $J_{-3/2}(x)$ in terms of trigonometric functions.*

From (18.95) we have

$$\begin{aligned}
J_{3/2}(x) &= \frac{1}{2x} J_{1/2}(x) - J'_{1/2}(x) \\
&= \frac{1}{2x}\left(\frac{2}{\pi x}\right)^{1/2} \sin x - \left(\frac{2}{\pi x}\right)^{1/2} \cos x + \frac{1}{2x}\left(\frac{2}{\pi x}\right)^{1/2} \sin x \\
&= \left(\frac{2}{\pi x}\right)^{1/2} \left(\frac{1}{x} \sin x - \cos x\right).
\end{aligned}$$

Similarly, from (18.94) we have

$$\begin{aligned}
J_{-3/2}(x) &= -\frac{1}{2x} J_{-1/2}(x) + J'_{-1/2}(x) \\
&= -\frac{1}{2x}\left(\frac{2}{\pi x}\right)^{1/2} \cos x - \left(\frac{2}{\pi x}\right)^{1/2} \sin x - \frac{1}{2x}\left(\frac{2}{\pi x}\right)^{1/2} \cos x \\
&= \left(\frac{2}{\pi x}\right)^{1/2} \left(-\frac{1}{x} \cos x - \sin x\right).
\end{aligned}$$

We see that, by repeated use of these recurrence relations, all Bessel functions $J_\nu(x)$ of half-integer order may be expressed in terms of trigonometric functions. From their definition (18.81), Bessel functions of the second kind, $Y_\nu(x)$, of half-integer order can be similarly expressed. ◄

Finally, we note that the relations (18.92) and (18.93) may be rewritten in integral form as

$$\int x^\nu J_{\nu-1}(x)\,dx = x^\nu J_\nu(x),$$
$$\int x^{-\nu} J_{\nu+1}(x)\,dx = -x^{-\nu} J_\nu(x).$$

If ν is an integer, the recurrence relations of this section may be proved using the generating function for Bessel functions discussed below. It may be shown that Bessel functions of the second kind, $Y_\nu(x)$, also satisfy the recurrence relations derived above.

Generating function

The Bessel functions $J_\nu(x)$, where $\nu = n$ is an integer, can be described by a generating function in a way similar to that discussed for Legendre polynomials in subsection 18.1.2. The generating function for Bessel functions of integer order is given by

$$G(x,h) = \exp\left[\frac{x}{2}\left(h - \frac{1}{h}\right)\right] = \sum_{n=-\infty}^{\infty} J_n(x)h^n. \tag{18.98}$$

By expanding the exponential as a power series, it is straightfoward to verify that the functions $J_n(x)$ defined by (18.98) are indeed Bessel functions of the first kind, as given by (18.79).

The generating function (18.98) is useful for finding, for Bessel functions of integer order, properties that can often be extended to the non-integer case. In particular, the Bessel function recurrence relations may be derived.

►*Use the generating function to prove, for integer ν, the recurrence relation (18.97), i.e.*

$$J_{\nu-1}(x) + J_{\nu+1}(x) = \frac{2\nu}{x} J_\nu(x).$$

Differentiating $G(x,h)$ with respect to h we obtain

$$\frac{\partial G(x,h)}{\partial h} = \frac{x}{2}\left(1 + \frac{1}{h^2}\right) G(x,h) = \sum_{n=-\infty}^{\infty} nJ_n(x)h^{n-1},$$

which can be written using (18.98) again as

$$\frac{x}{2}\left(1 + \frac{1}{h^2}\right) \sum_{n=-\infty}^{\infty} J_n(x)h^n = \sum_{n=-\infty}^{\infty} nJ_n(x)h^{n-1}.$$

Equating coefficients of h^n we obtain

$$\frac{x}{2}[J_n(x) + J_{n+2}(x)] = (n+1)J_{n+1}(x),$$

which, on replacing n by $\nu - 1$, gives the required recurrence relation. ◄

Integral representations

The generating function (18.98) is also useful for deriving *integral representations* of Bessel functions of integer order.

►*Show that for integer n the Bessel function $J_n(x)$ is given by*

$$J_n(x) = \frac{1}{\pi}\int_0^\pi \cos(n\theta - x\sin\theta)\,d\theta. \tag{18.99}$$

By expanding out the cosine term in the integrand in (18.99) we obtain the integral

$$I = \frac{1}{\pi}\int_0^\pi [\cos(x\sin\theta)\cos n\theta + \sin(x\sin\theta)\sin n\theta]\;d\theta. \tag{18.100}$$

Now, we may express $\cos(x\sin\theta)$ and $\sin(x\sin\theta)$ in terms of Bessel functions by setting $h = \exp i\theta$ in (18.98) to give

$$\exp\left[\frac{x}{2}(\exp i\theta - \exp(-i\theta))\right] = \exp(ix\sin\theta) = \sum_{m=-\infty}^{\infty} J_m(x)\exp im\theta.$$

Using de Moivre's theorem, $\exp i\theta = \cos\theta + i\sin\theta$, we then obtain

$$\exp(ix\sin\theta) = \cos(x\sin\theta) + i\sin(x\sin\theta) = \sum_{m=-\infty}^{\infty} J_m(x)(\cos m\theta + i\sin m\theta).$$

Equating the real and imaginary parts of this expression gives

$$\cos(x\sin\theta) = \sum_{m=-\infty}^{\infty} J_m(x)\cos m\theta,$$

$$\sin(x\sin\theta) = \sum_{m=-\infty}^{\infty} J_m(x)\sin m\theta.$$

Substituting these expressions into (18.100) then yields

$$I = \frac{1}{\pi}\sum_{m=-\infty}^{\infty}\int_0^{\pi} [J_m(x)\cos m\theta\cos n\theta + J_m(x)\sin m\theta\sin n\theta]\, d\theta.$$

However, using the orthogonality of the trigonometric functions [see equations (12.1)–(12.3)], we obtain

$$I = \frac{1}{\pi}\frac{\pi}{2}[J_n(x) + J_n(x)] = J_n(x),$$

which proves the integral representation (18.99). ◀

Finally, we mention the special case of the integral representation (18.99) for $n = 0$:

$$J_0(x) = \frac{1}{\pi}\int_0^{\pi}\cos(x\sin\theta)\, d\theta = \frac{1}{2\pi}\int_0^{2\pi}\cos(x\sin\theta)\, d\theta,$$

since $\cos(x\sin\theta)$ repeats itself in the range $\theta = \pi$ to $\theta = 2\pi$. However, $\sin(x\sin\theta)$ changes sign in this range and so

$$\frac{1}{2\pi}\int_0^{2\pi}\sin(x\sin\theta)\, d\theta = 0.$$

Using de Moivre's theorem, we can therefore write

$$J_0(x) = \frac{1}{2\pi}\int_0^{2\pi}\exp(ix\sin\theta)\, d\theta = \frac{1}{2\pi}\int_0^{2\pi}\exp(ix\cos\theta)\, d\theta.$$

There are in fact many other integral representations of Bessel functions; they can be derived from those given.

18.6 Spherical Bessel functions

When obtaining solutions of Helmholtz' equation $(\nabla^2 + k^2)u = 0$ in spherical polar coordinates (see section 21.3.2), one finds that, for solutions that are finite on the polar axis, the radial part $R(r)$ of the solution must satisfy the equation

$$r^2R'' + 2rR' + [k^2r^2 - \ell(\ell+1)]R = 0, \tag{18.101}$$

where ℓ is an integer. This equation looks very much like Bessel's equation and can in fact be reduced to it by writing $R(r) = r^{-1/2}S(r)$, in which case $S(r)$ then satisfies

$$r^2S'' + rS' + \left[k^2r^2 - \left(\ell + \tfrac{1}{2}\right)^2\right] S = 0.$$

On making the change of variable $x = kr$ and letting $y(x) = S(kr)$, we obtain

$$x^2y'' + xy' + [x^2 - (\ell + \tfrac{1}{2})^2]y = 0,$$

where the primes now denote d/dx. This is Bessel's equation of order $\ell + \frac{1}{2}$ and has as its solutions $y(x) = J_{\ell+1/2}(x)$ and $Y_{\ell+1/2}(x)$. The general solution of (18.101) can therefore be written

$$R(r) = r^{-1/2}[c_1J_{\ell+1/2}(kr) + c_2Y_{\ell+1/2}(kr)],$$

where c_1 and c_2 are constants that may be determined from the boundary conditions on the solution. In particular, for solutions that are finite at the origin we require $c_2 = 0$.

The functions $x^{-1/2}J_{\ell+1/2}(x)$ and $x^{-1/2}Y_{\ell+1/2}(x)$, when suitably normalised, are called *spherical Bessel functions* of the first and second kind, respectively, and are denoted as follows:

$$j_\ell(x) = \sqrt{\frac{\pi}{2x}}J_{\ell+1/2}(x), \qquad (18.102)$$

$$n_\ell(x) = \sqrt{\frac{\pi}{2x}}Y_{\ell+1/2}(x). \qquad (18.103)$$

For integer ℓ, we also note that $Y_{\ell+1/2}(x) = (-1)^{\ell+1}J_{-\ell-1/2}(x)$, as discussed in section 18.5.2. Moreover, in section 18.5.1, we noted that Bessel functions of the first kind, $J_\nu(x)$, of half-integer order are expressible in closed form in terms of trigonometric functions. Thus, all spherical Bessel functions of both the first and second kinds may be expressed in such a form. In particular, using the results of the worked example in section 18.5.1, we find that

$$j_0(x) = \frac{\sin x}{x}, \qquad (18.104)$$

$$n_0(x) = -\frac{\cos x}{x}. \qquad (18.105)$$

Expressions for higher-order spherical Bessel functions are most easily obtained by using the recurrence relations for Bessel functions.

►*Show that the ℓth spherical Bessel function is given by*

$$f_\ell(x) = (-1)^\ell x^\ell \left(\frac{1}{x}\frac{d}{dx}\right)^\ell f_0(x), \tag{18.106}$$

where $f_\ell(x)$ denotes either $j_\ell(x)$ or $n_\ell(x)$.

The recurrence relation (18.93) for Bessel functions of the first kind reads

$$J_{\nu+1}(x) = -x^\nu \frac{d}{dx}\left[x^{-\nu} J_\nu(x)\right].$$

Thus, on setting $\nu = \ell + \frac{1}{2}$ and rearranging, we find

$$x^{-1/2} J_{\ell+3/2}(x) = -x^\ell \frac{d}{dx}\left[\frac{x^{-1/2} J_{\ell+1/2}}{x^\ell}\right],$$

which on using (18.102) yields the recurrence relation

$$j_{\ell+1}(x) = -x^\ell \frac{d}{dx}[x^{-\ell} j_\ell(x)].$$

We now change $\ell + 1 \to \ell$ and iterate this result:

$$\begin{aligned} j_\ell(x) &= -x^{\ell-1}\frac{d}{dx}[x^{-\ell+1} j_{\ell-1}(x)] \\ &= -x^{\ell-1}\frac{d}{dx}\left\{x^{-\ell+1}(-1)x^{\ell-2}\frac{d}{dx}\left[x^{-\ell+2} j_{\ell-2}(x)\right]\right\} \\ &= (-1)^2\frac{x^\ell}{x}\frac{d}{dx}\left\{\frac{1}{x}\frac{d}{dx}\left[x^{-\ell+2} j_{\ell-2}(x)\right]\right\} \\ &= \cdots \\ &= (-1)^\ell x^\ell \left(\frac{1}{x}\frac{d}{dx}\right)^\ell j_0(x). \end{aligned}$$

This is the expression for $j_\ell(x)$ as given in (18.106). One may prove the result (18.106) for $n_\ell(x)$ in an analogous manner by setting $\nu = \ell - \frac{1}{2}$ in the recurrence relation (18.92) for Bessel functions of the first kind and using the relationship $Y_{\ell+1/2}(x) = (-1)^{\ell+1} J_{-\ell-1/2}(x)$. ◄

Using result (18.106) and the expressions (18.104) and (18.105), one quickly finds, for example,

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x}, \qquad j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x}\right)\sin x - \frac{3\cos x}{x^2},$$

$$n_1(x) = -\frac{\cos x}{x^2} - \frac{\sin x}{x}, \qquad n_2(x) = -\left(\frac{3}{x^3} - \frac{1}{x}\right)\cos x - \frac{3\sin x}{x^2}.$$

Finally, we note that the orthogonality properties of the spherical Bessel functions follow directly from the orthogonality condition (18.88) for Bessel functions of the first kind.

18.7 Laguerre functions

Laguerre's equation has the form

$$xy'' + (1 - x)y' + \nu y = 0; \tag{18.107}$$

it has a regular singularity at $x = 0$ and an essential singularity at $x = \infty$. The parameter ν is a given real number, although it nearly always takes an integer value in physical applications. The Laguerre equation appears in the description of the wavefunction of the hydrogen atom. Any solution of (18.107) is called a *Laguerre function.*

Since the point $x = 0$ is a regular singularity, we may find at least one solution in the form of a Frobenius series (see section 16.3):

$$y(x) = \sum_{m=0}^{\infty} a_m x^{m+\sigma}. \tag{18.108}$$

Substituting this series into (18.107) and dividing through by $x^{\sigma-1}$, we obtain

$$\sum_{m=0}^{\infty} [(m+\sigma)(m+\sigma-1) + (1-x)(m+\sigma) + \nu x]\, a_m x^m = 0. \tag{18.109}$$

Setting $x = 0$, so that only the $m = 0$ term remains, we obtain the indicial equation $\sigma^2 = 0$, which trivially has $\sigma = 0$ as its repeated root. Thus, Laguerre's equation has only one solution of the form (18.108), and it, in fact, reduces to a simple power series. Substituting $\sigma = 0$ into (18.109) and demanding that the coefficient of x^{m+1} vanishes, we obtain the recurrence relation

$$a_{m+1} = \frac{m-\nu}{(m+1)^2} a_m.$$

As mentioned above, in nearly all physical applications, the parameter ν takes integer values. Therefore, if $\nu = n$, where n is a non-negative integer, we see that $a_{n+1} = a_{n+2} = \cdots = 0$, and so our solution to Laguerre's equation is a polynomial of order n. It is conventional to choose $a_0 = 1$, so that the solution is given by

$$L_n(x) = \frac{(-1)^n}{n!}\left[x^n - \frac{n^2}{1!}x^{n-1} + \frac{n^2(n-1)^2}{2!}x^{n-2} - \cdots + (-1)^n n!\right] \tag{18.110}$$

$$= \sum_{m=0}^{n} (-1)^m \frac{n!}{(m!)^2(n-m)!} x^m, \tag{18.111}$$

where $L_n(x)$ is called the nth *Laguerre polynomial.* We note in particular that $L_n(0) = 1$. The first few Laguerre polynomials are given by

$$\begin{aligned}
L_0(x) &= 1, & 3!L_3(x) &= -x^3 + 9x^2 - 18x + 6,\\
L_1(x) &= -x + 1, & 4!L_4(x) &= x^4 - 16x^3 + 72x^2 - 96x + 24,\\
2!L_2(x) &= x^2 - 4x + 2, & 5!L_5(x) &= -x^5 + 25x^4 - 200x^3 + 600x^2 - 600x + 120.
\end{aligned}$$

The functions $L_0(x)$, $L_1(x)$, $L_2(x)$ and $L_3(x)$ are plotted in figure 18.7.

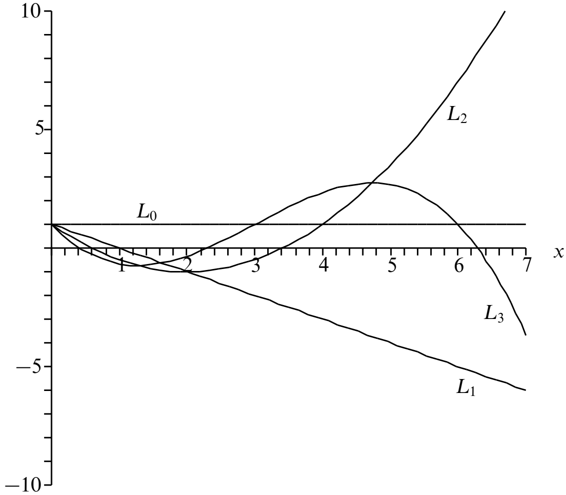


Figure 18.7 The first four Laguerre polynomials.

18.7.1 Properties of Laguerre polynomials

The Laguerre polynomials and functions derived from them are important in the analysis of the quantum mechanical behaviour of some physical systems. We therefore briefly outline their useful properties in this section.

Rodrigues' formula

The Laguerre polynomials can be expressed in terms of a Rodrigues' formula given by

$$L_n(x) = \frac{e^x}{n!}\frac{d^n}{dx^n}\left(x^n e^{-x}\right), \tag{18.112}$$

which may be proved straightforwardly by calculating the nth derivative explicitly using Leibnitz' theorem and comparing the result with (18.111). This is illustrated in the following example.

►*Prove that the expression (18.112) yields the nth Laguerre polynomial.*

Evaluating the nth derivative in (18.112) using Leibnitz' theorem, we find

$$\begin{aligned}
L_n(x) &= \frac{e^x}{n!}\sum_{r=0}^{n} {}^nC_r \frac{d^r x^n}{dx^r}\frac{d^{n-r}e^{-x}}{dx^{n-r}} \\
&= \frac{e^x}{n!}\sum_{r=0}^{n} \frac{n!}{r!(n-r)!}\frac{n!}{(n-r)!}x^{n-r}(-1)^{n-r}e^{-x} \\
&= \sum_{r=0}^{n}(-1)^{n-r}\frac{n!}{r!(n-r)!(n-r)!}x^{n-r}.
\end{aligned}$$

Relabelling the summation using the index $m = n - r$, we obtain

$$L_n(x) = \sum_{m=0}^{n}(-1)^m \frac{n!}{(m!)^2(n-m)!}x^m,$$

which is precisely the expression (18.111) for the nth Laguerre polynomial. ◄

Mutual orthogonality

In section 17.4, we noted that Laguerre's equation could be put into Sturm–Liouville form with $p = xe^{-x}$, $q = 0$, $\lambda = \nu$ and $\rho = e^{-x}$, and its natural interval is thus $[0, \infty]$. Since the Laguerre polynomials $L_n(x)$ are solutions of the equation and are regular at the end-points, they must be mutually orthogonal over this interval with respect to the weight function $\rho = e^{-x}$, i.e.

$$\int_0^\infty L_n(x)L_k(x)e^{-x}\,dx = 0 \qquad \text{if } n \neq k.$$

This result may also be proved directly using the Rodrigues' formula (18.112). Indeed, the normalisation, when $k = n$, is most easily found using this method.

►*Show that*

$$I \equiv \int_0^\infty L_n(x)L_n(x)e^{-x}\,dx = 1. \tag{18.113}$$

Using the Rodrigues' formula (18.112), we may write

$$I = \frac{1}{n!}\int_0^\infty L_n(x)\frac{d^n}{dx^n}(x^n e^{-x})\,dx = \frac{(-1)^n}{n!}\int_0^\infty \frac{d^n L_n}{dx^n}x^n e^{-x}\,dx,$$

where, in the second equality, we have integrated by parts n times and used the fact that the boundary terms all vanish. When d^nL_n/dx^n is evaluated using (18.111), only the derivative of the $m = n$ term survives and that has the value $[(-1)^n n!\,n!]/[(n!)^2\,0!] = (-1)^n$. Thus we have

$$I = \frac{1}{n!}\int_0^\infty x^n e^{-x}\,dx = 1,$$

where, in the second equality, we use the expression (18.153) defining the gamma function (see section 18.12). ◄

The above orthogonality and normalisation conditions allow us to expand any (reasonable) function in the interval $0 \le x < \infty$ in a series of the form

$$f(x) = \sum_{n=0}^{\infty} a_n L_n(x),$$

in which the coefficients a_n are given by

$$a_n = \int_0^{\infty} f(x) L_n(x) e^{-x}\, dx.$$

We note that it is sometimes convenient to define the *orthonormal Laguerre functions* $\phi_n(x) = e^{-x/2} L_n(x)$, which may also be used to produce a series expansion of a function in the interval $0 \le x < \infty$.

Generating function

The generating function for the Laguerre polynomials is given by

$$G(x,h) = \frac{e^{-xh/(1-h)}}{1-h} = \sum_{n=0}^{\infty} L_n(x) h^n. \tag{18.114}$$

We may prove this result by differentiating the generating function with respect to x and h, respectively, to obtain recurrence relations for the Laguerre polynomials, which may then be combined to show that the functions $L_n(x)$ in (18.114) do indeed satisfy Laguerre's equation (as discussed in the next subsection).

Recurrence relations

The Laguerre polynomials obey a number of useful recurrence relations. The three most important relations are as follows:

$$(n+1)L_{n+1}(x) = (2n+1-x)L_n(x) - nL_{n-1}(x), \tag{18.115}$$

$$L_{n-1}(x) = L'_{n-1}(x) - L'_n(x), \tag{18.116}$$

$$xL'_n(x) = nL_n(x) - nL_{n-1}(x). \tag{18.117}$$

The first two relations are easily derived from the generating function (18.114), and may be combined straightforwardly to yield the third result.

►*Derive the recurrence relations (18.115) and (18.116).*

Differentiating the generating function (18.114) with respect to h, we find

$$\frac{\partial G}{\partial h} = \frac{(1-x-h)e^{-xh/(1-h)}}{(1-h)^3} = \sum nL_n h^{n-1}.$$

Thus, we may write

$$(1-x-h)\sum L_n h^n = (1-h)^2 \sum nL_n h^{n-1},$$

and, on equating coefficients of h^n on each side, we obtain

$$(1-x)L_n - L_{n-1} = (n+1)L_{n+1} - 2nL_n + (n-1)L_{n-1},$$

which trivially rearranges to give the recurrence relation (18.115).

To obtain the recurrence relation (18.116), we begin by differentiating the generating function (18.114) with respect to x, which yields

$$\frac{\partial G}{\partial x} = -\frac{he^{-xh/(1-h)}}{(1-h)^2} = \sum L_n' h^n,$$

and thus we have

$$-h \sum L_n h^n = (1-h) \sum L_n' h^n.$$

Equating coefficients of h^n on each side then gives

$$-L_{n-1} = L_n' - L_{n-1}',$$

which immediately simplifies to give (18.116). ◄

18.8 Associated Laguerre functions

The associated Laguerre equation has the form

$$xy'' + (m+1-x)y' + ny = 0; \tag{18.118}$$

it has a regular singularity at $x = 0$ and an essential singularity at $x = \infty$. We restrict our attention to the situation in which the parameters n and m are both non-negative integers, as is the case in nearly all physical problems. The associated Laguerre equation occurs most frequently in quantum-mechanical applications. Any solution of (18.118) is called an *associated Laguerre function*.

Solutions of (18.118) for non-negative integers n and m are given by the *associated Laguerre polynomials*

$$L_n^m(x) = (-1)^m \frac{d^m}{dx^m} L_{n+m}(x), \tag{18.119}$$

where $L_n(x)$ are the ordinary Laguerre polynomials.[§]

►*Show that the functions $L_n^m(x)$ defined in (18.119) are solutions of (18.118).*

Since the Laguerre polynomials $L_n(x)$ are solutions of Laguerre's equation (18.107), we have

$$xL_{n+m}'' + (1-x)L_{n+m}' + (n+m)L_{n+m} = 0.$$

Differentiating this equation m times using Leibnitz' theorem and rearranging, we find

$$xL_{n+m}^{(m+2)} + (m+1-x)L_{n+m}^{(m+1)} + nL_{n+m}^{(m)} = 0.$$

On multiplying through by $(-1)^m$ and setting $L_n^m = (-1)^m L_{n+m}^{(m)}$, in accord with (18.119), we obtain

$$x(L_n^m)'' + (m+1-x)(L_n^m)' + nL_n^m = 0,$$

which shows that the functions L_n^m are indeed solutions of (18.118). ◄

[§] Note that some authors define the associated Laguerre polynomials as $\mathcal{L}_n^m(x) = (d^m/dx^m)L_n(x)$, which is thus related to our expression (18.119) by $L_n^m(x) = (-1)^m \mathcal{L}_{n+m}^m(x)$.

In particular, we note that $L_n^0(x) = L_n(x)$. As discussed in the previous section, $L_n(x)$ is a polynomial of order n and so it follows that $L_n^m(x)$ is also. The first few associated Laguerre polynomials are easily found using (18.119):

$$\begin{aligned}
L_0^m(x) &= 1,\\
L_1^m(x) &= -x + m + 1,\\
2!L_2^m(x) &= x^2 - 2(m+2)x + (m+1)(m+2),\\
3!L_3^m(x) &= -x^3 + 3(m+3)x^2 - 3(m+2)(m+3)x + (m+1)(m+2)(m+3).
\end{aligned}$$

Indeed, in the general case, one may show straightforwardly, from the definition (18.119) and the expression (18.111) for the ordinary Laguerre polynomials, that

$$L_n^m(x) = \sum_{k=0}^{n} (-1)^k \frac{(n+m)!}{k!(n-k)!(k+m)!} x^k. \tag{18.120}$$

18.8.1 Properties of associated Laguerre polynomials

The properties of the associated Laguerre polynomials follow directly from those of the ordinary Laguerre polynomials through the definition (18.119). We shall therefore only briefly outline the most useful results here.

Rodrigues' formula

A Rodrigues' formula for the associated Laguerre polynomials is given by

$$L_n^m(x) = \frac{e^x x^{-m}}{n!} \frac{d^n}{dx^n}(x^{n+m}e^{-x}). \tag{18.121}$$

It can be proved by evaluating the nth derivative using Leibnitz' theorem (see exercise 18.7).

Mutual orthogonality

In section 17.4, we noted that the associated Laguerre equation could be transformed into a Sturm–Liouville one with $p = x^{m+1}e^{-x}$, $q = 0$, $\lambda = n$ and $\rho = x^m e^{-x}$, and its natural interval is thus $[0, \infty]$. Since the associated Laguerre polynomials $L_n^m(x)$ are solutions of the equation and are regular at the end-points, those with the same m but differing values of the eigenvalue $\lambda = n$ must be mutually orthogonal over this interval with respect to the weight function $\rho = x^m e^{-x}$, i.e.

$$\int_0^\infty L_n^m(x) L_k^m(x) x^m e^{-x}\, dx = 0 \qquad \text{if } n \neq k.$$

This result may also be proved directly using the Rodrigues' formula (18.121), as may the normalisation condition when $k = n$.

►*Show that*

$$I \equiv \int_0^\infty L_n^m(x) L_n^m(x) x^m e^{-x}\, dx = \frac{(n+m)!}{n!}. \tag{18.122}$$

Using the Rodrigues' formula (18.121), we may write

$$I = \frac{1}{n!}\int_0^\infty L_n^m(x)\,\frac{d^n}{dx^n}(x^{n+m}e^{-x})\,dx = \frac{(-1)^n}{n!}\int_0^\infty \frac{d^n L_n^m}{dx^n}\,x^{n+m}e^{-x}\,dx,$$

where, in the second equality, we have integrated by parts n times and used the fact that the boundary terms all vanish. From (18.120) we see that $d^n L_n^m/dx^n = (-1)^n$. Thus we have

$$I = \frac{1}{n!}\int_0^\infty x^{n+m}e^{-x}\,dx = \frac{(n+m)!}{n!},$$

where, in the second equality, we use the expression (18.153) defining the gamma function (see section 18.12). ◄

The above orthogonality and normalisation conditions allow us to expand any (reasonable) function in the interval $0 \le x < \infty$ in a series of the form

$$f(x) = \sum_{n=0}^\infty a_n L_n^m(x),$$

in which the coefficients a_n are given by

$$a_n = \frac{n!}{(n+m)!}\int_0^\infty f(x) L_n^m(x) x^m e^{-x}\,dx.$$

We note that it is sometimes convenient to define the *orthogonal associated Laguerre functions* $\phi_n^m(x) = x^{m/2}e^{-x/2}L_n^m(x)$, which may also be used to produce a series expansion of a function in the interval $0 \le x < \infty$.

Generating function

The generating function for the associated Laguerre polynomials is given by

$$G(x,h) = \frac{e^{-xh/(1-h)}}{(1-h)^{m+1}} = \sum_{n=0}^\infty L_n^m(x)h^n. \tag{18.123}$$

This can be obtained by differentiating the generating function (18.114) for the ordinary Laguerre polynomials m times with respect to x, and using (18.119).

►*Use the generating function (18.123) to obtain an expression for $L_n^m(0)$.*

From (18.123), we have

$$\begin{aligned}\sum_{n=0}^\infty L_n^m(0)h^n &= \frac{1}{(1-h)^{m+1}}\\ &= 1 + (m+1)h + \frac{(m+1)(m+2)}{2!}h^2 + \cdots + \frac{(m+1)(m+2)\cdots(m+n)}{n!}h^n + \cdots,\end{aligned}$$

where, in the second equality, we have expanded the RHS using the binomial theorem. On equating coefficients of h^n, we immediately obtain

$$L_n^m(0) = \frac{(n+m)!}{n!m!}. \blacktriangleleft$$

Recurrence relations

The various recurrence relations satisfied by the associated Laguerre polynomials may be derived by differentiating the generating function (18.123) with respect to either or both of x and h, or by differentiating with respect to x the recurrence relations obeyed by the ordinary Laguerre polynomials, discussed in section 18.7.1. Of the many recurrence relations satisfied by the associated Laguerre polynomials, two of the most useful are as follows:

$$(n+1)L_{n+1}^m(x) = (2n+m+1-x)L_n^m(x) - (n+m)L_{n-1}^m(x), \qquad (18.124)$$

$$x(L_n^m)'(x) = nL_n^m(x) - (n+m)L_{n-1}^m(x). \qquad (18.125)$$

For proofs of these relations the reader is referred to exercise 18.7.

18.9 Hermite functions

Hermite's equation has the form

$$y'' - 2xy' + 2\nu y = 0, \qquad (18.126)$$

and has an essential singularity at $x = \infty$. The parameter ν is a given real number, although it nearly always takes an integer value in physical applications. The Hermite equation appears in the description of the wavefunction of the harmonic oscillator. Any solution of (18.126) is called a *Hermite function*.

Since $x = 0$ is an ordinary point of the equation, we may find two linearly independent solutions in the form of a power series (see section 16.2):

$$y(x) = \sum_{m=0}^{\infty} a_m x^m. \qquad (18.127)$$

Substituting this series into (18.107) yields

$$\sum_{m=0}^{\infty} \left[(m+1)(m+2)a_{m+2} + 2(\nu - m)a_m\right] x^m = 0.$$

Demanding that the coefficient of each power of x vanishes, we obtain the recurrence relation

$$a_{m+2} = -\frac{2(\nu - m)}{(m+1)(m+2)}a_m.$$

As mentioned above, in nearly all physical applications, the parameter ν takes integer values. Therefore, if $\nu = n$, where n is a non-negative integer, we see that

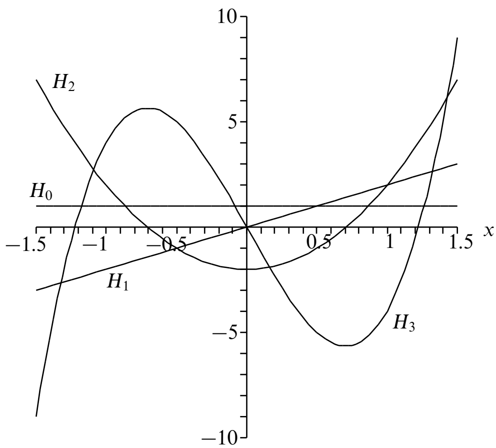


Figure 18.8 The first four Hermite polynomials.

$a_{n+2} = a_{n+4} = \cdots = 0$, and so one solution of Hermite's equation is a polynomial of order n. For even n, it is conventional to choose $a_0 = (-1)^{n/2} n!/(n/2)!$, whereas for odd n one takes $a_1 = (-1)^{(n-1)/2} 2n!/[\frac{1}{2}(n-1)]!$. These choices allow a general solution to be written as

$$H_n(x) = (2x)^n - n(n-1)(2x)^{n-1} + \frac{n(n-1)(n-2)(n-3)}{2!}(2x)^{n-4} - \cdots \quad (18.128)$$

$$= \sum_{m=0}^{[n/2]} (-1)^m \frac{n!}{m!(n-2m)!}(2x)^{n-2m}, \quad (18.129)$$

where $H_n(x)$ is called the nth *Hermite polynomial* and the notation $[n/2]$ denotes the integer part of $n/2$. We note in particular that $H_n(-x) = (-1)^n H_n(x)$. The first few Hermite polynomials are given by

$$H_0(x) = 1, \qquad H_3(x) = 8x^2 - 12x,$$
$$H_1(x) = 2x, \qquad H_4(x) = 16x^4 - 48x^2 + 12,$$
$$H_2(x) = 4x^2 - 2, \qquad H_5(x) = 32x^5 - 160x^3 + 120x.$$

The functions $H_0(x)$, $H_1(x)$, $H_2(x)$ and $H_3(x)$ are plotted in figure 18.8.

18.9.1 Properties of Hermite polynomials

The Hermite polynomials and functions derived from them are important in the analysis of the quantum mechanical behaviour of some physical systems. We therefore briefly outline their useful properties in this section.

Rodrigues' formula

The Rodrigues' formula for the Hermite polynomials is given by

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n}(e^{-x^2}). \tag{18.130}$$

This can be proved using Leibnitz' theorem.

►*Prove the Rodrigues' formula (18.130) for the Hermite polynomials.*

Letting $u = e^{-x^2}$ and differentiating with respect to x, we quickly find that

$$u' + 2xu = 0.$$

Differentiating this equation $n+1$ times using Leibnitz' theorem then gives

$$u^{(n+2)} + 2xu^{(n+1)} + 2(n+1)u^{(n)} = 0,$$

which, on introducing the new variable $v = (-1)^n u^{(n)}$, reduces to

$$v'' + 2xv' + 2(n+1)v = 0. \tag{18.131}$$

Now letting $y = e^{x^2} v$, we may write the derivatives of v as

$$v' = e^{-x^2}(y' - 2xy),$$
$$v'' = e^{-x^2}(y'' - 4xy' + 4x^2 y - 2y).$$

Substituting these expressions into (18.131), and dividing through by e^{-x^2}, finally yields Hermite's equation,

$$y'' - 2xy + 2ny = 0,$$

thus demonstrating that $y = (-1)^n e^{x^2} d^n(e^{-x^2})/dx^n$ is indeed a solution. Moreover, since this solution is clearly a polynomial of order n, it must be some multiple of $H_n(x)$. The normalisation is easily checked by noting that, from (18.130), the highest-order term is $(2x)^n$, which agrees with the expression (18.128). ◄

Mutual orthogonality

We saw in section 17.4 that Hermite's equation could be cast in Sturm–Liouville form with $p = e^{-x^2}$, $q = 0$, $\lambda = 2n$ and $\rho = e^{-x^2}$, and its natural interval is thus $[-\infty, \infty]$. Since the Hermite polynomials $H_n(x)$ are solutions of the equation and are regular at the end-points, they must be mutually orthogonal over this interval with respect to the weight function $\rho = e^{-x^2}$, i.e.

$$\int_{-\infty}^{\infty} H_n(x)H_k(x)e^{-x^2}\, dx = 0 \qquad \text{if } n \neq k.$$

This result may also be proved directly using the Rodrigues' formula (18.130). Indeed, the normalisation, when $k = n$, is most easily found in this way.

►*Show that*

$$I \equiv \int_{-\infty}^{\infty} H_n(x)H_n(x)e^{-x^2}\,dx = 2^n n!\sqrt{\pi}. \qquad (18.132)$$

Using the Rodrigues' formula (18.130), we may write

$$I = (-1)^n \int_0^{\infty} H_n(x)\,\frac{d^n}{dx^n}(e^{-x^2})\,dx = \int_{-\infty}^{\infty} \frac{d^n H_n}{dx^n}\,e^{-x^2}\,dx,$$

where, in the second equality, we have integrated by parts n times and used the fact that the boundary terms all vanish. From (18.128) we see that $d^nH_n/dx^n = 2^n n!$. Thus we have

$$I = 2^n n! \int_{-\infty}^{\infty} e^{-x^2}\,dx = 2^n n!\sqrt{\pi},$$

where, in the second equality, we use the standard result for the area under a Gaussian (see section 6.4.2). ◄

The above orthogonality and normalisation conditions allow any (reasonable) function in the interval $-\infty \leq x < \infty$ to be expanded in a series of the form

$$f(x) = \sum_{n=0}^{\infty} a_n H_n(x),$$

in which the coefficients a_n are given by

$$a_n = \frac{1}{2^n n!\sqrt{\pi}} \int_{-\infty}^{\infty} f(x)H_n(x)e^{-x^2}\,dx.$$

We note that it is sometimes convenient to define the *orthogonal Hermite functions* $\phi_n(x) = e^{-x^2/2}H_n(x)$; they also may be used to produce a series expansion of a function in the interval $-\infty \leq x < \infty$. Indeed, $\phi_n(x)$ is proportional to the wavefunction of a particle in the nth energy level of a quantum harmonic oscillator.

Generating function

The generating function equation for the Hermite polynomials reads

$$G(x,h) = e^{2hx-h^2} = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!}h^n, \qquad (18.133)$$

a result that may be proved using the Rodrigues' formula (18.130).

►*Show that the functions $H_n(x)$ in (18.133) are the Hermite polynomials.*

It is often more convenient to write the generating function (18.133) as

$$G(x,h) = e^{x^2}e^{-(x-h)^2} = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!}h^n.$$

Differentiating this form k times with respect to h gives

$$\sum_{n=k}^{\infty} \frac{H_n}{(n-k)!} h^{n-k} = \frac{\partial^k G}{\partial h^k} = e^{x^2} \frac{\partial^k}{\partial h^k} e^{-(x-h)^2} = (-1)^k e^{x^2} \frac{\partial^k}{\partial x^k} e^{-(x-h)^2}.$$

Relabelling the summation on the LHS using the new index $m = n - k$, we obtain

$$\sum_{m=0}^{\infty} \frac{H_{m+k}}{m!} h^m = (-1)^k e^{x^2} \frac{\partial^k}{\partial x^k} e^{-(x-h)^2}.$$

Setting $h = 0$ in this equation, we find

$$H_k(x) = (-1)^k e^{x^2} \frac{d^k}{dx^k} (e^{-x^2}),$$

which is the Rodrigues' formula (18.130) for the Hermite polynomials. ◀

The generating function (18.133) is also useful for determining special values of the Hermite polynomials. In particular, it is straightforward to show that $H_{2n}(0) = (-1)^n (2n)!/n!$ and $H_{2n+1}(0) = 0$.

Recurrence relations

The two most useful recurrence relations satisfied by the Hermite polynomials are given by

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x), \tag{18.134}$$

$$H_n'(x) = 2nH_{n-1}(x). \tag{18.135}$$

The first relation provides a simple iterative way of evaluating the nth Hermite polynomials at some point $x = x_0$, given the values of $H_0(x)$ and $H_1(x)$ at that point. For proofs of these recurrence relations, see exercise 18.5.

18.10 Hypergeometric functions

The hypergeometric equation has the form

$$x(1-x)y'' + [c - (a+b+1)x]y' - aby = 0, \tag{18.136}$$

and has three regular singular points, at $x = 0, 1, \infty$, but no essential singularities. The parameters a, b and c are given real numbers.

In our discussions of Legendre functions, associated Legendre functions and Chebyshev functions in sections 18.1, 18.2 and 18.4, respectively, it was noted that in each case the corresponding second-order differential equation had three regular singular points, at $x = -1, 1, \infty$, and no essential singularities. The hypergeometric equation can, in fact, be considered as the 'canonical form' for second-order differential equations with this number of singularities. It may be shown[§] that,

§ See, for example, J. Mathews and R. L. Walker, *Mathematical Methods of Physics*, 2nd edn (Reading MA: Addision–Wesley, 1971).

by making appropriate changes of the independent and dependent variables, any second-order differential equation with three regular singularities and an ordinary point at infinity can be transformed into the hypergeometric equation (18.136) with the singularities at $= -1$, 1 and ∞. As we discuss below, this allows Legendre functions, associated Legendre functions and Chebyshev functions, for example, to be written as particular cases of *hypergeometric functions*, which are the solutions to (18.136).

Since the point $x = 0$ is a regular singularity of (18.136), we may find at least one solution in the form of a Frobenius series (see section 16.3):

$$y(x) = \sum_{n=0}^{\infty} a_n x^{n+\sigma}. \tag{18.137}$$

Substituting this series into (18.136) and dividing through by $x^{\sigma-1}$, we obtain

$$\sum_{n=0}^{\infty} \{(1-x)(n+\sigma)(n+\sigma-1) + [c-(a+b+1)x](n+\sigma) - abx\}\, a_n x^n = 0. \tag{18.138}$$

Setting $x = 0$, so that only the $n = 0$ term remains, we obtain the indicial equation $\sigma(\sigma - 1) + c\sigma = 0$, which has the roots $\sigma = 0$ and $\sigma = 1 - c$. Thus, provided c is not an integer, one can obtain two linearly independent solutions of the hypergeometric equation in the form (18.137).

For $\sigma = 0$ the corresponding solution is a simple power series. Substituting $\sigma = 0$ into (18.138) and demanding that the coefficient of x^n vanishes, we find the recurrence relation

$$n[(n-1)+c]a_n - [(n-1)(a+b+n-1)+ab]a_{n-1} = 0, \tag{18.139}$$

which, on simplifying and replacing n by $n+1$, yields the recurrence relation

$$a_{n+1} = \frac{(a+n)(b+n)}{(n+1)(c+n)}a_n. \tag{18.140}$$

It is conventional to make the simple choice $a_0 = 1$. Thus, provided c is not a negative integer or zero, we may write the solution as follows:

$$F(a,b,c;x) = 1 + \frac{ab}{c}\frac{x}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{x^2}{2!} + \cdots \tag{18.141}$$

$$= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)}\frac{x^n}{n!}, \tag{18.142}$$

where $F(a,b,c;x)$ is known as the *hypergeometric function* or *hypergeometric series*, and in the second equality we have used the property (18.154) of the gamma function.§ It is straightforward to show that the hypergeometric series converges in the range $|x| < 1$. It also converges at $x = 1$ if $c > a + b$ and at $x = -1$ if $c > a + b - 1$. We also note that $F(a, b, c; x)$ is symmetric in the parameters a and b, i.e. $F(a, b, c; x) = F(b, a, c; x)$.

The hypergeometric function $y(x) = F(a, b, c; x)$ is clearly not the general solution to the hypergeometric equation (18.136), since we must also consider the second root of the indicial equation. Substituting $\sigma = 1 - c$ into (18.138) and demanding that the coefficient of x^n vanishes, we find that we must have

$$n(n + 1 - c)a_n - [(n - c)(a + b + n - c) + ab]a_{n-1} = 0,$$

which, on comparing with (18.139) and replacing n by $n+1$, yields the recurrence relation

$$a_{n+1} = \frac{(a - c + 1 + n)(b - c + 1 + n)}{(n + 1)(2 - c + n)} a_n.$$

We see that this recurrence relation has the same form as (18.140) if one makes the replacements $a \to a - c + 1$, $b \to b - c + 1$ and $c \to 2 - c$. Thus, provided c, $a - b$ and $c - a - b$ are all non-integers, the general solution to the hypergeometric equation, valid for $|x| < 1$, may be written as

$$y(x) = AF(a, b, c; x) + Bx^{1-c}F(a - c + 1, b - c + 1, 2 - c; x), \tag{18.143}$$

where A and B are arbitrary constants to be fixed by the boundary conditions on the solution. If the solution is to be regular at $x = 0$, one requires $B = 0$.

18.10.1 Properties of hypergeometric functions

Since the hypergeometric equation is so general in nature, it is not feasible to present a comprehensive account of the hypergeometric functions. Nevertheless, we outline here some of their most important properties.

Special cases

As mentioned above, the general nature of the hypergeometric equation allows us to write a large number of elementary functions in terms of the hypergeometric functions $F(a, b, c; x)$. Such identifications can be made from the series expansion (18.142) directly, or by transformation of the hypergeometric equation into a more familiar equation, the solutions to which are already known. Some particular examples of well known special cases of the hypergeometric function are as follows:

§ We note that it is also common to denote the hypergeometric function by ${}_2F_1(a, b, c; x)$. This slightly odd-looking notation is meant to signify that, in the coefficient of each power of x, there are two parameters (a and b) in the numerator and one parameter (c) in the denominator.

$$F(a,b,b;x) = (1-x)^{-a}, \qquad F(\tfrac{1}{2},\tfrac{1}{2},\tfrac{3}{2};x^2) = x^{-1}\sin^{-1}x,$$
$$F(1,1,2;-x) = x^{-1}\ln(1+x), \qquad F(\tfrac{1}{2},1,\tfrac{3}{2};-x^2) = x^{-1}\tan^{-1}x,$$
$$\lim_{m\to\infty} F(1,m,1;x/m) = e^x, \qquad F(\tfrac{1}{2},1,\tfrac{3}{2};x^2) = \tfrac{1}{2}x^{-1}\ln[(1+x)/(1-x)],$$
$$F(\tfrac{1}{2},-\tfrac{1}{2},\tfrac{1}{2};\sin^2 x) = \cos x, \qquad F(m+1,-m,1;(1-x)/2) = P_m(x),$$
$$F(\tfrac{1}{2},p,p;\sin^2 x) = \sec x, \qquad F(m,-m,\tfrac{1}{2};(1-x)/2) = T_m(x),$$

where m is an integer, $P_m(x)$ is the mth Legendre polynomial and $T_m(x)$ is the mth Chebyshev polynomial of the first kind. Some of these results are proved in exercise 18.11.

►*Show that* $F(m,-m,\frac{1}{2};(1-x)/2) = T_m(x)$.

Let us prove this result by transforming the hypergeometric equation. The form of the result suggests that we should make the substitution $x = (1-z)/2$ into (18.136), in which case $d/dx = -2d/dz$. Thus, letting $u(z) = y(x)$ and setting $a = m$, $b = -m$ and $c = 1/2$, (18.136) becomes

$$\frac{(1-z)}{2}\frac{(1+z)}{2}(-2)^2\frac{d^2u}{dz^2} + \left[\tfrac{1}{2} - (m-m+1)\frac{1-z}{2}\right](-2)\frac{du}{dz} - (m)(-m)u = 0.$$

On simplifying, we obtain

$$(1-z^2)\frac{d^2u}{dz^2} - z\frac{du}{dz} + m^2u = 0,$$

which has the form of Chebyshev's equation, (18.54). This equation has $u(z) = T_m(z)$ as its power series solution, and so $F(m,-m,\frac{1}{2};(1-z)/2)$ and $T_m(z)$ are equal to within a normalisation factor. On comparing the expressions (18.141) and (18.56) at $x = 0$, i.e. at $z = 1$, we see that they both have value 1. Hence, the normalisations already agree and we obtain the required result. ◄

Integral representation

One of the most useful representations for the hypergeometric functions is in terms of an integral, which may be derived using the properties of the gamma and beta functions discussed in section 18.12. The integral representation reads

$$F(a,b,c;x) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tx)^{-a}\,dt, \tag{18.144}$$

and requires $c > b > 0$ for the integral to converge.

►*Prove the result (18.144).*

From the series expansion (18.142), we have

$$\begin{aligned} F(a,b,c;x) &= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)}\sum_{n=0}^{\infty}\frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)}\frac{x^n}{n!} \\ &= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c-b)}\sum_{n=0}^{\infty}\Gamma(a+n)B(b+n,c-b)\frac{x^n}{n!}, \end{aligned}$$

where in the second equality we have used the expression (18.165) relating the gamma and beta functions. Using the definition (18.162) of the beta function, we then find

$$
\begin{aligned}
F(a,b,c;x) &= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c-b)}\sum_{n=0}^{\infty}\Gamma(a+n)\frac{x^n}{n!}\int_0^1 t^{b+n-1}(1-t)^{c-b-1}\,dt \\
&= \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 dt\, t^{b-1}(1-t)^{c-b-1}\sum_{n=0}^{\infty}\frac{\Gamma(a+n)}{\Gamma(a)}\frac{(tx)^n}{n!},
\end{aligned}
$$

where in the second equality we have rearranged the expression and reversed the order of integration and summation. Finally, one recognises the sum over n as being equal to $(1-tx)^{-a}$, and so we obtain the final result (18.144). ◀

The integral representation may be used to prove a wide variety of properties of the hypergeometric functions. As a simple example, on setting $x = 1$ in (18.144), and using properties of the beta function discussed in section 18.12.2, one quickly finds that, provided c is not a negative integer or zero and $c > a + b$,

$$
F(a,b,c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}.
$$

Relationships between hypergeometric functions

There exist a great many relationships between hypergeometric functions with different arguments. These are most easily derived by making use of the integral representation (18.144) or the series form (18.141). It is not feasible to list all the relationships here, so we simply note two useful examples, which read

$$
F(a,b,c;x) = (1-x)^{c-a-b}F(c-a,c-b,c;x), \tag{18.145}
$$

$$
F'(a,b,c;x) = \frac{ab}{c}F(a+1,b+1,c+1;x), \tag{18.146}
$$

where the prime in the second relation denotes d/dx. The first result follows straightforwardly from the integral representation using the substitution $t = (1-u)/(1-ux)$, whereas the second result may be proved more easily from the series expansion.

In addition to the above results, one may also derive relationships between $F(a,b,c;x)$ and any two of the six 'contiguous functions' $F(a\pm 1,b,c;x)$, $F(a,b\pm 1,c;x)$ and $F(a,b,c\pm 1;x)$. These 'contiguous relations' serve as the recurrence relations for the hypergeometric functions. An example of such a relationship is

$$
(c-a)F(a-1,b,c;x) + (2a-c-ax+bx)F(a,b,c;x) + a(x-1)F(a+1,b,c;x) = 0.
$$

Repeated application of such relationships allows one to express $F(a+l,b+m,c+n;x)$, where l,m,n are integers (with $c+n$ not equalling a negative integer or zero), as a linear combination of $F(a,b,c;x)$ and one of its contiguous functions.

18.11 Confluent hypergeometric functions

The confluent hypergeometric equation has the form

$$xy'' + (c - x)y' - ay = 0; \tag{18.147}$$

it has a regular singularity at $x = 0$ and an essential singularity at $x = \infty$. This equation can be obtained by merging two of the singularities of the ordinary hypergeometric equation (18.136). The parameters a and c are given real numbers.

►*Show that setting $x = z/b$ in the hypergeometric equation, and letting $b \to \infty$, yields the confluent hypergeometric equation.*

Substituting $x = z/b$ into (18.136), with $d/dx = b\,d/dz$, and letting $u(z) = y(x)$, we obtain

$$bz\left(1 - \frac{z}{b}\right)\frac{d^2u}{dz^2} + [bc - (a + b + 1)z]\frac{du}{dz} - abu = 0,$$

which clearly has regular singular points at $z = 0$, b and ∞. If we now merge the last two singularities by letting $b \to \infty$, we obtain

$$zu'' + (c - z)u' - au = 0,$$

where the primes denote d/dz. Hence $u(z)$ must satisfy the confluent hypergeometric equation. ◄

In our discussion of Bessel, Laguerre and associated Laguerre functions, it was noted that the corresponding second-order differential equation in each case had a single regular singular point at $x = 0$ and an essential singularity at $x = \infty$. From table 16.1, we see that this is also true for the confluent hypergeometric equation. Indeed, this equation can be considered as the 'canonical form' for second-order differential equations with this pattern of singularities. Consequently, as we mention below, the Bessel, Laguerre and associated Laguerre functions can all be written in terms of the *confluent hypergeometric functions*, which are the solutions of (18.147).

The solutions of the confluent hypergeometric equation are obtained from those of the ordinary hypergeometric equation by again letting $x \to x/b$ and carrying out the limiting process $b \to \infty$. Thus, from (18.141) and (18.143), two linearly independent solutions of (18.147) are (when c is not an integer)

$$y_1(x) = 1 + \frac{a}{c}\frac{x}{1!} + \frac{a(a+1)}{c(c+1)}\frac{z^2}{2!} + \cdots \equiv M(a, c; x), \tag{18.148}$$

$$y_2(x) = x^{1-c}M(a - c + 1, 2 - c; x), \tag{18.149}$$

where $M(a, c; x)$ is called the *confluent hypergeometric function* (or *Kummer function*).[§] It is worth noting, however, that $y_1(x)$ is singular when $c = 0, -1, -2, \ldots$ and $y_2(x)$ is singular when $c = 2, 3, 4, \ldots$. Thus, it is conventional to take the

[§] We note that an alternative notation for the confluent hypergeometric function is ${}_1F_1(a, c; x)$.

second solution to (18.147) as a linear combination of (18.148) and (18.149) given by

$$U(a,c;x) \equiv \frac{\pi}{\sin \pi c}\left[\frac{M(a,c;x)}{\Gamma(a-c+1)\Gamma(c)} - x^{1-c}\frac{M(a-c+1,2-c;x)}{\Gamma(a)\Gamma(2-c)}\right].$$

This has a well behaved limit as c approaches an integer.

18.11.1 Properties of confluent hypergeometric functions

The properties of confluent hypergeometric functions can be derived from those of ordinary hypergeometric functions by letting $x \to x/b$ and taking the limit $b \to \infty$, in the same way as both the equation and its solution were derived. A general procedure of this sort is called a *confluence* process.

Special cases

The general nature of the confluent hypergeometric equation allows one to write a large number of elementary functions in terms of the confluent hypergeometric functions $M(a,c;x)$. Once again, such identifications can be made from the series expansion (18.148) directly, or by transformation of the confluent hypergeometric equation into a more familiar equation for which the solutions are already known. Some particular examples of well known special cases of the confluent hypergeometric function are as follows:

$$M(a,a;x) = e^x, \qquad M(1,2;2x) = \frac{e^x \sinh x}{x},$$

$$M(-n,1;x) = L_n(x), \qquad M(-n,m+1;x) = \frac{n!m!}{(n+m)!}L_n^m(x),$$

$$M(-n,\tfrac{1}{2};x^2) = \frac{(-1)^n n!}{(2n)!}H_{2n}(x), \qquad M(-n,\tfrac{3}{2};x^2), = \frac{(-1)^n n!}{2(2n+1)!}\frac{H_{2n+1}(x)}{x},$$

$$M(v+\tfrac{1}{2},2v+1;2ix) = v!e^{ix}(\frac{x}{2})^{-v}J_v(x), \quad M(\tfrac{1}{2},\tfrac{3}{2};-x^2) = \frac{\sqrt{\pi}}{2x}\text{erf}(x),$$

where n and m are integers, $L_n^m(x)$ is an associated Legendre polynomial, $H_n(x)$ is a Hermite polynomial, $J_v(x)$ is a Bessel function and $\text{erf}(x)$ is the error function discussed in section 18.12.4.

Integral representation

Using the integral representation (18.144) of the ordinary hypergeometric function, exchanging a and b and carrying out the process of confluence gives

$$M(a,c,x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 e^{tx}t^{a-1}(1-t)^{c-a-1}\,dt, \tag{18.150}$$

which converges provided $c > a > 0$.

►*Prove the result (18.150).*

Since $F(a,b,c;x)$ is unchanged by swapping a and b, we may write its integral representation (18.144) as

$$F(a,b,c;x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 t^{a-1}(1-t)^{c-a-1}(1-tx)^{-b}\, dt.$$

Setting $x = z/b$ and taking the limit $b \to \infty$, we obtain

$$M(a,c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 t^{a-1}(1-t)^{c-a-1} \lim_{b\to\infty}\left(1-\frac{tz}{b}\right)^{-b} dt.$$

Since the limit is equal to e^{tz}, we obtain result (18.150). ◄

Relationships between confluent hypergeometric functions

A large number of relationships exist between confluent hypergeometric functions with different arguments. These are straightforwardly derived using the integral representation (18.150) or the series form (18.148). Here, we simply note two useful examples, which read

$$M(a,c;x) = e^x M(c-a,c;-x), \tag{18.151}$$

$$M'(a,c;x) = \frac{a}{c}M(a+1,c+1;x), \tag{18.152}$$

where the prime in the second relation denotes d/dx. The first result follows straightforwardly from the integral representation, and the second result may be proved from the series expansion (see exercise 18.19).

In an analogous manner to that used for the ordinary hypergeometric functions, one may also derive relationships between $M(a,c;x)$ and any two of the four 'contiguous functions' $M(a \pm 1, c; x)$ and $M(a, c \pm 1; x)$. These serve as the recurrence relations for the confluent hypergeometric functions. An example of such a relationship is

$$(c-a)M(a-1,c;x) + (2a-c+x)M(a,c;x) - aM(a+1,c;x) = 0.$$

18.12 The gamma function and related functions

Many times in this chapter, and often throughout the rest of the book, we have made mention of the gamma function and related functions such as the beta and error functions. Although not derived as the solutions of important second-order ODEs, these convenient functions appear in a number of contexts, and so here we gather together some of their properties. This final section should be regarded merely as a reference containing some useful relations obeyed by these functions; a minimum of formal proofs is given.

18.12.1 The gamma function

The *gamma function* $\Gamma(n)$ is defined by

$$\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}\,dx, \tag{18.153}$$

which converges for $n > 0$, where in general n is a real number. Replacing n by $n+1$ in (18.153) and integrating the RHS by parts, we find

$$\begin{aligned}\Gamma(n+1) &= \int_0^\infty x^n e^{-x}\,dx \\ &= \left[-x^n e^{-x}\right]_0^\infty + \int_0^\infty nx^{n-1}e^{-x}\,dx \\ &= n\int_0^\infty x^{n-1}e^{-x}\,dx,\end{aligned}$$

from which we obtain the important result

$$\Gamma(n+1) = n\Gamma(n). \tag{18.154}$$

From (18.153), we see that $\Gamma(1) = 1$, and so, if n is a positive integer,

$$\Gamma(n+1) = n!. \tag{18.155}$$

In fact, equation (18.155) serves as a definition of the factorial function even for non-integer n. For negative n the factorial function is defined by

$$n! = \frac{(n+m)!}{(n+m)(n+m-1)\cdots(n+1)}, \tag{18.156}$$

where m is any positive integer that makes $n+m > 0$. Different choices of m ($> -n$) do not lead to different values for $n!$. A plot of the gamma function is given in figure 18.9, where it can be seen that the function is infinite for negative integer values of n, in accordance with (18.156). For an extension of the factorial function to complex arguments, see exercise 18.15.

By letting $x = y^2$ in (18.153), we immediately obtain another useful representation of the gamma function given by

$$\Gamma(n) = 2\int_0^\infty y^{2n-1}e^{-y^2}\,dy. \tag{18.157}$$

Setting $n = \frac{1}{2}$ we find the result

$$\Gamma\left(\tfrac{1}{2}\right) = 2\int_0^\infty e^{-y^2}\,dy = \int_{-\infty}^\infty e^{-y^2}\,dy = \sqrt{\pi},$$

where have used the standard integral discussed in section 6.4.2. From this result, $\Gamma(n)$ for half-integral n can be found using (18.154). Some immediately derivable factorial values of half integers are

$$\left(-\tfrac{3}{2}\right)! = -2\sqrt{\pi}, \qquad \left(-\tfrac{1}{2}\right)! = \sqrt{\pi}, \qquad \left(\tfrac{1}{2}\right)! = \tfrac{1}{2}\sqrt{\pi}, \qquad \left(\tfrac{3}{2}\right)! = \tfrac{3}{4}\sqrt{\pi}.$$

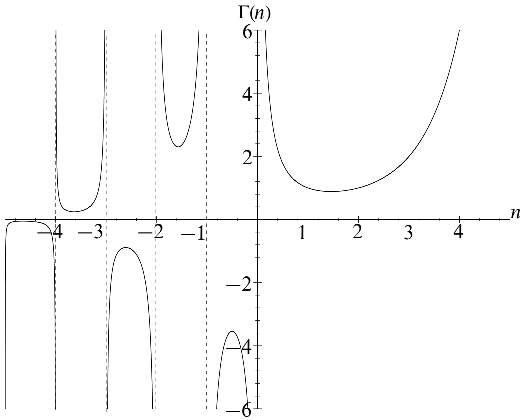


Figure 18.9 The gamma function $\Gamma(n)$.

Moreover, it may be shown for non-integral n that the gamma function satisfies the important identity

$$\Gamma(n)\Gamma(1-n) = \frac{\pi}{\sin n\pi}. \tag{18.158}$$

This is proved for a restricted range of n in the next section, once the beta function has been introduced.

It can also be shown that the gamma function is given by

$$\Gamma(n+1) = \sqrt{2\pi n}\; n^n e^{-n} \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51\,840n^3} + \ldots\right) = n!, \tag{18.159}$$

which is known as *Stirling's asymptotic series*. For large n the first term dominates, and so

$$n! \approx \sqrt{2\pi n}\; n^n e^{-n}; \tag{18.160}$$

this is known as *Stirling's approximation*. This approximation is particularly useful in statistical thermodynamics, when arrangements of a large number of particles are to be considered.

► *Prove Stirling's approximation $n! \approx \sqrt{2\pi n}\; n^n e^{-n}$ for large n.*

From (18.153), the extended definition of the factorial function (which is valid for $n > -1$) is given by

$$n! = \int_0^\infty x^n e^{-x}\,dx = \int_0^\infty e^{n\ln x - x}\,dx. \tag{18.161}$$

If we let $x = n + y$, then

$$\ln x = \ln n + \ln\left(1 + \frac{y}{n}\right)$$
$$= \ln n + \frac{y}{n} - \frac{y^2}{2n^2} + \frac{y^3}{3n^3} - \cdots .$$

Substituting this result into (18.161), we obtain

$$n! = \int_{-n}^{\infty} \exp\left[n\left(\ln n + \frac{y}{n} - \frac{y^2}{2n^2} + \cdots\right) - n - y\right] dy.$$

Thus, when n is sufficiently large, we may approximate $n!$ by

$$n! \approx e^{n\ln n - n} \int_{-\infty}^{\infty} e^{-y^2/(2n)}\, dy = e^{n\ln n - n}\sqrt{2\pi n} = \sqrt{2\pi n}\, n^n e^{-n},$$

which is Stirling's approximation (18.160). ◄

18.12.2 The beta function

The *beta function* is defined by

$$B(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1}\, dx, \tag{18.162}$$

which converges for $m > 0$, $n > 0$, where m and n are, in general, real numbers. By letting $x = 1 - y$ in (18.162) it is easy to show that $B(m,n) = B(n,m)$. Other useful representations of the beta function may be obtained by suitable changes of variable. For example, putting $x = (1+y)^{-1}$ in (18.162), we find that

$$B(m,n) = \int_0^{\infty} \frac{y^{n-1}\, dy}{(1+y)^{m+n}}. \tag{18.163}$$

Alternatively, if we let $x = \sin^2\theta$ in (18.162), we obtain immediately

$$B(m,n) = 2\int_0^{\pi/2} \sin^{2m-1}\theta\, \cos^{2n-1}\theta\, d\theta. \tag{18.164}$$

The beta function may also be written in terms of the gamma function as

$$B(m,n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}. \tag{18.165}$$

►*Prove the result (18.165).*

Using (18.157), we have

$$\Gamma(n)\Gamma(m) = 4\int_0^{\infty} x^{2n-1}e^{-x^2}\, dx \int_0^{\infty} y^{2m-1}e^{-y^2}\, dy$$
$$= 4\int_0^{\infty}\int_0^{\infty} x^{2n-1}y^{2m-1}e^{-(x^2+y^2)}\, dx\, dy.$$

Changing variables to plane polar coordinates (ρ, ϕ) given by $x = \rho\cos\phi$, $y = \rho\sin\phi$, we obtain

$$\begin{aligned}\Gamma(n)\Gamma(m) &= 4\int_0^{\pi/2}\int_0^{\infty} \rho^{2(m+n-1)}e^{-\rho^2}\sin^{2m-1}\phi\,\cos^{2n-1}\phi\,\rho\,d\rho\,d\phi \\ &= 4\int_0^{\pi/2}\sin^{2m-1}\phi\,\cos^{2n-1}\phi\,d\phi\int_0^{\infty}\rho^{2(m+n)-1}e^{-\rho^2}\,d\rho \\ &= B(m,n)\Gamma(m+n),\end{aligned}$$

where in the last line we have used the results (18.157) and (18.164). ◄

The result (18.165) is useful in proving the identity (18.158) satisfied by the gamma function, since

$$\Gamma(n)\Gamma(1-n) = B(1-n, n) = \int_0^{\infty}\frac{y^{n-1}\,dy}{1+y},$$

where, in the second equality, we have used the integral representation (18.163). For $0 < n < 1$ this integral can be evaluated using contour integration and has the value $\pi/(\sin n\pi)$ (see exercise 24.19), thereby proving result (18.158) for this range of n. Extensions to other ranges require more sophisticated methods.

18.12.3 The incomplete gamma function

In the definition (18.153) of the gamma function, we may divide the range of integration into two parts and write

$$\Gamma(n) = \int_0^x u^{n-1}e^{-u}\,du + \int_x^{\infty} u^{n-1}e^{-u}\,du \equiv \gamma(n,x) + \Gamma(n,x), \tag{18.166}$$

whereby we have defined the *incomplete gamma functions* $\gamma(n,x)$ and $\Gamma(n,x)$, respectively. The choice of which of these two functions to use is merely a matter of convenience.

►*Show that if n is a positive integer*

$$\Gamma(n,x) = (n-1)!e^{-x}\sum_{k=0}^{n-1}\frac{x^k}{k!}.$$

From (18.166), on integrating by parts we find

$$\begin{aligned}\Gamma(n,x) = \int_x^{\infty} u^{n-1}e^{-u}du &= x^{n-1}e^{-x} + (n-1)\int_x^{\infty} u^{n-2}e^{-u}du \\ &= x^{n-1}e^{-x} + (n-1)\Gamma(n-1,x),\end{aligned}$$

which is valid for arbitrary n. If n is an integer, however, we obtain

$$\begin{aligned}\Gamma(n,x) &= e^{-x}[x^{n-1} + (n-1)x^{n-2} + (n-1)(n-2)x^{n-3} + \cdots + (n-1)!] \\ &= (n-1)!\,e^{-x}\sum_{k=0}^{n-1}\frac{x^k}{k!},\end{aligned}$$

which is the required result. ◄

We note that is it conventional to define, in addition, the functions

$$P(a,x) \equiv \frac{\gamma(a,x)}{\Gamma(a)}, \qquad Q(a,x) \equiv \frac{\Gamma(a,x)}{\Gamma(a)},$$

which are also often called incomplete gamma functions; it is clear that $Q(a,x) = 1 - P(a,x)$.

18.12.4 The error function

Finally, we mention the *error function*, which is encountered in probability theory and in the solutions of some partial differential equations. The error function is related to the incomplete gamma function by $\text{erf}(x) = \gamma(\frac{1}{2}, x^2)/\sqrt{\pi}$ and is thus given by

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}\, du = 1 - \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2}\, du. \tag{18.167}$$

From this definition we can easily see that

$$\text{erf}(0) = 0, \qquad \text{erf}(\infty) = 1, \qquad \text{erf}(-x) = -\text{erf}(x).$$

By making the substitution $y = \sqrt{2}u$ in (18.167), we find

$$\text{erf}(x) = \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{2}x} e^{-y^2/2}\, dy.$$

The cumulative probability function $\Phi(x)$ for the standard Gaussian distribution (discussed in section 30.9.1) may be written in terms of the error function as follows:

$$\begin{aligned}\Phi(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2}\, dy \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2}\, dy \\ &= \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{x}{\sqrt{2}}\right).\end{aligned}$$

It is also sometimes useful to define the *complementary error function*

$$\text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2}\, du = \frac{\Gamma(\frac{1}{2}, x^2)}{\sqrt{\pi}}. \tag{18.168}$$

18.13 Exercises

18.1 Use the explicit expressions

$$Y_0^0 = \sqrt{\tfrac{1}{4\pi}}, \qquad Y_1^0 = \sqrt{\tfrac{3}{4\pi}}\cos\theta,$$

$$Y_1^{\pm 1} = \mp\sqrt{\tfrac{3}{8\pi}}\sin\theta\exp(\pm i\phi), \qquad Y_2^0 = \sqrt{\tfrac{5}{16\pi}}(3\cos^2\theta - 1),$$

$$Y_2^{\pm 1} = \mp\sqrt{\tfrac{15}{8\pi}}\sin\theta\cos\theta\exp(\pm i\phi), \quad Y_2^{\pm 2} = \sqrt{\tfrac{15}{32\pi}}\sin^2\theta\exp(\pm 2i\phi),$$

to verify for $\ell = 0, 1, 2$ that

$$\sum_{m=-\ell}^{\ell} |Y_\ell^m(\theta, \phi)|^2 = \frac{2\ell + 1}{4\pi},$$

and so is independent of the values of θ and ϕ. This is true for any ℓ, but a general proof is more involved. This result helps to reconcile intuition with the apparently arbitrary choice of polar axis in a general quantum mechanical system.

18.2 Express the function

$$f(\theta, \phi) = \sin\theta[\sin^2(\theta/2)\cos\phi + i\cos^2(\theta/2)\sin\phi] + \sin^2(\theta/2)$$

as a sum of spherical harmonics.

18.3 Use the generating function for the Legendre polynomials $P_n(x)$ to show that

$$\int_0^1 P_{2n+1}(x)\,dx = (-1)^n \frac{(2n)!}{2^{2n+1} n!(n+1)!}$$

and that, except for the case $n = 0$,

$$\int_0^1 P_{2n}(x)\,dx = 0.$$

18.4 Carry through the following procedure as a proof of the result

$$I_n = \int_{-1}^{1} P_n(z)P_n(z)\,dz = \frac{2}{2n+1}.$$

(a) Square both sides of the generating-function definition of the Legendre polynomials,

$$(1 - 2zh + h^2)^{-1/2} = \sum_{n=0}^{\infty} P_n(z)h^n.$$

(b) Express the RHS as a sum of powers of h, obtaining expressions for the coefficients.
(c) Integrate the RHS from -1 to 1 and use the orthogonality property of the Legendre polynomials.
(d) Similarly integrate the LHS and expand the result in powers of h.
(e) Compare coefficients.

18.5 The Hermite polynomials $H_n(x)$ may be defined by

$$\Phi(x, h) = \exp(2xh - h^2) = \sum_{n=0}^{\infty} \frac{1}{n!} H_n(x)h^n.$$

Show that

$$\frac{\partial^2\Phi}{\partial x^2} - 2x\frac{\partial\Phi}{\partial x} + 2h\frac{\partial\Phi}{\partial h} = 0,$$

and hence that the $H_n(x)$ satisfy the Hermite equation

$$y'' - 2xy' + 2ny = 0,$$

where n is an integer ≥ 0.

Use Φ to prove that

(a) $H_n'(x) = 2nH_{n-1}(x)$,
(b) $H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0$.

18.6 A charge $+2q$ is situated at the origin and charges of $-q$ are situated at distances $\pm a$ from it along the polar axis. By relating it to the generating function for the Legendre polynomials, show that the electrostatic potential Φ at a point (r, θ, ϕ) with $r > a$ is given by

$$\Phi(r, \theta, \phi) = \frac{2q}{4\pi\epsilon_0 r} \sum_{s=1}^{\infty} \left(\frac{a}{r}\right)^{2s} P_{2s}(\cos\theta).$$

18.7 For the associated Laguerre polynomials, carry through the following exercises.

(a) Prove the Rodrigues' formula

$$L_n^m(x) = \frac{e^x x^{-m}}{n!} \frac{d^n}{dx^n}(x^{n+m} e^{-x}),$$

taking the polynomials to be defined by

$$L_n^m(x) = \sum_{k=0}^{n} (-1)^k \frac{(n+m)!}{k!(n-k)!(k+m)!} x^k.$$

(b) Prove the recurrence relations

$$(n+1)L_{n+1}^m(x) = (2n+m+1-x)L_n^m(x) - (n+m)L_{n-1}^m(x),$$

$$x(L_n^m)'(x) = nL_n^m(x) - (n+m)L_{n-1}^m(x),$$

but this time taking the polynomial as defined by

$$L_n^m(x) = (-1)^m \frac{d^m}{dx^m} L_{n+m}(x)$$

or the generating function.

18.8 The quantum mechanical wavefunction for a one-dimensional simple harmonic oscillator in its nth energy level is of the form

$$\psi(x) = \exp(-x^2/2)H_n(x),$$

where $H_n(x)$ is the nth Hermite polynomial. The generating function for the polynomials is

$$G(x, h) = e^{2hx - h^2} = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!} h^n.$$

(a) Find $H_i(x)$ for $i = 1, 2, 3, 4$.
(b) Evaluate by direct calculation

$$\int_{-\infty}^{\infty} e^{-x^2} H_p(x) H_q(x)\, dx,$$

(i) for $p = 2$, $q = 3$; (ii) for $p = 2$, $q = 4$; (iii) for $p = q = 3$. Check your answers against the expected values $2^p\, p! \sqrt{\pi}\, \delta_{pq}$.

[You will find it convenient to use

$$\int_{-\infty}^{\infty} x^{2n} e^{-x^2}\, dx = \frac{(2n)!\sqrt{\pi}}{2^{2n} n!}$$

for integer $n \geq 0$.]

18.9 By initially writing $y(x)$ as $x^{1/2} f(x)$ and then making subsequent changes of variable, reduce Stokes' equation,

$$\frac{d^2 y}{dx^2} + \lambda x y = 0,$$

to Bessel's equation. Hence show that a solution that is finite at $x = 0$ is a multiple of $x^{1/2} J_{1/3}(\frac{2}{3}\sqrt{\lambda x^3})$.

18.10 By choosing a suitable form for h in their generating function,

$$G(z, h) = \exp\left[\frac{z}{2}\left(h - \frac{1}{h}\right)\right] = \sum_{n=-\infty}^{\infty} J_n(z) h^n,$$

show that integral repesentations of the Bessel functions of the first kind are given, for integral m, by

$$J_{2m}(z) = \frac{(-1)^m}{\pi} \int_0^{2\pi} \cos(z\cos\theta)\, \cos 2m\theta\, d\theta \qquad m \geq 1,$$

$$J_{2m+1}(z) = \frac{(-1)^{m+1}}{\pi} \int_0^{2\pi} \cos(z\cos\theta)\, \sin(2m+1)\theta\, d\theta \qquad m \geq 0.$$

18.11 Identify the series for the following hypergeometric functions, writing them in terms of better known functions:

(a) $F(a, b, b; z)$,
(b) $F(1, 1, 2; -x)$,
(c) $F(\frac{1}{2}, 1, \frac{3}{2}; -x^2)$,
(d) $F(\frac{1}{2}, \frac{1}{2}, \frac{3}{2}; x^2)$,
(e) $F(-a, a, \frac{1}{2}; \sin^2 x)$; this is a much more difficult exercise.

18.12 By making the substitution $z = (1 - x)/2$ and suitable choices for a, b and c, convert the hypergeometric equation,

$$z(1-z)\frac{d^2 u}{dz^2} + [\,c - (a + b + 1)z\,]\frac{du}{dz} - abu = 0,$$

into the Legendre equation,

$$(1 - x^2)\frac{d^2 y}{dx^2} - 2x\frac{dy}{dx} + \ell(\ell + 1)y = 0.$$

Hence, using the hypergeometric series, generate the Legendre polynomials $P_\ell(x)$ for the integer values $\ell = 0, 1, 2, 3$. Comment on their normalisations.

18.13 Find a change of variable that will allow the integral

$$I = \int_1^{\infty} \frac{\sqrt{u - 1}}{(u + 1)^2}\, du$$

to be expressed in terms of the beta function, and so evaluate it.

18.14 Prove that, if m and n are both greater than -1, then

$$I = \int_0^{\infty} \frac{u^m}{(au^2 + b)^{(m+n+2)/2}}\, du = \frac{\Gamma[\frac{1}{2}(m+1)]\,\Gamma[\frac{1}{2}(n+1)]}{2a^{(m+1)/2}\, b^{(n+1)/2}\, \Gamma[\frac{1}{2}(m+n+2)]}.$$

Deduce the value of

$$J = \int_0^\infty \frac{(u+2)^2}{(u^2+4)^{5/2}}\, du.$$

18.15 The complex function $z!$ is defined by

$$z! = \int_0^\infty u^z e^{-u}\, du \qquad \text{for Re } z > -1.$$

For Re $z \le -1$ it is defined by

$$z! = \frac{(z+n)!}{(z+n)(z+n-1)\cdots(z+1)},$$

where n is any (positive) integer $> -\text{Re } z$. Being the ratio of two polynomials, $z!$ is analytic everywhere in the finite complex plane except at the poles that occur when z is a negative integer.

(a) Show that the definition of $z!$ for Re $z \le -1$ is independent of the value of n chosen.
(b) Prove that the residue of $z!$ at the pole $z = -m$, where m is an integer > 0, is $(-1)^{m-1}/(m-1)!$.

18.16 For $-1 < \text{Re } z < 1$, use the definition and value of the beta function to show that

$$z!(-z)! = \int_0^\infty \frac{u^z}{(1+u)^2}\, du.$$

Contour integration gives the value of the integral on the RHS of the above equation as $\pi z \operatorname{cosec} \pi z$. Use this to deduce the value of $(-\frac{1}{2})!$.

18.17 The integral

$$I = \int_{-\infty}^{\infty} \frac{e^{-k^2}}{k^2+a^2}\, dk, \qquad (*)$$

in which $a > 0$, occurs in some statistical mechanics problems. By first considering the integral

$$J = \int_0^\infty e^{iu(k+ia)}\, du,$$

and a suitable variation of it, show that $I = (\pi/a)\exp(a^2)\operatorname{erfc}(a)$, where $\operatorname{erfc}(x)$ is the complementary error function.

18.18 Consider two series expansions of the error function as follows.

(a) Obtain a series expansion of the error function $\operatorname{erf}(x)$ in ascending powers of x. How many terms are needed to give a value correct to four significant figures for $\operatorname{erf}(1)$?
(b) Obtain an asymptotic expansion that can be used to estimate $\operatorname{erfc}(x)$ for large x (> 0) in the form of a series

$$\operatorname{erfc}(x) = R(x) = e^{-x^2} \sum_{n=0}^{\infty} \frac{a_n}{x^n}.$$

Consider what bounds can be put on the estimate and at what point the infinite series should be terminated in a practical estimate. In particular, estimate $\operatorname{erfc}(1)$ and test the answer for compatibility with that in part (a).

18.19 For the functions $M(a, c; z)$ that are the solutions of the confluent hypergeometric equation,

(a) use their series representation to prove that

$$b\frac{d}{dz}M(a,c;z) = a\,M(a+1,c+1;z);$$

(b) use an integral representation to prove that

$$M(a,c;z) = e^z M(c-a,c;-z).$$

18.20 The Bessel function $J_\nu(z)$ can be considered as a special case of the solution $M(a,c;z)$ of the confluent hypergeometric equation, the connection being

$$\lim_{a\to\infty}\frac{M(a,\nu+1;-z/a)}{\Gamma(\nu+1)} = z^{-\nu/2}J_\nu(2\sqrt{z}).$$

Prove this equality by writing each side in terms of an infinite series and showing that the series are the same.

18.21 Find the differential equation satisfied by the function $y(x)$ defined by

$$y(x) = Ax^{-n}\int_0^x e^{-t}t^{n-1}\,dt \equiv Ax^{-n}\gamma(n,x),$$

and, by comparing it with the confluent hypergeometric function, express y as a multiple of the solution $M(a,c;z)$ of that equation. Determine the value of A that makes y equal to M.

18.22 Show, from its definition, that the Bessel function of the second kind, and of integral order ν, can be written as

$$Y_\nu(z) = \frac{1}{\pi}\left[\frac{\partial J_\mu(z)}{\partial\mu} - (-1)^\nu\frac{\partial J_{-\mu}(z)}{\partial\mu}\right]_{\mu=\nu}.$$

Using the explicit series expression for $J_\mu(z)$, show that $\partial J_\mu(z)/\partial\mu$ can be written as

$$J_\nu(z)\ln\left(\frac{z}{2}\right) + g(\nu,z),$$

and deduce that $Y_\nu(z)$ can be expressed as

$$Y_\nu(z) = \frac{2}{\pi}J_\nu(z)\ln\left(\frac{z}{2}\right) + h(\nu,z),$$

where $h(\nu,z)$, like $g(\nu,z)$, is a power series in z.

18.23 Prove two of the properties of the incomplete gamma function $P(a,x^2)$ as follows.

(a) By considering its form for a suitable value of a, show that the error function can be expressed as a particular case of the incomplete gamma function.

(b) The Fresnel integrals, of importance in the study of the diffraction of light, are given by

$$C(x) = \int_0^x \cos\left(\frac{\pi}{2}t^2\right)dt, \qquad S(x) = \int_0^x \sin\left(\frac{\pi}{2}t^2\right)dt.$$

Show that they can be expressed in terms of the error function by

$$C(x) + iS(x) = A\,\mathrm{erf}\left[\frac{\sqrt{\pi}}{2}(1-i)x\right],$$

where A is a (complex) constant, which you should determine. Hence express $C(x)+iS(x)$ in terms of the incomplete gamma function.

18.24 The solutions $y(x, a)$ of the equation

$$\frac{d^2y}{dx^2} - (\tfrac{1}{4}x^2 + a)y = 0 \qquad (*)$$

are known as parabolic cylinder functions.

(a) If $y(x, a)$ is a solution of $(*)$, determine which of the following are also solutions: (i) $y(a, -x)$, (ii) $y(-a, x)$, (iii) $y(a, ix)$ and (iv) $y(-a, ix)$.

(b) Show that one solution of $(*)$, even in x, is

$$y_1(x, a) = e^{-x^2/4} M(\tfrac{1}{2}a + \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{2}x^2),$$

where $M(\alpha, c, z)$ is the confluent hypergeometric function satisfying

$$z\frac{d^2M}{dz^2} + (c - z)\frac{dM}{dz} - \alpha M = 0.$$

You may assume (or prove) that a second solution, odd in x, is given by $y_2(x, a) = xe^{-x^2/4} M(\tfrac{1}{2}a + \tfrac{3}{4}, \tfrac{3}{2}, \tfrac{1}{2}x^2)$.

(c) Find, as an infinite series, an explicit expression for $e^{x^2/4} y_1(x, a)$.

(d) Using the results from part (a), show that $y_1(x, a)$ can also be written as

$$y_1(x, a) = e^{x^2/4} M(-\tfrac{1}{2}a + \tfrac{1}{4}, \tfrac{1}{2}, -\tfrac{1}{2}x^2).$$

(e) By making a suitable choice for a deduce that

$$1 + \sum_{n=1}^{\infty} \frac{b_n x^{2n}}{(2n)!} = e^{x^2/2}\left(1 + \sum_{n=1}^{\infty} \frac{(-1)^n b_n x^{2n}}{(2n)!}\right),$$

where $b_n = \prod_{r=1}^{n}(2r - \tfrac{3}{2})$.

18.14 Hints and answers

18.1 Note that taking the square of the modulus eliminates all mention of ϕ.

18.3 Integrate both sides of the generating function definition from $x = 0$ to $x = 1$, and then expand the resulting term, $(1 + h^2)^{1/2}$, using a binomial expansion. Show that ${}^{1/2}C_m$ can be written as $[(-1)^{m-1}(2m-2)!]/[2^{2m-1}m!(m-1)!]$.

18.5 Prove the stated equation using the explicit closed form of the generating function. Then substitute the series and require the coefficient of each power of h to vanish. (b) Differentiate result (a) and then use (a) again to replace the derivatives.

18.7 (a) Write the result of using Leibnitz' theorem on the product of x^{n+m} and e^{-x} as a finite sum, evaluate the separated derivatives, and then re-index the summation. (b) For the first recurrence relation, differentiate the generating function with respect to h and then use the generating function again to replace the exponential. Equating coefficients of h^n then yields the result. For the second, differentiate the corresponding relationship for the ordinary Laguerre polynomials m times.

18.9 $x^2 f'' + xf' + (\lambda x^3 - \tfrac{1}{4})f = 0$. Then, in turn, set $x^{3/2} = u$, and $\tfrac{2}{3}\lambda^{1/2}u = v$; then v satisfies Bessel's equation with $\nu = \tfrac{1}{3}$.

18.11 (a) $(1 - z)^{-a}$. (b) $x^{-1}\ln(1 + x)$. (c) Compare the calculated coefficients with those of $\tan^{-1} x$. $F(\tfrac{1}{2}, 1, \tfrac{3}{2}; -x^2) = x^{-1}\tan^{-1} x$. (d) $x^{-1}\sin^{-1} x$. (e) Note that a term containing x^{2n} can only arise from the first $n+1$ terms of an expansion in powers of $\sin^2 x$; make a few trials. $F(-a, a, \tfrac{1}{2}; \sin^2 x) = \cos 2ax$.

18.13 Looking for $f(x) = u$ such that $u + 1$ is an inverse power of x with $f(0) = \infty$ and $f(1) = 1$ leads to $f(x) = 2x^{-1} - 1$. $I = B(\tfrac{1}{2}, \tfrac{3}{2})/\sqrt{2} = \pi/(2\sqrt{2})$.

18.15 (a) Show that the ratio of two definitions based on m and n, with $m > n > -\text{Re}\, z$, is unity, independent of the actual values of m and n.
(b) Consider the limit as $z \to -m$ of $(z+m)z!$, with the definition of $z!$ based on n where $n > m$.

18.17 Express the integrand in partial fractions and use J, as given, and $J' = \int_0^\infty \exp[-iu(k-ia)]\,du$ to express I as the sum of two double integral expressions. Reduce them using the standard Gaussian integral, and then make a change of variable $2v = u + 2a$.

18.19 (b) Using the representation

$$M(a,b;z) = \frac{\Gamma(b)}{\Gamma(b-a)\,\Gamma(a)} \int_0^1 e^{zt}\, t^{a-1}\,(1-t)^{b-a-1}\,dt$$

allows the equality to be established, without actual integration, by changing the integration variable to $s = 1 - t$.

18.21 Calculate $y'(x)$ and $y''(x)$ and then eliminate $x^{-1}e^{-x}$ to obtain $xy'' + (n+1+x)y' + ny = 0$; $M(n, n+1; -x)$. Comparing the expansion of the hypergeometric series with the result of term by term integration of the expansion of the integrand shows that $A = n$.

18.23 (a) If the dummy variable in the incomplete gamma function is t, make the change of variable $y = +\sqrt{t}$. Now choose a so that $2(a-1)+1 = 0$; $\text{erf}(x) = P(\frac{1}{2}, x^2)$.
(b) Change the integration variable u in the standard representation of the RHS to s, given by $u = \frac{1}{2}\sqrt{\pi}(1-i)s$, and note that $(1-i)^2 = -2i$. $A = (1+i)/2$. From part (a), $C(x) + iS(x) = \frac{1}{2}(1+i)P(\frac{1}{2}, -\frac{1}{2}\pi i\, x^2)$.

19

Quantum operators

Although the previous chapter was principally concerned with the use of linear operators and their eigenfunctions in connection with the solution of given differential equations, it is of interest to study the properties of the operators themselves and determine which of them follow purely from the nature of the operators, without reference to specific forms of eigenfunctions.

19.1 Operator formalism

The results we will obtain in this chapter have most of their applications in the field of quantum mechanics and our descriptions of the methods will reflect this. In particular, when we discuss a function ψ that depends upon variables such as space coordinates and time, and possibly also on some non-classical variables, ψ will usually be a quantum-mechanical wavefunction that is being used to describe the state of a physical system. For example, the value of $|\psi|^2$ for a particular set of values of the variables is interpreted in quantum mechanics as being the probability that the system's variables have that set of values.

To this end, we will be no more specific about the functions involved than attaching just enough labels to them that a particular function, or a particular set of functions, is identified. A convenient notation for this kind of approach is that already hinted at, but not specifically stated, in subsection 17.1, where the definition of an inner product is given. This notation, often called the Dirac notation, denotes a state whose wavefunction is ψ by $|\,\psi\rangle$; since ψ belongs to a vector space of functions, $|\,\psi\rangle$ is known as a *ket vector*. Ket vectors, or simply kets, must not be thought of as completely analogous to physical vectors. Quantum mechanics associates the same physical state with $ke^{i\theta}|\,\psi\rangle$ as it does with $|\,\psi\rangle$ for all real k and θ and so there is no loss of generality in taking k as 1 and θ as 0. On the other hand, the combination $c_1|\,\psi_1\rangle + c_2|\,\psi_2\rangle$, where $|\,\psi_1\rangle$ and $|\,\psi_2\rangle$

represent different states, is a ket that represents a continuum of different states as the complex numbers c_1 and c_2 are varied.

If we need to specify a state more closely – say we know that it corresponds to a plane wave with a wave number whose magnitude is k – then we indicate this with a label; the corresponding ket vector would be written as $|\,k\rangle$. If we also knew the direction of the wave then $|\,\mathbf{k}\rangle$ would be the appropriate form. Clearly, in general, the more labels we include, the more precisely the corresponding state is specified.

The Dirac notation for the Hermitian conjugate (dual vector) of the ket vector $|\,\psi\rangle$ is written as $\langle\psi|$ and is known as a *bra vector*; the wavefunction describing this state is ψ^*, the complex conjugate of ψ. The inner product of two wavefunctions $\int \psi^* \phi\, dv$ is then denoted by $\langle\psi|\,\phi\rangle$ or, more generally if a non-unit weight function ρ is involved, by

$$\langle\psi|\,\rho|\,\phi\rangle, \quad \text{evaluated as} \quad \int \psi^*(\mathbf{r})\phi(\mathbf{r})\rho(\mathbf{r})\,d\mathbf{r}. \tag{19.1}$$

Given the (contrived) names for the two sorts of vectors, an inner product like $\langle\psi|\,\phi\rangle$ becomes a particular type of 'bra(c)ket'. Despite its somewhat whimsical construction, this type of quantity has a fundamental role to play in the interpretation of quantum theory, because expectation values, probabilities and transition rates are all expressed in terms of them. For physical states the inner product of the corresponding ket with itself, with or without an explicit weight function, is non-zero, and it is usual to take

$$\langle\psi|\psi\rangle = 1.$$

Although multiplying a ket vector by a constant does not change the state described by the vector, acting upon it with a more general linear operator A results (in general) in a ket describing a different state. For example, if ψ is a state that is described in one-dimensional x-space by the wavefunction $\psi(x) = \exp(-x^2)$ and A is the differential operator $\partial/\partial x$, then

$$|\,\psi_1\rangle = A|\,\psi\rangle \equiv |\,A\psi\rangle$$

is the ket associated with the state whose wavefunction is $\psi_1(x) = -2x\exp(-x^2)$, clearly a different state. This allows us to attach a meaning to an expression such as $\langle\phi|A|\,\psi\rangle$ through the equation

$$\langle\phi|A|\,\psi\rangle = \langle\phi|\,\psi_1\rangle, \tag{19.2}$$

i.e. it is the inner product of $|\,\psi_1\rangle$ and $|\,\phi\rangle$. We have already used this notation in equation (19.1), but there the effect of the operator A was merely multiplication by a weight function.

If it should happen that the effect of an operator acting upon a particular ket is to produce a scalar multiple of that ket, i.e.

$$A|\psi\rangle = \lambda|\psi\rangle, \tag{19.3}$$

then, just as for matrices and differential equations, $|\psi\rangle$ is called an *eigenket* or, more usually, an *eigenstate* of A, with corresponding eigenvalue λ; to mark this special property the state will normally be denoted by $|\lambda\rangle$, rather than by the more general $|\psi\rangle$. Taking the Hermitian conjugate of this ket vector eigenequation gives a bra vector equation,

$$\langle\psi|A^\dagger = \lambda^*\langle\psi|. \tag{19.4}$$

It should be noted that the complex conjugate of the eigenvalue appears in this equation. Should the action of A on $|\psi\rangle$ produce an unphysical state (usually one whose wavefunction is identically zero, and is therefore unacceptable as a quantum-mechanical wavefunction because of the required probability interpretation) we denote the result either by 0 or by the ket vector $|\emptyset\rangle$ according to context. Formally, $|\emptyset\rangle$ can be considered as an eigenket of any operator, but one for which the eigenvalue is always zero.

If an operator A is Hermitian ($A^\dagger = A$) then its eigenvalues are real and the eigenstates can be chosen to be orthogonal; this can be shown in the same way as in chapter 17 (but using a different notation). As indicated there, the reality of their eigenvalues is one reason why Hermitian operators form the basis of measurement in quantum mechanics; in that formulation of physics, the eigenvalues of an operator are the *only* possible values that can be obtained when a measurement of the physical quantity corresponding to the operator is made. Actual individual measurements must always result in real values, even if they are combined in a complex form ($x + iy$ or $re^{i\theta}$) for final presentation or analysis, and using only Hermitian operators ensures this. The proof of the reality of the eigenvalues using the Dirac notation is given below in a worked example.

In the same notation the Hermitian property of an operator A is represented by the double equality

$$\langle A\,\phi|\psi\rangle = \langle\phi|A|\psi\rangle = \langle\phi|A\,\psi\rangle.$$

It should be remembered that the definition of an Hermitian operator involves specifying boundary conditions that the wavefunctions considered must satisfy. Typically, they are that the wavefunctions vanish for large values of the spatial variables upon which they depend; this deals with most physical systems since they are nearly all formally infinite in extent. Some model systems require the wavefunction to be periodic or to vanish at finite values of a spatial variable.

Depending on the nature of the physical system, the eigenvalues of a particular linear operator may be discrete, part of a continuum, or a mixture of both. For example, the energy levels of the bound proton–electron system (the hydrogen atom) are discrete, but if the atom is ionised and the electron is free, the energy spectrum of the system is continuous. This system has discrete negative and continuous positive eigenvalues for the operator corresponding to the total energy (the Hamiltonian).

►*Using the Dirac notation, show that the eigenvalues of an Hermitian operator are real.*

Let $|\,a\rangle$ be an eigenstate of Hermitian operator A corresponding to eigenvalue a, then

$$\begin{aligned} A|\,a\rangle &= a|\,a\rangle, \\ \Rightarrow \quad \langle a|A|\,a\rangle &= \langle a|a|\,a\rangle = a\langle a|\,a\rangle, \\ &\text{and} \\ \langle a|A^\dagger &= a^*\langle a|, \\ \Rightarrow \quad \langle a|A^\dagger|\,a\rangle &= a^*\langle a|\,a\rangle, \\ \langle a|A|\,a\rangle &= a^*\langle a|\,a\rangle, \quad \text{since } A \text{ is Hermitian.} \end{aligned}$$

Hence,

$$\begin{aligned} (a - a^*)\langle a|\,a\rangle &= 0, \\ \Rightarrow \quad a &= a^*, \ \text{since } \langle a|\,a\rangle \neq 0. \end{aligned}$$

Thus a is real. ◄

It is not our intention to describe the complete axiomatic basis of quantum mechanics, but rather to show what can be learned about linear operators, and in particular about their eigenvalues, without recourse to explicit wavefunctions on which the operators act.

Before we proceed to do that, we close this subsection with a number of results, expressed in Dirac notation, that the reader should verify by inspection or by following the lines of argument sketched in the statements. Where a sum over a complete set of eigenvalues is shown, it should be replaced by an integral for those parts of the eigenvalue spectrum that are continuous. With the notation that $|a_n\rangle$ is an eigenstate of Hermitian operator A with non-degenerate eigenvalue a_n (or, if a_n is k-fold degenerate, then a set of k mutually orthogonal eigenstates has been constructed and the states relabelled), we have the following results.

$$A|\,a_n\rangle = a_n|\,a_n\rangle,$$

$$\langle a_m|a_n\rangle = \delta_{mn} \quad \text{(orthonormality of eigenstates)}, \tag{19.5}$$

$$A(c_n|\,a_n\rangle + c_m|\,a_m\rangle) = c_n a_n|\,a_n\rangle + c_m a_m|\,a_m\rangle \quad \text{(linearity)}. \tag{19.6}$$

The definitions of the sum and product of two operators are

$$(A + B)|\,\psi\rangle \equiv A|\,\psi\rangle + B|\,\psi\rangle, \tag{19.7}$$

$$AB|\,\psi\rangle \equiv A(B|\,\psi\rangle) \quad (\neq BA|\,\psi\rangle \text{ in general}), \tag{19.8}$$

$$\Rightarrow \quad A^p|\,a_n\rangle = a_n^p|\,a_n\rangle. \tag{19.9}$$

If $A|\,a_n\rangle = a|\,a_n\rangle$ for all $N_1 \leq n \leq N_2$, then

$$|\,\psi\rangle = \sum_{n=N_1}^{N_2} d_n|\,a_n\rangle \text{ satisfies } A|\,\psi\rangle = a|\,\psi\rangle \text{ for any set of } d_i.$$

For a general state $|\,\psi\rangle$,

$$|\,\psi\rangle = \sum_{n=0}^{\infty} c_n|\,a_n\rangle, \text{ where } c_n = \langle a_n|\psi\rangle. \tag{19.10}$$

This can also be expressed as the operator identity,

$$1 = \sum_{n=0}^{\infty} |\,a_n\rangle\,\langle a_n|, \tag{19.11}$$

in the sense that

$$|\,\psi\rangle = 1\,|\,\psi\rangle = \sum_{n=0}^{\infty} |\,a_n\rangle\,\langle a_n|\psi\rangle = \sum_{n=0}^{\infty} c_n|\,a_n\rangle.$$

It also follows that

$$1 = \langle\psi|\psi\rangle = \left(\sum_{m=0}^{\infty} c_m^*\langle a_m|\right)\left(\sum_{n=0}^{\infty} c_n|a_n\rangle\right) = \sum_{m,n}^{\infty} c_m^* c_n \delta_{m\,n} = \sum_{n=0}^{\infty} |c_n|^2. \tag{19.12}$$

Similarly, the expectation value of the physical variable corresponding to A is

$$\begin{aligned}\langle\psi|A|\,\psi\rangle &= \sum_{m,n}^{\infty} c_m^*\langle a_m|\,A|\,a_n\rangle c_n = \sum_{m,n}^{\infty} c_m^*\langle a_m|\,a_n|\,a_n\rangle c_n \\ &= \sum_{m,n}^{\infty} c_m^* c_n a_n \delta_{m\,n} = \sum_{n=0}^{\infty} |c_n|^2 a_n.\end{aligned} \tag{19.13}$$

19.1.1 Commutation and commutators

As has been noted above, the product AB of two linear operators may or may not be equal to the product BA. That is

$$AB|\,\psi\rangle \text{ is not necessarily equal to } BA|\,\psi\rangle.$$

If A and B are both purely multiplicative operators, multiplication by $f(\mathbf{r})$ and $g(\mathbf{r})$ say, then clearly the order of the operations is immaterial, the result $|\,f(\mathbf{r})g(\mathbf{r})\psi\rangle$ being obtained in both cases. However, consider a case in which A is the differential operator $\partial/\partial x$ and B is the operator 'multiply by x'. Then the wavefunction describing $AB|\,\psi\rangle$ is

$$\frac{\partial}{\partial x}\,(x\psi(x)) = \psi(x) + x\frac{\partial\psi}{\partial x},$$

whilst that for $BA|\psi\rangle$ is simply

$$x\frac{\partial\psi}{\partial x},$$

which is not the same.

If the result

$$AB|\psi\rangle = BA|\psi\rangle$$

is true for *all* ket vectors $|\psi\rangle$, then A and B are said to *commute*; otherwise they are non-commuting operators.

A convenient way to express the commutation properties of two linear operators is to define their *commutator*, $[A, B]$, by

$$[A, B]|\psi\rangle \equiv AB|\psi\rangle - BA|\psi\rangle. \tag{19.14}$$

Clearly two operators that commute have a zero commutator. But, for the example given above we have that

$$\left[\frac{\partial}{\partial x}, x\right]\psi(x) = \left(\psi(x) + x\frac{\partial\psi}{\partial x}\right) - \left(x\frac{\partial\psi}{\partial x}\right) = \psi(x) = 1 \times \psi$$

or, more simply, that

$$\left[\frac{\partial}{\partial x}, x\right] = 1; \tag{19.15}$$

in words, the commutator of the differential operator $\partial/\partial x$ and the multiplicative operator x is the multiplicative operator 1. It should be noted that the order of the linear operators is important and that

$$[A, B] = -[B, A]. \tag{19.16}$$

Clearly any linear operator commutes with itself and some other obvious zero commutators (when operating on wavefunctions with 'reasonable' properties) are:

$[A, I]$, where I is the identity operator;

$[A^n, A^m]$, for any positive integers n and m;

$[A, p(A)]$, where $p(x)$ is any polynomial in x;

$[A, c]$, where A is any linear operator and c is any constant;

$[f(x), g(x)]$, where the functions are mutiplicative;

$[A(x), B(y)]$, where the operators act on different variables, with

$\left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right]$ as a specific example.

Simple identities amongst commutators include the following:

$$[A, B+C] = [A, B] + [A, C], \tag{19.17}$$

$$[A+B, C] = [A, C] + [B, C], \tag{19.18}$$

$$\begin{aligned}[A, BC] &= ABC - BCA + BAC - BAC \\ &= (AB - BA)C + B(AC - CA) \\ &= [A, B]\,C + B\,[A, C],\end{aligned} \tag{19.19}$$

$$[AB, C] = A\,[B, C] + [A, C]\,B. \tag{19.20}$$

►*If A and B are two linear operators that both commute with their commutator, prove that* $[A, B^n] = nB^{n-1}[A, B]$ *and that* $[A^n, B] = nA^{n-1}[A, B]$.

Define C_n by $C_n = [A, B^n]$. We aim to find a reduction formula for C_n:

$$\begin{aligned}C_n &= \left[A, B\,B^{n-1}\right] \\ &= [A, B]\,B^{n-1} + B\left[A, B^{n-1}\right], \text{ using (19.19)}, \\ &= B^{n-1}[A, B] + B\left[A, B^{n-1}\right], \text{ since } [[A, B], B] = 0, \\ &= B^{n-1}[A, B] + BC_{n-1}, \text{ the required reduction formula}, \\ &= B^{n-1}[A, B] + B\{B^{n-2}[A, B] + BC_{n-2}\}, \text{ applying the formula}, \\ &= 2B^{n-1}[A, B] + B^2C_{n-2} \\ &= \cdots \\ &= nB^{n-1}[A, B] + B^nC_0.\end{aligned}$$

However, $C_0 = [A, I] = 0$ and so $C_n = nB^{n-1}[A, B]$.

Using equation (19.16) and interchanging A and B in the result just obtained, we find

$$[A^n, B] = -[B, A^n] = -nA^{n-1}[B, A] = nA^{n-1}[A, B],$$

as stated in the question. ◄

As the power of a linear operator can be defined, so can its exponential; this situation parallels that for matrices, which are of course a particular set of operators that act upon state functions represented by vectors. The definition follows that for the exponential of a scalar or matrix, namely

$$\exp A = \sum_{n=0}^{\infty} \frac{A^n}{n!}. \tag{19.21}$$

Related functions of A, such as $\sin A$ and $\cos A$, can be defined in a similar way.

Since any linear operator commutes with itself, when two functions of it are combined in some way, the result takes a form similar to that for the corresponding functions of scalar quantities. Consider, for example, the function $f(A)$ defined by $f(A) = 2\sin A\cos A$. Expressing $\sin A$ and $\cos A$ in terms of their defining series, we have

$$f(A) = 2\sum_{m=0}^{\infty} \frac{(-1)^m A^{2m+1}}{(2m+1)!} \sum_{n=0}^{\infty} \frac{(-1)^n A^{2n}}{(2n)!}.$$

Writing $m+n$ as r and replacing n by s, we have

$$\begin{aligned} f(A) &= 2\sum_{r=0}^{\infty} A^{2r+1} \left(\sum_{s=0}^{r} \frac{(-1)^{r-s}}{(2r-2s+1)!} \frac{(-1)^s}{(2s)!} \right) \\ &= 2\sum_{r=0}^{\infty} (-1)^r c_r A^{2r+1}, \end{aligned}$$

where

$$c_r = \sum_{s=0}^{r} \frac{1}{(2r-2s+1)!\,(2s)!} = \frac{1}{(2r+1)!} \sum_{s=0}^{r} {}^{2r+1}C_{2s}.$$

By adding the binomial expansions of $2^{2r+1} = (1+1)^{2r+1}$ and $0 = (1-1)^{2r+1}$, it can easily be shown that

$$2^{2r+1} = 2\sum_{s=0}^{r} {}^{2r+1}C_{2s} \quad \Rightarrow \quad c_r = \frac{2^{2r}}{(2r+1)!}.$$

It then follows that

$$2\sin A \cos A = 2\sum_{r=0}^{\infty} \frac{(-1)^r A^{2r+1}\, 2^{2r}}{(2r+1)!} = \sum_{r=0}^{\infty} \frac{(-1)^r (2A)^{2r+1}}{(2r+1)!} = \sin 2A,$$

a not unexpected result.

However, if two (or more) linear operators that do not commute are involved, combining functions of them is more complicated and the results less intuitively obvious. We take as a particular case the product of two exponential functions and, even then, take the simplified case in which each linear operator commutes with their commutator (so that we may use the results from the previous worked example).

►*If A and B are two linear operators that both commute with their commutator, show that*

$$\exp(A)\exp(B) = \exp(A + B + \tfrac{1}{2}[A, B]).$$

We first find the commutator of A and $\exp \lambda B$, where λ is a scalar quantity introduced for later algebraic convenience:

$$\begin{aligned}
\left[A, e^{\lambda B}\right] &= \left[A, \sum_{n=0}^{\infty} \frac{(\lambda B)^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} [A, B^n] \\
&= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} n B^{n-1} [A, B], \text{ using the earlier result,} \\
&= \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} n B^{n-1} [A, B] \\
&= \lambda \sum_{m=0}^{\infty} \frac{\lambda^m B^m}{m!} [A, B], \text{ writing } m = n-1, \\
&= \lambda e^{\lambda B} [A, B].
\end{aligned}$$

Now consider the derivative with respect to λ of the function

$$f(\lambda) = e^{\lambda A} e^{\lambda B} e^{-\lambda(A+B)}.$$

In the following calculation we use the fact that the derivative of $e^{\lambda C}$ is $Ce^{\lambda C}$; this is the same as $e^{\lambda C}C$, since any two functions of the same operator commute. Differentiating the three-factor product gives

$$\begin{aligned}
\frac{df}{d\lambda} &= e^{\lambda A} A e^{\lambda B} e^{-\lambda(A+B)} + e^{\lambda A} e^{\lambda B} B e^{-\lambda(A+B)} + e^{\lambda A} e^{\lambda B} (-A - B) e^{-\lambda(A+B)} \\
&= e^{\lambda A} (e^{\lambda B} A + \lambda e^{\lambda B} [A, B]) e^{-\lambda(A+B)} + e^{\lambda A} e^{\lambda B} B e^{-\lambda(A+B)} \\
&\quad - e^{\lambda A} e^{\lambda B} A e^{-\lambda(A+B)} - e^{\lambda A} e^{\lambda B} B e^{-\lambda(A+B)} \\
&= e^{\lambda A} \lambda e^{\lambda B} [A, B] e^{-\lambda(A+B)} \\
&= \lambda [A, B] f(\lambda).
\end{aligned}$$

In the second line we have used the result obtained above to replace $Ae^{\lambda B}$, and in the last line have used the fact that $[A, B]$ commutes with each of A and B, and hence with any function of them.

Integrating this scalar differential equation with respect to λ and noting that $f(0) = 1$, we obtain

$$\ln f = \tfrac{1}{2}\lambda^2 [A, B] \quad \Rightarrow \quad e^{\lambda A} e^{\lambda B} e^{-\lambda(A+B)} = f(\lambda) = e^{\frac{1}{2}\lambda^2 [A,B]}.$$

Finally, post-multiplying both sides of the equation by $e^{\lambda(A+B)}$ and setting $\lambda = 1$ yields

$$e^A e^B = e^{\frac{1}{2}[A,B]+A+B}. \blacktriangleleft$$

19.2 Physical examples of operators

We now turn to considering some of the specific linear operators that play a part in the description of physical systems. In particular, we will examine the properties of some of those that appear in the quantum-mechanical description of the physical world.

As stated earlier, the operators corresponding to physical observables are restricted to Hermitian operators (which have real eigenvalues) as this ensures the reality of predicted values for experimentally measured quantities. The two basic quantum-mechanical operators are those corresponding to position $\mathbf{r}$ and momentum $\mathbf{p}$. One prescription for making the transition from classical to quantum mechanics is to express classical quantities in terms of these two variables in Cartesian coordinates and then make the component by component substitutions

$$\mathbf{r} \to \text{multiplicative operator } \mathbf{r} \quad \text{and} \quad \mathbf{p} \to \text{differential operator } -i\hbar\nabla. \tag{19.22}$$

This generates the quantum operators corresponding to the classical quantities. For the sake of completeness, we should add that if the classical quantity contains a product of factors whose corresponding operators A and B do not commute, then the operator $\frac{1}{2}(AB + BA)$ is to be substituted for the product.

The substitutions (19.22) invoke operators that are closely connected with the two that we considered at the start of the previous subsection, namely x and $\partial/\partial x$. One, x, corresponds exactly to the x-component of the prescribed quantum position operator; the other, however, has been multiplied by the imaginary constant $-i\hbar$, where $\hbar$ is the Planck constant divided by 2π. This has the (subtle) effect of converting the differential operator into the x-component of an *Hermitian* operator; this is easily verified using integration by parts to show that it satisfies equation (17.16). Without the extra imaginary factor (which changes sign under complex conjugation) the two sides of the equation differ by a minus sign.

Making the differential operator Hermitian does not change in any essential way its commutation properties, and the commutation relation of the two basic quantum operators reads

$$[\,p_x, x\,] = \left[-i\hbar\frac{\partial}{\partial x}, x\right] = -i\hbar. \tag{19.23}$$

Corresponding results hold when x is replaced, in both operators, by y or z. However, it should be noted that if different Cartesian coordinates appear in the two operators then the operators commute, i.e.

$$[\,p_x, y\,] = [\,p_x, z\,] = \left[\,p_y, x\,\right] = \left[\,p_y, z\,\right] = [\,p_z, x\,] = [\,p_z, y\,] = 0. \tag{19.24}$$

As an illustration of the substitution rules, we now construct the Hamiltonian (the quantum-mechanical energy operator) H for a particle of mass m moving in a potential $V(x, y, z)$ when it has one of its allowed energy values, i.e its energy is E_n, where $H|\psi_n\rangle = E_n|\psi_n\rangle$. This latter equation when expressed in a particular coordinate system is the Schrödinger equation for the particle. In terms of position and momentum, the total classical energy of the particle is given by

$$E = \frac{p^2}{2m} + V(x, y, z) = \frac{p_x^2 + p_y^2 + p_z^2}{2m} + V(x, y, z).$$

Substituting $-i\hbar\partial/\partial x$ for p_x (and similarly for p_y and p_z) in the first term on the

RHS gives

$$\frac{(-i\hbar)^2}{2m}\frac{\partial}{\partial x}\frac{\partial}{\partial x}+\frac{(-i\hbar)^2}{2m}\frac{\partial}{\partial y}\frac{\partial}{\partial y}+\frac{(-i\hbar)^2}{2m}\frac{\partial}{\partial z}\frac{\partial}{\partial z}.$$

The potential energy V, being a function of position only, becomes a purely multiplicative operator, thus creating the full expression for the Hamiltonian,

$$H=-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+V(x,y,z),$$

and giving the corresponding Schrödinger equation as

$$H\psi_n=-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi_n}{\partial x^2}+\frac{\partial^2\psi_n}{\partial y^2}+\frac{\partial^2\psi_n}{\partial z^2}\right)+V(x,y,z)\psi_n=E_n\psi_n.$$

We are not so much concerned in this section with solving such differential equations, but with the commutation properties of the operators from which they are constructed. To this end, we now turn our attention to the topic of angular momentum, the operators for which can be constructed in a straightforward manner from the two basic sets.

19.2.1 Angular momentum operators

As required by the substitution rules, we start by expressing angular momentum in terms of the classical quantities $\mathbf{r}$ and $\mathbf{p}$, namely $\mathbf{L}=\mathbf{r}\times\mathbf{p}$ with Cartesian components

$$L_z=xp_y-yp_x,\quad L_x=yp_z-zp_y,\quad L_y=zp_x-xp_z.$$

Making the substitutions (19.22) yields as the corresponding quantum-mechanical operators

$$\begin{aligned}L_z&=-i\hbar\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right),\\L_x&=-i\hbar\left(y\frac{\partial}{\partial z}-z\frac{\partial}{\partial y}\right),\\L_y&=-i\hbar\left(z\frac{\partial}{\partial x}-x\frac{\partial}{\partial z}\right).\end{aligned}\tag{19.25}$$

It should be noted that for xp_y, say, x and $\partial/\partial y$ commute, and there is no ambiguity about the way it is to be carried into its quantum form. Further, since the operators corresponding to each of its factors commute and are Hermitian, the operator corresponding to the product is Hermitian. This was shown directly for matrices in exercise 8.7, and can be verified using equation (17.16).

The first question that arises is whether or not these three operators commute.

Consider first

$$\begin{aligned} L_x L_y &= -\hbar^2 \left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y} \right) \left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z} \right) \\ &= -\hbar^2 \left(y\frac{\partial}{\partial x} + yz\frac{\partial^2}{\partial z \partial x} - yx\frac{\partial^2}{\partial z^2} - z^2\frac{\partial^2}{\partial y \partial x} + zx\frac{\partial^2}{\partial y \partial z} \right). \end{aligned}$$

Now consider

$$\begin{aligned} L_y L_x &= -\hbar^2 \left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z} \right) \left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y} \right) \\ &= -\hbar^2 \left(zy\frac{\partial^2}{\partial x \partial z} - z^2\frac{\partial^2}{\partial x \partial y} - xy\frac{\partial^2}{\partial z^2} + x\frac{\partial}{\partial y} + xz\frac{\partial^2}{\partial z \partial y} \right). \end{aligned}$$

These two expressions are *not* the same. The difference between them, i.e. the commutator of L_x and L_y, is given by

$$\left[L_x, L_y \right] = L_x L_y - L_y L_x = \hbar^2 \left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} \right) = i\hbar L_z. \tag{19.26}$$

This, and two similar results obtained by permutting x, y and z cyclically, summarise the commutation relationships between the quantum operators corresponding to the three Cartesian components of angular momentum:

$$\begin{aligned} \left[L_x, L_y \right] &= i\hbar L_z, \\ \left[L_y, L_z \right] &= i\hbar L_x, \\ \left[L_z, L_x \right] &= i\hbar L_y. \end{aligned} \tag{19.27}$$

As well as its separate components of angular momentum, the total angular momentum associated with a particular state $|\psi\rangle$ is a physical quantity of interest. This is measured by the operator corresponding to the sum of squares of its components,

$$L^2 = L_x^2 + L_y^2 + L_z^2. \tag{19.28}$$

This is an Hermitian operator, as each term in it is the product of two Hermitian operators that (trivially) commute. It might seem natural to want to 'take the square root' of this operator, but such a process is undefined and we will not pursue the matter.

We next show that, although no two of its components commute, the total angular momentum operator does commute with each of its components. In the proof we use some of the properties (19.17) to (19.20) and result (19.27). We begin with

$$\begin{aligned}\left[L^2, L_z\right] &= \left[L_x^2 + L_y^2 + L_z^2, L_z\right] \\ &= L_x\left[L_x, L_z\right] + \left[L_x, L_z\right]L_x \\ &\qquad + L_y\left[L_y, L_z\right] + \left[L_y, L_z\right]L_y + \left[L_z^2, L_z\right] \\ &= L_x(-i\hbar)L_y + (-i\hbar)L_yL_x + L_y(i\hbar)L_x + (i\hbar)L_xL_y + 0 \\ &= 0.\end{aligned}$$

Thus operators L^2 and L_z commute and, continuing in the same way, it can be shown that

$$\left[L^2, L_x\right] = \left[L^2, L_y\right] = \left[L^2, L_z\right] = 0. \tag{19.29}$$

Eigenvalues of the angular momentum operators

We will now use the commutation relations for L^2 and its components to find the eigenvalues of L^2 and L_z, without reference to any specific wavefunction. In other words, the eigenvalues of the operators follow from the structure of their commutators. There is nothing particular about L_z, and L_x or L_y could equally well have been chosen, though, in general, it is not possible to find states that are simultaneously eigenstates of two or more of L_x, L_y and L_z.

To help with the calculation, it is convenient to define the two operators

$$U \equiv L_x + iL_y \quad \text{and} \quad D \equiv L_x - iL_y.$$

These operators are not Hermitian; they are in fact Hermitian conjugates, in that $U^\dagger = D$ and $D^\dagger = U$, but they do not represent measurable physical quantities. We first note their multiplication and commutation properties:

$$\begin{aligned}UD &= (L_x + iL_y)(L_x - iL_y) = L_x^2 + L_y^2 + i\left[L_y, L_x\right] \\ &= L^2 - L_z^2 + \hbar L_z, \end{aligned}\tag{19.30}$$

$$\begin{aligned}DU &= (L_x - iL_y)(L_x + iL_y) = L_x^2 + L_y^2 - i\left[L_y, L_x\right] \\ &= L^2 - L_z^2 - \hbar L_z, \end{aligned}\tag{19.31}$$

$$\left[L_z, U\right] = \left[L_z, L_x\right] + i\left[L_z, L_y\right] = i\hbar L_y + \hbar L_x = \hbar U, \tag{19.32}$$

$$\left[L_z, D\right] = \left[L_z, L_x\right] - i\left[L_z, L_y\right] = i\hbar L_y - \hbar L_x = -\hbar D. \tag{19.33}$$

In the same way as was shown for matrices, it can be demonstrated that if two operators commute they have a common set of eigenstates. Since L^2 and L_z commute they possess such a set; let one of the set be $|\psi\rangle$ with

$$L^2|\psi\rangle = a|\psi\rangle \quad \text{and} \quad L_z|\psi\rangle = b|\psi\rangle.$$

Now consider the state $|\psi'\rangle = U|\psi\rangle$ and the actions of L^2 and L_z upon it.

Consider first $L^2|\psi'\rangle$, recalling that L^2 commutes with both L_x and L_y and hence with U:

$$L^2|\psi'\rangle = L^2U|\psi\rangle = UL^2|\psi\rangle = Ua|\psi\rangle = aU|\psi\rangle = a|\psi'\rangle.$$

Thus, $|\psi'\rangle$ is also an eigenstate of L^2, corresponding to the same eigenvalue as $|\psi\rangle$. Now consider the action of L_z:

$$\begin{aligned}
L_z|\psi'\rangle &= L_zU|\psi\rangle \\
&= (UL_z + \hbar U)|\psi\rangle, \text{ using } [\,L_z, U\,] = \hbar U, \\
&= Ub|\psi\rangle + \hbar U|\psi\rangle \\
&= (b+\hbar)U|\psi\rangle \\
&= (b+\hbar)|\psi'\rangle.
\end{aligned}$$

Thus, $|\psi'\rangle$ is also an eigenstate of L_z, but with eigenvalue $b + \hbar$.

In summary, the effect of U acting upon $|\psi\rangle$ is to produce a new state that has the same eigenvalue for L^2 and is still an eigenstate of L_z, though with that eigenvalue increased by $\hbar$. An exactly analogous calculation shows that the effect of D acting upon $|\psi\rangle$ is to produce another new state, one that also has the same eigenvalue for L^2 and is also still an eigenstate of L_z, though with the eigenvalue decreased by $\hbar$ in this case. For these reasons, U and D are usually known as *ladder* operators.

It is clear that, by starting from any arbitrary eigenstate and repeatedly applying either U or D, we could generate a series of eigenstates, all of which have the eigenvalue a for L^2, but increment in their L_z eigenvalues by $\pm\hbar$. However, we also have the physical requirement that, for real values of the z-component, its square cannot exceed the square of the total angular momentum, i.e. $b^2 \le a$. Thus b has a maximum value c that satisfies

$$c^2 \le a \quad \text{but} \quad (c+\hbar)^2 > a;$$

let the corresponding eigenstate be $|\psi_u\rangle$ with $L_z|\psi_u\rangle = c|\psi_u\rangle$. Now it is still true that

$$L_zU|\psi_u\rangle = (c+\hbar)U|\psi_u\rangle,$$

and, to make this compatible with the physical constraint, we must have that $U|\psi_u\rangle$ is the zero ket vector $|\,\emptyset\,\rangle$. Now, using result (19.31), we have

$$\begin{aligned}
DU|\psi_u\rangle &= (L^2 - L_z^2 - \hbar L_z)|\psi_u\rangle, \\
\Rightarrow \quad 0|\,\emptyset\,\rangle = D|\,\emptyset\,\rangle &= (a^2 - c^2 - \hbar c)|\psi_u\rangle, \\
\Rightarrow \quad a &= c(c+\hbar).
\end{aligned}$$

This gives the relationship between a and c. We now establish the possible forms for c.

If we start with eigenstate $|\psi_u\rangle$, which has the highest eigenvalue c for L_z, and operate repeatedly on it with the (down) ladder operator D, we will generate a state $|\psi_d\rangle$ which, whilst still an eigenstate of L^2 with eigenvalue a, has the lowest physically possible value, d say, for the eigenvalue of L_z. If this happens after n operations we will have that $d = c - n\hbar$ and

$$L_z|\psi_d\rangle = (c - n\hbar)|\psi_d\rangle.$$

Arguing in the same way as previously that $D|\psi_d\rangle$ must be an unphysical ket vector, we conclude that

$$\begin{aligned} 0|\,\emptyset\,\rangle = U|\,\emptyset\,\rangle &= UD|\psi_d\rangle \\ &= (L^2 - L_z^2 + \hbar L_z)|\psi_d\rangle, \text{ using (19.30)}, \\ &= [\,a - (c - n\hbar)^2 + \hbar(c - n\hbar)\,]|\psi_d\rangle \\ \Rightarrow \quad a &= (c - n\hbar)^2 - \hbar(c - n\hbar). \end{aligned}$$

Equating the two results for a gives

$$\begin{aligned} c^2 + c\hbar &= c^2 - 2cn\hbar + n^2\hbar^2 - c\hbar + n\hbar^2, \\ 2c(n+1) &= n(n+1)\hbar, \\ c &= \tfrac{1}{2}n\hbar. \end{aligned}$$

Since n is necessarily integral, c is an integer multiple of $\frac{1}{2}\hbar$. This result is valid irrespective of which eigenstate $|\psi\rangle$ we started with, though the actual value of the integer n depends on $|\psi_u\rangle$ and hence upon $|\psi\rangle$.

Denoting $\frac{1}{2}n$ by ℓ we can say that the possible eigenvalues of the operator L_z, and hence the possible results of a measurement of the z-component of the angular momentum of a system, are given by

$$\ell\hbar, \quad (\ell - 1)\hbar, \quad (\ell - 2)\hbar, \quad \ldots, \quad -\ell\hbar.$$

The value of a for all $2\ell + 1$ of the corresponding states,

$$|\psi_u\rangle, \quad D|\psi_u\rangle, \quad D^2|\psi_u\rangle, \quad \ldots, \quad D^{2\ell}|\psi_u\rangle,$$

is $\ell(\ell + 1)\hbar^2$.

The similarity of form between this eigenvalue and that appearing in Legendre's equation is not an accident. It is intimately connected with the facts (i) that L^2 is a measure of the rotational kinetic energy of a particle in a system centred on the origin, and (ii) that in spherical polar coordinates L^2 has the same form as the angle-dependent part of ∇^2, which, as we have seen, is itself proportional to the quantum-mechanical kinetic energy operator. Legendre's equation and the associated Legendre equation arise naturally when $\nabla^2\psi = f(r)$ is solved in spherical polar coordinates using the method of separation of variables discussed in chapter 21.

The derivation of the eigenvalues $\ell(\ell+1)\hbar^2$ and $m\hbar$, with $-\ell \leq m \leq \ell$, depends only on the commutation relationships between the corresponding operators. Any other set of four operators with the same commutation structure would result in the same eigenvalue spectrum. In fact, quantum mechanically, orbital angular momentum is restricted to cases in which n is even and so ℓ is an integer; this is in accord with the requirement placed on ℓ if solutions to $\nabla^2\psi = f(r)$ that are finite on the polar axis are to be obtained. The non-classical notion of internal angular momentum (spin) for a particle provides a set of operators that are able to take both integral and half-integral multiples of $\hbar$ as their eigenvalues.

We have already seen that, for a state $|\ell, m\rangle$ that has a z-component of angular momentum $m\hbar$, the state $U|\ell, m\rangle$ is one with its z-component of angular momentum equal to $(m+1)\hbar$. But the new state ket vector so produced is not necessarily normalised so as to make $\langle \ell, m+1 \,|\, \ell, m+1\rangle = 1$. We will conclude this discussion of angular momentum by calculating the coefficients μ_m and ν_m in the equations

$$U|\ell, m\rangle = \mu_m|\ell, m+1\rangle \quad \text{and} \quad D|\ell, m\rangle = \nu_m|\ell, m-1\rangle$$

on the basis that $\langle \ell, r \,|\, \ell, r\rangle = 1$ for all ℓ and r.

To do so, we consider the inner product $I = \langle \ell, m\,|DU|\,\ell, m\rangle$, evaluated in two different ways. We have already noted that U and D are Hermitian conjugates and so I can be written as

$$I = \langle \ell, m\,|U^\dagger U|\,\ell, m\rangle = \mu_m^*\langle \ell, m\,|\,\ell, m\rangle\mu_m = |\mu_m|^2.$$

But, using equation (19.31), it can also be expressed as

$$\begin{aligned} I &= \langle \ell, m\,|L^2 - L_z^2 - \hbar L_z|\,\ell, m\rangle \\ &= \langle \ell, m\,|\ell(\ell+1)\hbar^2 - m^2\hbar^2 - m\hbar^2|\,\ell, m\rangle \\ &= [\,\ell(\ell+1)\hbar^2 - m^2\hbar^2 - m\hbar^2\,]\,\langle \ell, m\,|\,\ell, m\rangle \\ &= [\,\ell(\ell+1) - m(m+1)\,]\,\hbar^2. \end{aligned}$$

Thus we are required to have

$$|\mu_m|^2 = [\,\ell(\ell+1) - m(m+1)\,]\,\hbar^2,$$

but can choose that all μ_m are real and non-negative (recall that $|m| \leq \ell$). A similar calculation can be used to calculate ν_m. The results are summarised in the equations

$$U|\,\ell, m\rangle = \sqrt{\ell(\ell+1) - m(m+1)}\,\hbar|\,\ell, m+1\rangle, \tag{19.34}$$

$$D|\,\ell, m\rangle = \sqrt{\ell(\ell+1) - m(m-1)}\,\hbar|\,\ell, m-1\rangle. \tag{19.35}$$

It can easily be checked that $U|\ell, \ell\rangle = |\,\emptyset\,\rangle = D|\ell, -\ell\rangle$.

19.2.2 Uncertainty principles

The next topic we explore is the quantitative consequences of a non-zero commutator for two quantum (Hermitian) operators that correspond to physical variables.

As previously noted, the expectation value in a state $|\psi\rangle$ of the physical quantity A corresponding to the operator A is $E[A] = \langle\psi|\,A\,|\psi\rangle$. Any one measurement of A can only yield one of the eigenvalues of A. But if repeated measurements could be made on a large number of identical systems, a discrete or continuous range of values would be obtained. It is a natural extension of normal data analysis to measure the uncertainty in the value of A by the observed variance in the measured values of A, denoted by $(\Delta A)^2$ and calculated as the average value of $(A - E[A])^2$. The expected value of this variance for the state $|\psi\rangle$ is given by $\langle\psi\,|(A - E[A]\,)^2|\psi\rangle$.

We now give a mathematical proof that there is a theoretical lower limit for the product of the uncertainties in any two physical quantities, and we start by proving a result similar to the Schwarz inequality. Let $|u\rangle$ and $|v\rangle$ be any two state vectors and let λ be any *real* scalar. Then consider the vector $|w\rangle = |u\rangle + \lambda|v\rangle$ and, in particular, note that

$$0 \le \langle w\,|\,w\rangle = \langle u\,|\,u\rangle + \lambda(\langle u\,|\,v\rangle + \langle v\,|\,u\rangle) + \lambda^2\langle v\,|\,v\rangle.$$

This is a quadratic inequality in λ and therefore the quadratic equation formed by equating the RHS to zero must have no real roots. The coefficient of λ is $(\langle u\,|\,v\rangle + \langle v\,|\,u\rangle) = 2\,\mathrm{Re}\,\langle u\,|\,v\rangle$ and its square is thus ≥ 0. The condition for no real roots of the quadratic is therefore

$$0 \le (\langle u\,|\,v\rangle + \langle v\,|\,u\rangle)^2 \le 4\langle u\,|\,u\rangle\,\langle v\,|\,v\rangle. \tag{19.36}$$

This result will now be applied to state vectors constructed from $|\psi\rangle$, the state vector of the particular system for which we wish to establish a relationship between the uncertainties in the two physical variables corresponding to (Hermitian) operators A and B. We take

$$|u\rangle = (A - E[A])\,|\,\psi\rangle \quad \text{and} \quad |v\rangle = i(B - E[B])\,|\,\psi\rangle. \tag{19.37}$$

Then

$$\begin{aligned}\langle u\,|\,u\rangle &= \langle\psi\,|(A - E[A]\,)^2|\psi\rangle = (\Delta A)^2,\\ \langle v\,|\,v\rangle &= \langle\psi\,|(B - E[B]\,)^2|\psi\rangle = (\Delta B)^2.\end{aligned}$$

Further,

$$\begin{aligned}\langle u\,|\,v\rangle &= \langle\psi\,|\,(A - E[A])i(B - E[B])\,|\,\psi\rangle\\ &= i\langle\psi\,|\,AB\,|\,\psi\rangle - iE[A]\langle\psi\,|\,B\,|\,\psi\rangle - iE[B]\langle\psi\,|\,A\,|\,\psi\rangle + iE[A]E[B]\langle\psi\,|\,\psi\rangle\\ &= i\langle\psi\,|\,AB\,|\,\psi\rangle - iE[A]E[B].\end{aligned}$$

In the second line, we have moved expectation values, which are purely numbers, out of the inner products and used the normalisation condition $\langle\psi|\psi\rangle = 1$. Similarly

$$\langle v \mid u\rangle = -i\langle\psi \mid BA \mid \psi\rangle + iE[A]E[B].$$

Adding these two results gives

$$\langle u \mid v\rangle + \langle v \mid u\rangle = i\langle\psi \mid AB - BA \mid\psi\rangle,$$

and substitution into (19.36) yields

$$0 \le (i\langle\psi \mid AB - BA \mid\psi\rangle)^2 \le 4(\Delta A)^2(\Delta B)^2$$

At first sight, the middle term of this inequality might appear to be negative, but this is not so. Since A and B are Hermitian, $AB-BA$ is anti-Hermitian, as is easily demonstrated. Since i is also anti-Hermitian, the quantity in the parentheses in the middle term is real and its square non-negative. Rearranging the equation and expressing it in terms of the commutator of A and B gives the generalised form of the *Uncertainty Principle*. For any particular state $|\psi\rangle$ of a system, this provides the quantitative relationship between the minimum value that the product of the uncertainties in A and B can have and the expectation value, in that state, of their commutator,

$$(\Delta A)^2(\Delta B)^2 \ge \tfrac{1}{4}|\,\langle\psi \mid [\,A, B\,] \mid \psi\rangle\,|^2. \tag{19.38}$$

Immediate observations include the following:

(i) If A and B commute there is no absolute restriction on the accuracy with which the corresponding physical quantities may be known. That is not to say that ΔA and ΔB will always be zero, only that they may be.
(ii) If the commutator of A and B is a constant, $k \neq 0$, then the RHS of equation (19.38) is necessarily equal to $\frac{1}{4}|k|^2$, whatever the form of $|\psi\rangle$, and it is not possible to have $\Delta A = \Delta B = 0$.
(iii) Since the RHS depends upon $|\psi\rangle$, it is possible, even for two operators that do not commute, for the lower limit of $(\Delta A)^2(\Delta B)^2$ to be zero. This will occur if the commutator $[\,A, B\,]$ is itself an operator whose expectation value in the particular state $|\psi\rangle$ happens to be zero.

To illustrate the third of these, we might consider the components of angular momentum discussed in the previous subsection. There, in equation (19.27), we found that the commutator of the operators corresponding to the x- and y-components of angular momentum is non-zero; in fact, it has the value $i\hbar L_z$. This means that if the state $|\psi\rangle$ of a system happened to be such that $\langle\psi|L_z|\psi\rangle = 0$, as it would if, for example, it were the eigenstate of L_z, $|\psi\rangle = |\ell, 0\rangle$, then there would be no fundamental reason why the physical values of both L_x and L_y should not be known exactly. Indeed, if the state were spherically symmetric, and hence formally an eigenstate of L^2 with $\ell = 0$, all three components of angular momentum could be (and are) known to be zero.

►*Working in one dimension, show that the minimum value of the product $\Delta p_x \times \Delta x$ for a particle is $\frac{1}{2}\hbar$. Find the form of the wavefunction that attains this minimum value for a particle whose expectation values for position and momentum are $\bar{x}$ and $\bar{p}$, respectively.*

We have already seen, in (19.23) that the commutator of p_x and x is $-i\hbar$, a constant. Therefore, irrespective of the actual form of $|\psi\rangle$, the RHS of (19.38) is $\frac{1}{4}\hbar^2$ (see observation (ii) above). Thus, since all quantities are positive, taking the square roots of both sides of the equation shows directly that

$$\Delta p_x \times \Delta x \geq \tfrac{1}{2}\hbar.$$

Returning to the derivation of the Uncertainty Principle, we see that the inequality becomes an equality only when

$$(\langle u \,|\, v\rangle + \langle v \,|\, u\rangle)^2 = 4\langle u \,|\, u\rangle \,\langle v \,|\, v\rangle.$$

The RHS of this equality has the value $4||u||^2||v||^2$ and so, by virtue of Schwarz's inequality, we have

$$\begin{aligned}
4\|u\|^2\|v\|^2 &= (\langle u \,|\, v\rangle + \langle v \,|\, u\rangle)^2 \\
&\leq (|\langle u \,|\, v\rangle| + |\langle v \,|\, u\rangle|)^2 \\
&\leq (\|u\| \,\|v\| + \|v\| \,\|u\|)^2 \\
&= 4\|u\|^2\|v\|^2.
\end{aligned}$$

Since the LHS is less than or equal to something that has the same value as itself, all of the inequalities are, in fact, equalities. Thus $\langle u|v\rangle = \|u\| \,\|v\|$, showing that $|u\rangle$ and $|v\rangle$ are parallel vectors, i.e. $|u\rangle = \mu|v\rangle$ for some scalar μ.

We now transform this condition into a constraint that the wavefunction $\psi = \psi(x)$ must satisfy. Recalling the definitions (19.37) of $|u\rangle$ and $|v\rangle$ in terms of $|\psi\rangle$, we have

$$\begin{aligned}
\left(-i\hbar\frac{d}{dx} - \bar{p}\right)\psi &= \mu i(x - \bar{x})\psi, \\
\frac{d\psi}{dx} + \frac{1}{\hbar}[\,\mu(x - \bar{x}) - i\bar{p}\,]\psi &= 0.
\end{aligned}$$

The IF for this equation is $\exp\left[\dfrac{\mu(x-\bar{x})^2}{2\hbar} - \dfrac{i\bar{p}x}{\hbar}\right]$, giving

$$\frac{d}{dx}\left\{\psi \exp\left[\frac{\mu(x-\bar{x})^2}{2\hbar} - \frac{i\bar{p}x}{\hbar}\right]\right\} = 0,$$

which, in turn, leads to

$$\psi(x) = A\exp\left[-\frac{\mu(x-\bar{x})^2}{2\hbar}\right]\exp\left(\frac{i\bar{p}x}{\hbar}\right).$$

From this it is apparent that the minimum uncertainty product $\Delta p_x \times \Delta x$ is obtained when the probability density $|\psi(x)|^2$ has the form of a Gaussian distribution centred on $\bar{x}$. The value of μ is not fixed by this consideration and it could be anything (positive); a large value for μ would yield a small value for Δx but a correspondingly large one for Δp_x. ◄

19.2.3 Annihilation and creation operators

As a final illustration of the use of operator methods in physics we consider their application to the quantum mechanics of a simple harmonic oscillator (s.h.o.). Although we will start with the conventional description of a one-dimensional oscillator, using its position and momentum, we will recast the description in terms of two operators and their commutator and show that many important conclusions can be reached from studying these alone.

The Hamiltonian for a particle of mass m with momentum p moving in a one-dimensional parabolic potential $V(x) = \frac{1}{2}kx^2$ is

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2 = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2,$$

where its classical frequency of oscillation ω is given by $\omega^2 = k/m$. We recall that the corresponding operators, p and x, do not commute and that $[p, x] = -i\hbar$.

In analogy with the ladder operators used when discussing angular momentum, we define two new operators:

$$A \equiv \sqrt{\frac{m\omega}{2}}\,x + \frac{ip}{\sqrt{2m\omega}} \quad \text{and} \quad A^\dagger \equiv \sqrt{\frac{m\omega}{2}}\,x - \frac{ip}{\sqrt{2m\omega}}. \tag{19.39}$$

Since both x and p are Hermitian, A and $A^\dagger$ are Hermitian conjugates, though neither is Hermitian and they do not represent physical quantities that can be measured.

Now consider the two products $A^\dagger A$ and $AA^\dagger$:

$$\begin{aligned} A^\dagger A &= \frac{m\omega}{2}x^2 - \frac{ipx}{2} + \frac{ixp}{2} + \frac{p^2}{2m\omega} = \frac{H}{\omega} - \frac{i}{2}[p, x] = \frac{H}{\omega} - \frac{\hbar}{2}, \\ AA^\dagger &= \frac{m\omega}{2}x^2 + \frac{ipx}{2} - \frac{ixp}{2} + \frac{p^2}{2m\omega} = \frac{H}{\omega} + \frac{i}{2}[p, x] = \frac{H}{\omega} + \frac{\hbar}{2}. \end{aligned}$$

From these it follows that

$$H = \tfrac{1}{2}\omega(A^\dagger A + AA^\dagger) \tag{19.40}$$

and that

$$\left[A, A^\dagger\right] = \hbar. \tag{19.41}$$

Further,

$$\begin{aligned} [H, A] &= \left[\tfrac{1}{2}\omega(A^\dagger A + AA^\dagger), A\right] \\ &= \tfrac{1}{2}\omega\left(A^\dagger\, 0 + \left[A^\dagger, A\right]A + A\left[A^\dagger, A\right] + 0\, A^\dagger\right) \\ &= \tfrac{1}{2}\omega(-\hbar A - A\hbar) = -\hbar\omega A. \end{aligned} \tag{19.42}$$

Similarly,

$$\left[H, A^\dagger\right] = \hbar\omega A^\dagger \tag{19.43}$$

Before we apply these relationships to the question of the energy spectrum of the s.h.o., we need to prove one further result. This is that if B is an Hermitian operator then $\langle\psi\,|\,B^2\,|\,\psi\rangle \geq 0$ for *any* $|\psi\rangle$. The proof, which involves introducing an arbitrary complete set of orthonormal base states $|\phi_i\rangle$ and using equation (19.11), is as follows:

$$\begin{aligned}\langle\psi\,|\,B^2\,|\,\psi\rangle &= \langle\psi\,|\,B\times 1\ \times B\,|\,\psi\,\rangle \\ &= \sum_i \langle\psi\,|\,B\,|\phi_i\rangle\langle\phi_i\,|\,B\,|\psi\rangle \\ &= \sum_i \langle\psi\,|\,B\,|\phi_i\rangle\left(\langle\phi_i\,|\,B\,|\psi\rangle^*\right)^* \\ &= \sum_i \langle\psi\,|\,B\,|\phi_i\rangle\left(\langle\psi\,|\,B^\dagger\,|\phi_i\rangle\right)^* \\ &= \sum_i \langle\psi\,|\,B\,|\phi_i\rangle\langle\psi\,|\,B\,|\phi_i\rangle^*, \qquad \text{since } B \text{ is Hermitian,} \\ &= \sum_i |\,\langle\psi\,|\,B\,|\phi_i\rangle\,|^2 \geq 0.\end{aligned}$$

We note, for future reference, that the Hamiltonian H for the s.h.o. is the sum of two terms each of this form and therefore conclude that $\langle\psi|H|\psi\rangle \geq 0$ for all $|\psi\rangle$.

The energy spectrum of the simple harmonic oscillator

Let the normalised ket vector $|n\rangle$ (or $|E_n\rangle$) denote the nth energy state of the s.h.o. with energy E_n. Then it must be an eigenstate of the (Hermitian) Hamiltonian H and satisfy

$$H|n\rangle = E_n|n\rangle \text{ with } \langle m|n\rangle = \delta_{mn}.$$

Now consider the state $A|n\rangle$ and the effect of H upon it:

$$\begin{aligned}HA|n\rangle &= AH|n\rangle - \hbar\omega A|n\rangle, \qquad \text{using (19.42),} \\ &= AE_n|n\rangle - \hbar\omega A|n\rangle \\ &= (E_n - \hbar\omega)A|n\rangle.\end{aligned}$$

Thus $A|n\rangle$ is an eigenstate of H corresponding to energy $E_n - \hbar\omega$ and must be some multiple of the normalised ket vector $|E_n - \hbar\omega\rangle$, i.e.

$$A|\,E_n\rangle \equiv A|n\rangle = c_n|E_n - \hbar\omega\rangle,$$

where c_n is not necessarily of unit modulus. Clearly, A is an operator that generates a new state that is lower in energy by $\hbar\omega$; it can thus be compared to the operator D, which has a similar effect in the context of the z-component of angular momentum. Because it possesses the property of reducing the energy of the state by $\hbar\omega$, which, as we will see, is one quantum of excitation energy for the oscillator, the operator A is called an *annihilation operator*. Repeated application of A, m times say, will produce a state whose energy is $m\hbar\omega$ lower than that of the original:

$$A^m|E_n\rangle = c_nc_{n-1}\cdots c_{n-m+1}|E_n - m\hbar\omega\rangle. \tag{19.44}$$

In a similar way it can be shown that $A^\dagger$ parallels the operator U of our angular momentum discussion and creates an additional quantum of energy each time it is applied:

$$(A^\dagger)^m|E_n\rangle = d_n d_{n+1}\cdots d_{n+m-1}|E_n + m\hbar\omega\rangle. \tag{19.45}$$

It is therefore known as a *creation operator*.

As noted earlier, the expectation value of the oscillator's energy operator $\langle\psi|H|\psi\rangle$ must be non-negative, and therefore it must have a lowest value. Let this be E_0, with corresponding eigenstate $|0\rangle$. Since the energy-lowering property of A applies to any eigenstate of H, in order to avoid a contradiction we must have that $A|0\rangle = |\emptyset\rangle$. It then follows from (19.40) that

$$\begin{aligned}
H|0\rangle &= \tfrac{1}{2}\omega(A^\dagger A + AA^\dagger)|0\rangle \\
&= \tfrac{1}{2}\omega A^\dagger A|0\rangle + \tfrac{1}{2}\omega(A^\dagger A + \hbar)|0\rangle, \qquad \text{using (19.41)}, \\
&= 0 + 0 + \tfrac{1}{2}\hbar\omega|0\rangle.
\end{aligned} \tag{19.46}$$

This shows that the commutator structure of the operators and the form of the Hamiltonian imply that the lowest energy (its ground-state energy) is $\frac{1}{2}\hbar\omega$; this is a result that has been derived without explicit reference to the corresponding wavefunction. This non-zero lowest value for the energy, known as the zero-point energy of the oscillator, and the discrete values for the allowed energy states are quantum-mechanical in origin; classically such an oscillator could have any non-negative energy, including zero.

Working back from this result, we see that the energy levels of the s.h.o. are $\frac{1}{2}\hbar\omega,\ \frac{3}{2}\hbar\omega,\ \frac{5}{2}\hbar\omega, \ldots, (m+\frac{1}{2})\hbar\omega, \ldots$, and that the corresponding (unnormalised) ket vectors can be written as

$$|0\rangle, \qquad A^\dagger|0\rangle, \qquad (A^\dagger)^2|0\rangle, \qquad \ldots, \qquad (A^\dagger)^m|0\rangle, \qquad \ldots.$$

This notation, and elaborations of it, are often used in the quantum treatment of classical fields such as the electromagnetic field. Thus, as the reader should verify, $A(A^\dagger)^3A^2A^\dagger A(A^\dagger)^4|0\rangle$ is a state with energy $\frac{9}{2}\hbar\omega$, whilst $A(A^\dagger)^3A^5A^\dagger A(A^\dagger)^4|0\rangle$ is not a physical state at all.

The normalisation of the eigenstates

In order to make quantitative calculations using the previous results we need to establish the values of the c_n and d_n that appear in equations (19.44) and (19.45). To do this, we first establish the operator recurrence relation

$$A^m(A^\dagger)^m = A^{m-1}(A^\dagger)^m A + m\hbar A^{m-1}(A^\dagger)^{m-1}. \tag{19.47}$$

The proof, which makes repeated use of $\left[A, A^\dagger\right] = \hbar$, is as follows:

$$\begin{aligned}
A^m(A^\dagger)^m &= A^{m-1}AA^\dagger(A^\dagger)^{m-1} \\
&= A^{m-1}(A^\dagger A + \hbar)(A^\dagger)^{m-1} \\
&= A^{m-1}A^\dagger A(A^\dagger)^{m-1} + \hbar A^{m-1}(A^\dagger)^{m-1} \\
&= A^{m-1}A^\dagger(A^\dagger A + \hbar)(A^\dagger)^{m-2} + \hbar A^{m-1}(A^\dagger)^{m-1} \\
&= A^{m-1}(A^\dagger)^2 A(A^\dagger)^{m-2} + A^{m-1}A^\dagger \hbar(A^\dagger)^{m-2} + \hbar A^{m-1}(A^\dagger)^{m-1} \\
&= A^{m-1}(A^\dagger)^2(A^\dagger A + \hbar)(A^\dagger)^{m-3} + 2\hbar A^{m-1}(A^\dagger)^{m-1} \\
&\;\;\vdots \\
&= A^{m-1}(A^\dagger)^m A + m\hbar A^{m-1}(A^\dagger)^{m-1}.
\end{aligned}$$

Now we take the expectation values in the ground state $|0\rangle$ of both sides of this operator equation and note that the first term on the RHS is zero since it contains the term $A|0\rangle$. The non-vanishing terms are

$$\langle 0 \,|\, A^m(A^\dagger)^m \,|\, 0\rangle = m\hbar\langle 0 \,|\, A^{m-1}(A^\dagger)^{m-1} \,|\, 0\rangle.$$

The LHS is the square of the norm of $(A^\dagger)^m|0\rangle$, and, from equation (19.45), it is equal to

$$|d_0|^2|d_1|^2 \cdots |d_{m-1}|^2 \,\langle 0 \,|\, 0\rangle.$$

Similarly, the RHS is equal to

$$m\hbar\,|d_0|^2|d_1|^2 \cdots |d_{m-2}|^2 \,\langle 0 \,|\, 0\rangle.$$

It follows that $|d_{m-1}|^2 = m\hbar$ and, taking all coefficients as real, $d_m = \sqrt{(m+1)\hbar}$. Thus the correctly normalised state of energy $(n + \frac{1}{2})\hbar$, obtained by repeated application of $A^\dagger$ to the ground state, is given by

$$|n\rangle = \frac{(A^\dagger)^n}{(n!\,\hbar^n)^{1/2}}\,|0\rangle. \tag{19.48}$$

To evaluate the c_n, we note that, from the commutator of A and $A^\dagger$,

$$\begin{aligned}
\left[A, A^\dagger\right]\,|n\rangle &= AA^\dagger|n\rangle - A^\dagger A|n\rangle \\
\hbar\,|n\rangle &= \sqrt{(n+1)\hbar}\,A\,|n+1\rangle - c_n\,A^\dagger\,|n-1\rangle \\
&= \sqrt{(n+1)\hbar}\,c_{n+1}\,|n\rangle - c_n\sqrt{n\hbar}\,|n\rangle, \\
\hbar &= \sqrt{(n+1)\hbar}\,c_{n+1} - c_n\sqrt{n\hbar},
\end{aligned}$$

which has the obvious solution $c_n = \sqrt{n\hbar}$. To summarise:

$$c_n = \sqrt{n\hbar} \qquad \text{and} \qquad d_n = \sqrt{(n+1)\hbar}. \tag{19.49}$$

We end this chapter with another worked example. This one illustrates how the operator formalism that we have developed can be used to obtain results that would involve a number of non-trivial integrals if tackled using explicit wavefunctions.

►*Given that the first-order change in the ground-state energy of a quantum system when it is perturbed by a small additional term H' in the Hamiltonian is $\langle 0|H'|0\rangle$, find the first-order change in the energy of a simple harmonic oscillator in the presence of an additional potential $V'(x) = \lambda x^3 + \mu x^4$.*

From the definitions of A and $A^\dagger$, equation (19.39), we can write

$$x = \frac{1}{\sqrt{2m\omega}}(A + A^\dagger) \quad \Rightarrow \quad H' = \frac{\lambda}{(2m\omega)^{3/2}}(A + A^\dagger)^3 + \frac{\mu}{(2m\omega)^2}(A + A^\dagger)^4.$$

We now compute successive values of $(A + A^\dagger)^n |0\rangle$ for $n = 1, 2, 3, 4$, remembering that

$$A|n\rangle = \sqrt{n\hbar}\,|n-1\rangle \qquad \text{and} \qquad A^\dagger|n\rangle = \sqrt{(n+1)\hbar}\,|n+1\rangle :$$

$$\begin{aligned}
(A + A^\dagger)|0\rangle &= 0 + \hbar^{1/2}|1\rangle, \\
(A + A^\dagger)^2|0\rangle &= \hbar|0\rangle + \sqrt{2}\,\hbar|2\rangle, \\
(A + A^\dagger)^3|0\rangle &= 0 + \hbar^{3/2}|1\rangle + 2\hbar^{3/2}|1\rangle + \sqrt{6}\,\hbar^{3/2}|3\rangle \\
&= 3\hbar^{3/2}|1\rangle + \sqrt{6}\,\hbar^{3/2}|3\rangle, \\
(A + A^\dagger)^4|0\rangle &= 3\hbar^2|0\rangle + \sqrt{18}\,\hbar^2|2\rangle + \sqrt{18}\,\hbar^2|2\rangle + \sqrt{24}\,\hbar^2|4\rangle.
\end{aligned}$$

To find the energy shift we need to form the inner product of each of these state vectors with $|0\rangle$. But $|0\rangle$ is orthogonal to all $|n\rangle$ if $n \neq 0$. Consequently, the term $\langle 0|(A + A^\dagger)^3|0\rangle$ in the expectation value is zero, and in the expression for $\langle 0|(A + A^\dagger)^4|0\rangle$ only the first term is non-zero; its value is $3\hbar^2$. The perturbation energy is thus given by

$$\langle 0|H'|0\rangle = \frac{3\mu\hbar^2}{(2m\omega)^2}.$$

It could have been anticipated on symmetry grounds that the expectation of λx^3, an odd function of x, would be zero, but the calculation gives this result automatically. The contribution of the quadratic term in the perturbation would have been much harder to anticipate! ◄

19.3 Exercises

19.1 Show that the commutator of two operators that correspond to two physical observables cannot itself correspond to another physical observable.

19.2 By expressing the operator L_z, corresponding to the z-component of angular momentum, in spherical polar coordinates (r, θ, ϕ), show that the angular momentum of a particle about the polar axis cannot be known at the same time as its azimuthal position around that axis.

19.3 In quantum mechanics, the time dependence of the state function $|\psi\rangle$ of a system is given, as a further postulate, by the equation

$$i\hbar\frac{\partial}{\partial t}|\psi\rangle = H|\psi\rangle,$$

where H is the Hamiltonian of the system. Use this to find the time dependence of the expectation value $\langle A\rangle$ of an operator A that itself has no explicit time dependence. Hence show that operators that commute with the Hamiltonian correspond to the classical 'constants of the motion'.

For a particle of mass m moving in a one-dimensional potential $V(x)$, prove Ehrenfest's theorem:

$$\frac{d\langle p_x\rangle}{dt} = -\left\langle \frac{dV}{dx}\right\rangle \quad \text{and} \quad \frac{d\langle x\rangle}{dt} = \frac{\langle p_x\rangle}{m}.$$

19.4 Show that the Pauli matrices

$$\mathsf{S}_x = \tfrac{1}{2}\hbar\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathsf{S}_y = \tfrac{1}{2}\hbar\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \mathsf{S}_z = \tfrac{1}{2}\hbar\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

which are used as the operators corresponding to intrinsic spin of $\frac{1}{2}\hbar$ in non-relativistic quantum mechanics, satisfy $\mathsf{S}_x^2 = \mathsf{S}_y^2 = \mathsf{S}_z^2 = \frac{1}{4}\hbar^2\mathsf{I}$, and have the same commutation properties as the components of orbital angular momentum. Deduce that any state $|\psi\rangle$ represented by the column vector $(a,\ b)^{\mathrm{T}}$ is an eigenstate of S^2 with eigenvalue $3\hbar^2/4$.

19.5 Find closed-form expressions for $\cos\mathsf{C}$ and $\sin\mathsf{C}$, where C is the matrix

$$\mathsf{C} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Demonstrate that the 'expected' relationships

$$\cos^2\mathsf{C} + \sin^2\mathsf{C} = \mathsf{I} \quad \text{and} \quad \sin 2\mathsf{C} = 2\sin\mathsf{C}\cos\mathsf{C}$$

are valid.

19.6 Operators A and B anticommute. Evaluate $(A+B)^{2n}$ for a few values of n and hence propose an expression for c_{nr} in the expansion

$$(A+B)^{2n} = \sum_{r=0}^{n} c_{nr}\, A^{2n-2r}\, B^{2r}.$$

Prove your proposed formula for general values of n, using the method of induction.

Show that

$$\cos(A+B) = \sum_{n=0}^{\infty}\sum_{r=0}^{n} d_{nr}\, A^{2n-2r} B^{2r},$$

where the d_{nr} are constants whose values you should determine.

By taking as A the matrix $\mathsf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, confirm that your answer is consistent with that obtained in exercise 19.5.

19.7 Expressed in terms of the annihilation and creation operators A and $A^\dagger$ discussed in the text, a system has an unperturbed Hamiltonian $H_0 = \hbar\omega A^\dagger A$. The system is disturbed by the addition of a perturbing Hamiltonian $H_1 = g\hbar\omega(A + A^\dagger)$, where g is real. Show that the effect of the perturbation is to move the whole energy spectrum of the system down by $g^2\hbar\omega$.

19.8 For a system of N electrons in their ground state $|0\rangle$, the Hamiltonian is

$$H = \sum_{n=1}^{N} \frac{p_{x_n}^2 + p_{y_n}^2 + p_{z_n}^2}{2m} + \sum_{n=1}^{N} V(x_n, y_n, z_n).$$

Show that $\left[p_{x_n}^2, x_n\right] = -2i\hbar p_{x_n}$, and hence that the expectation value of the double commutator $[\,[x, H\,], x]$, where $x = \sum_{n=1}^{N} x_n$, is given by

$$\langle 0|\,[\,[x, H\,], x]\,|0\rangle = \frac{N\hbar^2}{m}.$$

Now evaluate the expectation value using the eigenvalue properties of H, namely $H|r\rangle = E_r|r\rangle$, and deduce the *sum rule for oscillation strengths*,

$$\sum_{r=0}^{\infty}(E_r - E_0)|\langle r\,|\,x\,|\,0\rangle|^2 = \frac{N\hbar^2}{2m}.$$

19.9 By considering the function

$$F(\lambda) = \exp(\lambda A)B\exp(-\lambda A),$$

where A and B are linear operators and λ is a parameter, and finding its derivatives with respect to λ, prove that

$$e^A B e^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \frac{1}{3!}[A, [A, [A, B]]] + \cdots.$$

Use this result to express

$$\exp\left(\frac{iL_x\theta}{\hbar}\right) L_y \exp\left(\frac{-iL_x\theta}{\hbar}\right)$$

as a linear combination of the angular momentum operators L_x, L_y and L_z.

19.10 For a system containing more than one particle, the total angular momentum J and its components are represented by operators that have completely analogous commutation relations to those for the operators for a single particle, i.e. J^2 has eigenvalue $j(j+1)\hbar^2$ and J_z has eigenvalue $m_j\hbar$ for the state $|j, m_j\rangle$. The usual orthonormality relationship $\langle j', m'_j\,|\,j, m_j\rangle = \delta_{j'j}\,\delta_{m'_j m_j}$ is also valid.

A system consists of two (distinguishable) particles A and B. Particle A is in an $\ell = 3$ state and can have state functions of the form $|A, 3, m_A\rangle$, whilst B is in an $\ell = 2$ state with possible state functions $|B, 2, m_B\rangle$. The range of possible values for j is $|3-2| \le j \le |3+2|$, i.e. $1 \le j \le 5$, and the overall state function can be written as

$$|j, m_j\rangle = \sum_{m_A+m_B=m_j} C^{j\,m_j}_{m_A\,m_B}\,|\,A, 3, m_A\rangle\,|\,B, 2, m_B\rangle.$$

The numerical coefficients $C^{j\,m_j}_{m_A\,m_B}$ are known as *Clebsch–Gordon* coefficients.

Assume (as can be shown) that the ladder operators $U(AB)$ and $D(AB)$ for the system can be written as $U(A) + U(B)$ and $D(A) + D(B)$, respectively, and that they lead to relationships equivalent to (19.34) and (19.35) with ℓ replaced by j and m by m_j.

(a) Apply the operators to the (obvious) relationship

$$|AB, 5, 5\rangle = |A, 3, 3\rangle\,|B, 2, 2\rangle$$

to show that

$$|AB, 5, 4\rangle = \sqrt{\tfrac{6}{10}}\,|A, 3, 2\rangle\,|B, 2, 2\rangle + \sqrt{\tfrac{4}{10}}\,|A, 3, 3\rangle\,|B, 2, 1\rangle.$$

(b) Find, to within an overall sign, the real coefficients c and d in the expansion

$$|AB, 4, 4\rangle = c|A, 3, 2\rangle\,|B, 2, 2\rangle + d|A, 3, 3\rangle\,|B, 2, 1\rangle$$

by requiring it to be orthogonal to $|AB, 5, 4\rangle$. Check your answer by considering $U(AB)|AB, 4, 4\rangle$.

(c) Find, to within an overall sign, and as efficiently as possible, an expression for $|AB, 4, -3\rangle$ as a sum of products of the form $|A, 3, m_A\rangle\,|B, 2, m_B\rangle$.

19.4 Hints and answers

19.1 Show that the commutator is anti-Hermitian.

19.3 Use the Hermitian conjugate of the given equation to obtain the time dependence of $\langle\psi|$. The rate of change of $\langle\psi|A|\psi\rangle$ is $i\langle\psi|\,[H, A]\,|\psi\rangle$. Note that $[H, p_x] = [V, p_x]$ and $[H, x] = \left[p_x^2, x\right]/2m$.

19.5 Show that $\mathsf{C}^2 = 2\mathsf{I}$.

$$\cos\mathsf{C} = \cos\sqrt{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sin\mathsf{C} = \frac{\sin\sqrt{2}}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

19.7 Express the total Hamiltonian in terms of $B = A + gI$ and determine the value of $\left[B, B^\dagger\right]$.

19.9 Show that, if $F^{(n)}$ is the nth derivative of $F(\lambda)$, then $F^{(n+1)} = \left[A, F^{(n)}\right]$. Use a Taylor series in λ to evaluate $F(1)$, using derivatives evaluated at $\lambda = 0$. Successively reduce the level of nesting of each multiple commutator by using the result of evaluating the previous term. The given expression reduces to $\cos\theta\, L_y - \sin\theta\, L_z$.

20

Partial differential equations: general and particular solutions

In this chapter and the next the solution of differential equations of types typically encountered in the physical sciences and engineering is extended to situations involving more than one independent variable. A partial differential equation (PDE) is an equation relating an unknown function (the dependent variable) of two or more variables to its partial derivatives with respect to those variables. The most commonly occurring independent variables are those describing position and time, and so we will couch our discussion and examples in notation appropriate to them.

As in other chapters we will focus our attention on the equations that arise most often in physical situations. We will restrict our discussion, therefore, to linear PDEs, i.e. those of first degree in the dependent variable. Furthermore, we will discuss primarily second-order equations. The solution of first-order PDEs will necessarily be involved in treating these, and some of the methods discussed can be extended without difficulty to third- and higher-order equations. We shall also see that many ideas developed for ordinary differential equations (ODEs) can be carried over directly into the study of PDEs.

In this chapter we will concentrate on general solutions of PDEs in terms of arbitrary functions and the particular solutions that may be derived from them in the presence of boundary conditions. We also discuss the existence and uniqueness of the solutions to PDEs under given boundary conditions.

In the next chapter the methods most commonly used in practice for obtaining solutions to PDEs subject to given boundary conditions will be considered. These methods include the separation of variables, integral transforms and Green's functions. This division of material is rather arbitrary and has been made only to emphasise the general usefulness of the latter methods. In particular, it will be readily apparent that some of the results of the present chapter are in fact solutions in the form of separated variables, but arrived at by a different approach.

20.1 Important partial differential equations

Most of the important PDEs of physics are second-order and linear. In order to gain familiarity with their general form, some of the more important ones will now be briefly discussed. These equations apply to a wide variety of different physical systems.

Since, in general, the PDEs listed below describe three-dimensional situations, the independent variables are $\mathbf{r}$ and t, where $\mathbf{r}$ is the position vector and t is time. The actual variables used to specify the position vector $\mathbf{r}$ are dictated by the coordinate system in use. For example, in Cartesian coordinates the independent variables of position are x, y and z, whereas in spherical polar coordinates they are r, θ and ϕ. The equations may be written in a coordinate-independent manner, however, by the use of the Laplacian operator ∇^2.

20.1.1 The wave equation

The wave equation

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} \tag{20.1}$$

describes as a function of position and time the displacement from equilibrium, $u(\mathbf{r}, t)$, of a vibrating string or membrane or a vibrating solid, gas or liquid. The equation also occurs in electromagnetism, where u may be a component of the electric or magnetic field in an elecromagnetic wave or the current or voltage along a transmission line. The quantity c is the speed of propagation of the waves.

► *Find the equation satisfied by small transverse displacements $u(x,t)$ of a uniform string of mass per unit length ρ held under a uniform tension T, assuming that the string is initially located along the x-axis in a Cartesian coordinate system.*

Figure 20.1 shows the forces acting on an elemental length Δs of the string. If the tension T in the string is uniform along its length then the net upward vertical force on the element is

$$\Delta F = T\sin\theta_2 - T\sin\theta_1.$$

Assuming that the angles θ_1 and θ_2 are both small, we may make the approximation $\sin\theta \approx \tan\theta$. Since at any point on the string the slope $\tan\theta = \partial u/\partial x$, the force can be written

$$\Delta F = T\left[\frac{\partial u(x+\Delta x,t)}{\partial x} - \frac{\partial u(x,t)}{\partial x}\right] \approx T\frac{\partial^2 u(x,t)}{\partial x^2}\,\Delta x,$$

where we have used the definition of the partial derivative to simplify the RHS.

This upward force may be equated, by Newton's second law, to the product of the mass of the element and its upward acceleration. The element has a mass $\rho\,\Delta s$, which is approximately equal to $\rho\,\Delta x$ if the vibrations of the string are small, and so we have

$$\rho\,\Delta x\,\frac{\partial^2 u(x,t)}{\partial t^2} = T\frac{\partial^2 u(x,t)}{\partial x^2}\,\Delta x.$$

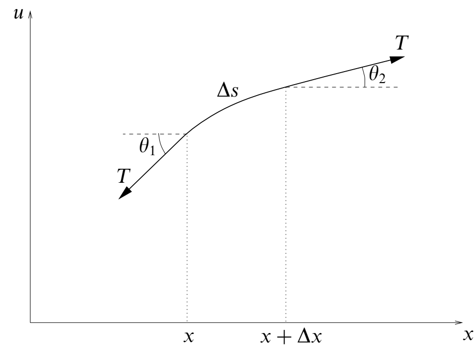


Figure 20.1 The forces acting on an element of a string under uniform tension T.

Dividing both sides by Δx we obtain, for the vibrations of the string, the one-dimensional wave equation

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2},$$

where $c^2 = T/\rho$. ◀

The longitudinal vibrations of an elastic rod obey a very similar equation to that derived in the above example, namely

$$\frac{\partial^2 u}{\partial x^2} = \frac{\rho}{E}\frac{\partial^2 u}{\partial t^2};$$

here ρ is the mass per unit volume and E is Young's modulus.

The wave equation can be generalised slightly. For example, in the case of the vibrating string, there could also be an external upward vertical force $f(x,t)$ per unit length acting on the string at time t. The transverse vibrations would then satisfy the equation

$$T\frac{\partial^2 u}{\partial x^2} + f(x,t) = \rho\frac{\partial^2 u}{\partial t^2},$$

which is clearly of the form 'upward force per unit length = mass per unit length × upward acceleration'.

Similar examples, but involving two or three spatial dimensions rather than one, are provided by the equation governing the transverse vibrations of a stretched membrane subject to an external vertical force density $f(x,y,t)$,

$$T\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + f(x,y,t) = \rho(x,y)\frac{\partial^2 u}{\partial t^2},$$

where ρ is the mass per unit area of the membrane and T is the tension.

20.1.2 The diffusion equation

The diffusion equation

$$\kappa\nabla^2 u = \frac{\partial u}{\partial t} \tag{20.2}$$

describes the temperature u in a region containing no heat sources or sinks; it also applies to the diffusion of a chemical that has a concentration $u(\mathbf{r}, t)$. The constant κ is called the diffusivity. The equation is clearly second order in the three spatial variables, but first order in time.

►*Derive the equation satisfied by the temperature $u(\mathbf{r}, t)$ at time t for a material of uniform thermal conductivity k, specific heat capacity s and density ρ. Express the equation in Cartesian coordinates.*

Let us consider an arbitrary volume V lying within the solid and bounded by a surface S (this may coincide with the surface of the solid if so desired). At any point in the solid the rate of heat flow per unit area in any given direction $\hat{\mathbf{r}}$ is proportional to minus the component of the temperature gradient in that direction and so is given by $(-k\nabla u)\cdot\hat{\mathbf{r}}$. The total flux of heat *out* of the volume V per unit time is given by

$$\begin{aligned} -\frac{dQ}{dt} &= \iint_S (-k\nabla u)\cdot\hat{\mathbf{n}}\,dS \\ &= \iiint_V \nabla\cdot(-k\nabla u)\,dV, \end{aligned} \tag{20.3}$$

where Q is the total heat energy in V at time t and $\hat{\mathbf{n}}$ is the outward-pointing unit normal to S; note that we have used the divergence theorem to convert the surface integral into a volume integral.

We can also express Q as a volume integral over V,

$$Q = \iiint_V s\rho u\,dV,$$

and its rate of change is then given by

$$\frac{dQ}{dt} = \iiint_V s\rho\frac{\partial u}{\partial t}\,dV, \tag{20.4}$$

where we have taken the derivative with respect to time inside the integral (see section 5.12).

Comparing (20.3) and (20.4), and remembering that the volume V is arbitrary, we obtain the three-dimensional diffusion equation

$$\kappa\nabla^2 u = \frac{\partial u}{\partial t},$$

where the diffusion coefficient $\kappa = k/(s\rho)$. To express this equation in Cartesian coordinates, we simply write ∇^2 in terms of x, y and z to obtain

$$\kappa\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) = \frac{\partial u}{\partial t}. \blacktriangleleft$$

The diffusion equation just derived can be generalised to

$$k\nabla^2 u + f(\mathbf{r}, t) = s\rho\frac{\partial u}{\partial t}.$$

The second term, $f(\mathbf{r}, t)$, represents a varying density of heat sources throughout the material but is often not required in physical applications. In the most general case, k, s and ρ may depend on position $\mathbf{r}$, in which case the first term becomes $\nabla \cdot (k\nabla u)$. However, in the simplest application the heat flow is one-dimensional with no heat sources, and the equation becomes (in Cartesian coordinates)

$$\frac{\partial^2 u}{\partial x^2} = \frac{s\rho}{k}\frac{\partial u}{\partial t}.$$

20.1.3 Laplace's equation

Laplace's equation,

$$\nabla^2 u = 0, \tag{20.5}$$

may be obtained by setting $\partial u/\partial t = 0$ in the diffusion equation (20.2), and describes (for example) the *steady-state* temperature distribution in a solid in which there are no heat sources – i.e. the temperature distribution after a long time has elapsed.

Laplace's equation also describes the gravitational potential in a region containing no matter or the electrostatic potential in a charge-free region. Further, it applies to the flow of an incompressible fluid with no sources, sinks or vortices; in this case u is the velocity potential, from which the velocity is given by $v = \nabla u$.

20.1.4 Poisson's equation

Poisson's equation,

$$\nabla^2 u = \rho(\mathbf{r}), \tag{20.6}$$

describes the same physical situations as Laplace's equation, but in regions containing matter, charges or sources of heat or fluid. The function $\rho(\mathbf{r})$ is called the source density and in physical applications usually contains some multiplicative physical constants. For example, if u is the electrostatic potential in some region of space, in which case ρ is the density of electric charge, then $\nabla^2 u = -\rho(\mathbf{r})/\epsilon_0$, where ϵ_0 is the permittivity of free space. Alternatively, u might represent the gravitational potential in some region where the matter density is given by ρ; then $\nabla^2 u = 4\pi G\rho(\mathbf{r})$, where G is the gravitational constant.

20.1.5 Schrødinger's equation

The Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2 u + V(\mathbf{r})u = i\hbar\frac{\partial u}{\partial t}, \tag{20.7}$$

describes the quantum mechanical wavefunction $u(\mathbf{r}, t)$ of a non-relativistic particle of mass m; $\hbar$ is Planck's constant divided by 2π. Like the diffusion equation it is second order in the three spatial variables and first order in time.

20.2 General form of solution

Before turning to the methods by which we may hope to solve PDEs such as those listed in the previous section, it is instructive, as for ODEs in chapter 14, to study how PDEs may be formed from a set of possible solutions. Such a study can provide an indication of how equations obtained not from possible solutions but from physical arguments might be solved.

For definiteness let us suppose we have a set of functions involving two independent variables x and y. Without further specification this is of course a very wide set of functions, and we could not expect to find a useful equation that they all satisfy. However, let us consider a type of function $u_i(x, y)$ in which x and y appear in a particular way, such that u_i can be written as a function (however complicated) *of a single variable* p, itself a simple function of x and y.

Let us illustrate this by considering the three functions

$$\begin{aligned}
u_1(x, y) &= x^4 + 4(x^2 y + y^2 + 1), \\
u_2(x, y) &= \sin x^2 \cos 2y + \cos x^2 \sin 2y, \\
u_3(x, y) &= \frac{x^2 + 2y + 2}{3x^2 + 6y + 5}.
\end{aligned}$$

These are all fairly complicated functions of x and y and a single differential equation of which each one is a solution is not obvious. However, if we observe that in fact each can be expressed as a function of the variable $p = x^2 + 2y$ alone (with no other x or y involved) then a great simplification takes place. Written in terms of p the above equations become

$$\begin{aligned}
u_1(x, y) &= (x^2 + 2y)^2 + 4 = p^2 + 4 = f_1(p), \\
u_2(x, y) &= \sin(x^2 + 2y) = \sin p = f_2(p), \\
u_3(x, y) &= \frac{(x^2 + 2y) + 2}{3(x^2 + 2y) + 5} = \frac{p + 2}{3p + 5} = f_3(p).
\end{aligned}$$

Let us now form, for each u_i, the partial derivatives $\partial u_i/\partial x$ and $\partial u_i/\partial y$. In each case these are (writing both the form for general p and the one appropriate to our particular case, $p = x^2 + 2y$)

$$\begin{aligned}
\frac{\partial u_i}{\partial x} &= \frac{df_i(p)}{dp}\frac{\partial p}{\partial x} = 2xf_i', \\
\frac{\partial u_i}{\partial y} &= \frac{df_i(p)}{dp}\frac{\partial p}{\partial y} = 2f_i',
\end{aligned}$$

for $i = 1, 2, 3$. All reference to the form of f_i can be eliminated from these equations by cross-multiplication, obtaining

$$\frac{\partial p}{\partial y}\frac{\partial u_i}{\partial x} = \frac{\partial p}{\partial x}\frac{\partial u_i}{\partial y},$$

or, for our specific form, $p = x^2 + 2y$,

$$\frac{\partial u_i}{\partial x} = x\frac{\partial u_i}{\partial y}. \tag{20.8}$$

It is thus apparent that not only are the three functions u_1, u_2 u_3 solutions of the PDE (20.8) but so also is *any arbitrary function* $f(p)$ of which the argument p has the form $x^2 + 2y$.

20.3 General and particular solutions

In the last section we found that the first-order PDE (20.8) has as a solution *any* function of the variable $x^2 + 2y$. This points the way for the solution of PDEs of other orders, as follows. It is *not* generally true that an nth-order PDE can always be considered as resulting from the elimination of n arbitrary *functions* from its solution (as opposed to the elimination of n arbitrary *constants* for an nth-order ODE, see section 14.1). However, given specific PDEs we can try to solve them by seeking combinations of variables in terms of which the solutions may be expressed as arbitrary functions. Where this is possible we may expect n combinations to be involved in the solution.

Naturally, the exact functional form of the solution for any particular situation must be determined by some set of boundary conditions. For instance, if the PDE contains two independent variables x and y then for complete determination of its solution the boundary conditions will take a form equivalent to specifying $u(x, y)$ along a suitable continuum of points in the xy-plane (usually along a line).

We now discuss the general and particular solutions of first- and second-order PDEs. In order to simplify the algebra, we will restrict our discussion to equations containing just two independent variables x and y. Nevertheless, the method presented below may be extended to equations containing several independent variables.

20.3.1 First-order equations

Although most of the PDEs encountered in physical contexts are second order (i.e. they contain $\partial^2 u/\partial x^2$ or $\partial^2 u/\partial x\partial y$, etc.), we now discuss first-order equations to illustrate the general considerations involved in the form of the solution and in satisfying any boundary conditions on the solution.

The most general first-order linear PDE (containing two independent variables) is of the form

$$A(x,y)\frac{\partial u}{\partial x} + B(x,y)\frac{\partial u}{\partial y} + C(x,y)u = R(x,y), \tag{20.9}$$

where $A(x,y)$, $B(x,y)$, $C(x,y)$ and $R(x,y)$ are given functions. Clearly, if either $A(x,y)$ or $B(x,y)$ is zero then the PDE may be solved straightforwardly as a first-order linear ODE (as discussed in chapter 14), the only modification being that the arbitrary constant of integration becomes an *arbitrary function* of x or y respectively.

►*Find the general solution $u(x,y)$ of*

$$x\frac{\partial u}{\partial x} + 3u = x^2.$$

Dividing through by x we obtain

$$\frac{\partial u}{\partial x} + \frac{3u}{x} = x,$$

which is a linear equation with integrating factor (see subsection 14.2.4)

$$\exp\left(\int \frac{3}{x}\,dx\right) = \exp(3\ln x) = x^3.$$

Multiplying through by this factor we find

$$\frac{\partial}{\partial x}(x^3 u) = x^4,$$

which, on integrating with respect to x, gives

$$x^3 u = \frac{x^5}{5} + f(y),$$

where $f(y)$ is an *arbitrary function* of y. Finally, dividing through by x^3, we obtain the solution

$$u(x,y) = \frac{x^2}{5} + \frac{f(y)}{x^3}. \blacktriangleleft$$

When the PDE contains partial derivatives with respect to both independent variables then, of course, we cannot employ the above procedure but must seek an alternative method. Let us for the moment restrict our attention to the special case in which $C(x,y) = R(x,y) = 0$ and, following the discussion of the previous section, look for solutions of the form $u(x,y) = f(p)$ where p is some, at present unknown, combination of x and y. We then have

$$\frac{\partial u}{\partial x} = \frac{df(p)}{dp}\frac{\partial p}{\partial x},$$
$$\frac{\partial u}{\partial y} = \frac{df(p)}{dp}\frac{\partial p}{\partial y},$$

which, when substituted into the PDE (20.9), give

$$\left[A(x,y)\frac{\partial p}{\partial x} + B(x,y)\frac{\partial p}{\partial y}\right]\frac{df(p)}{dp} = 0.$$

This removes all reference to the actual form of the function $f(p)$ since for non-trivial p we must have

$$A(x,y)\frac{\partial p}{\partial x} + B(x,y)\frac{\partial p}{\partial y} = 0. \qquad (20.10)$$

Let us now consider the necessary condition for $f(p)$ to remain constant as x and y vary; this is that p itself remains constant. Thus for f to remain constant implies that x and y must vary in such a way that

$$dp = \frac{\partial p}{\partial x}\,dx + \frac{\partial p}{\partial y}\,dy = 0. \qquad (20.11)$$

The forms of (20.10) and (20.11) are very alike and become the same if we require that

$$\frac{dx}{A(x,y)} = \frac{dy}{B(x,y)}. \qquad (20.12)$$

By integrating this expression the form of p can be found.

►*For*

$$x\frac{\partial u}{\partial x} - 2y\frac{\partial u}{\partial y} = 0, \qquad (20.13)$$

find (i) *the solution that takes the value* $2y+1$ *on the line* $x=1$, *and* (ii) *a solution that has the value* 4 *at the point* $(1,1)$.

If we seek a solution of the form $u(x,y) = f(p)$, we deduce from (20.12) that $u(x,y)$ will be constant along lines of (x,y) that satisfy

$$\frac{dx}{x} = \frac{dy}{-2y},$$

which on integrating gives $x = cy^{-1/2}$. Identifying the constant of integration c with $p^{1/2}$ (to avoid fractional powers), we conclude that $p = x^2y$. Thus the general solution of the PDE (20.13) is

$$u(x,y) = f(x^2y),$$

where f is an arbitrary function.

We must now find the particular solutions that obey each of the imposed boundary conditions. For boundary condition (i) a little thought shows that the particular solution required is

$$u(x,y) = 2(x^2y) + 1 = 2x^2y + 1. \qquad (20.14)$$

For boundary condition (ii) some obviously acceptable solutions are

$$u(x,y) = x^2y + 3,$$
$$u(x,y) = 4x^2y,$$
$$u(x,y) = 4.$$

Each is a valid solution (the freedom of choice of form arises from the fact that u is specified at only one point $(1,1)$, and not along a continuum (say), as in boundary condition (i)). All three are particular examples of the general solution, which may be written, for example, as

$$u(x,y) = x^2y + 3 + g(x^2y),$$

where $g = g(x^2y) = g(p)$ is an arbitrary function subject only to $g(1) = 0$. For this example, the forms of g corresponding to the particular solutions listed above are $g(p) = 0$, $g(p) = 3p - 3$, $g(p) = 1 - p$. ◄

As mentioned above, in order to find a solution of the form $u(x,y) = f(p)$ we require that the original PDE contains no term in u, but only terms containing its partial derivatives. If a term in u is present, so that $C(x,y) \neq 0$ in (20.9), then the procedure needs some modification, since we cannot simply divide out the dependence on $f(p)$ to obtain (20.10). In such cases we look instead for a solution of the form $u(x,y) = h(x,y)f(p)$. We illustrate this method in the following example.

►*Find the general solution of*

$$x\frac{\partial u}{\partial x} + 2\frac{\partial u}{\partial y} - 2u = 0. \tag{20.15}$$

We seek a solution of the form $u(x,y) = h(x,y)f(p)$, with the consequence that

$$\frac{\partial u}{\partial x} = \frac{\partial h}{\partial x}f(p) + h\frac{df(p)}{dp}\frac{\partial p}{\partial x},$$
$$\frac{\partial u}{\partial y} = \frac{\partial h}{\partial y}f(p) + h\frac{df(p)}{dp}\frac{\partial p}{\partial y}.$$

Substituting these expressions into the PDE (20.15) and rearranging, we obtain

$$\left(x\frac{\partial h}{\partial x} + 2\frac{\partial h}{\partial y} - 2h\right)f(p) + \left(x\frac{\partial p}{\partial x} + 2\frac{\partial p}{\partial y}\right)h\frac{df(p)}{dp} = 0.$$

The first factor in parentheses is just the original PDE with u replaced by h. Therefore, if h is *any* solution of the PDE, *however simple*, this term will vanish, to leave

$$\left(x\frac{\partial p}{\partial x} + 2\frac{\partial p}{\partial y}\right)h\frac{df(p)}{dp} = 0,$$

from which, as in the previous case, we obtain

$$x\frac{\partial p}{\partial x} + 2\frac{\partial p}{\partial y} = 0.$$

From (20.11) and (20.12) we see that $u(x,y)$ will be constant along lines of (x,y) that satisfy

$$\frac{dx}{x} = \frac{dy}{2},$$

which integrates to give $x = c\exp(y/2)$. Identifying the constant of integration c with p we find $p = x\exp(-y/2)$. Thus the general solution of (20.15) is

$$u(x,y) = h(x,y)f(x\exp(-\tfrac{1}{2}y)),$$

where $f(p)$ is any arbitrary function of p and $h(x,y)$ is any solution of (20.15).

If we take, for example, $h(x, y) = \exp y$, which clearly satisfies (20.15), then the general solution is

$$u(x, y) = (\exp y)f(x \exp(-\tfrac{1}{2}y)).$$

Alternatively, $h(x, y) = x^2$ also satisfies (20.15) and so the general solution to the equation can also be written

$$u(x, y) = x^2 g(x \exp(-\tfrac{1}{2}y)),$$

where g is an arbitrary function of p; clearly $g(p) = f(p)/p^2$. ◀

20.3.2 Inhomogeneous equations and problems

Let us discuss in a more general form the particular solutions of (20.13) found in the second example of the previous subsection. It is clear that, so far as this equation is concerned, if $u(x, y)$ is a solution then so is any multiple of $u(x, y)$ or any linear sum of separate solutions $u_1(x, y) + u_2(x, y)$. However, when it comes to fitting the boundary conditions this is not so.

For example, although $u(x, y)$ in (20.14) satisfies the PDE and the boundary condition $u(1, y) = 2y + 1$, the function $u_1(x, y) = 4u(x, y) = 8xy + 4$, whilst satisfying the PDE, takes the value $8y+4$ on the line $x = 1$ and so does not satisfy the required boundary condition. Likewise the function $u_2(x, y) = u(x, y)+f_1(x^2y)$, for arbitrary f_1, satisfies (20.13) but takes the value $u_2(1, y) = 2y + 1 + f_1(y)$ on the line $x = 1$, and so is not of the required form unless f_1 is identically zero.

Thus we see that when treating the superposition of solutions of PDEs two considerations arise, one concerning the equation itself and the other connected to the boundary conditions. The *equation* is said to be homogeneous if the fact that $u(x, y)$ is a solution implies that $\lambda u(x, y)$, for any constant λ, is also a solution. However, the *problem* is said to be homogeneous if, in addition, the boundary conditions are such that if they are satisfied by $u(x, y)$ then they are also satisfied by $\lambda u(x, y)$. The last requirement itself is referred to as that of *homogeneous boundary conditions*.

For example, the PDE (20.13) is homogeneous but the general first-order equation (20.9) would not be homogeneous unless $R(x, y) = 0$. Furthermore, the boundary condition (i) imposed on the solution of (20.13) in the previous subsection is not homogeneous though, in this case, the boundary condition

$$u(x, y) = 0 \qquad \text{on the line } y = 4x^{-2}$$

would be, since $u(x, y) = \lambda(x^2y - 4)$ satisfies this condition for any λ and, being a function of x^2y, satisfies (20.13).

The reason for discussing the homogeneity of PDEs and their boundary conditions is that in linear PDEs there is a close parallel to the complementary-function and particular-integral property of ODEs. The general solution of an inhomogeneous problem can be written as the sum of *any* particular solution of the problem and the general solution of the corresponding homogeneous problem (as for ODEs, we require that the particular solution is not already contained in the general solution of the homogeneous problem). Thus, for example, the general solution of

$$\frac{\partial u}{\partial x} - x\frac{\partial u}{\partial y} + au = f(x, y), \tag{20.16}$$

subject to, say, the boundary condition $u(0, y) = g(y)$, is given by

$$u(x, y) = v(x, y) + w(x, y),$$

where $v(x, y)$ is any solution (however simple) of (20.16) such that $v(0, y) = g(y)$ and $w(x, y)$ is the general solution of

$$\frac{\partial w}{\partial x} - x\frac{\partial w}{\partial y} + aw = 0, \tag{20.17}$$

with $w(0, y) = 0$. If the boundary conditions are sufficiently specified then the only possible solution of (20.17) will be $w(x, y) \equiv 0$ and $v(x, y)$ will be the complete solution by itself.

Alternatively, we may begin by finding the general solution of the inhomogeneous equation (20.16) *without* regard for any boundary conditions; it is just the sum of the general solution to the homogeneous equation and a particular integral of (20.16), both without reference to the boundary conditions. The boundary conditions can then be used to find the appropriate particular solution from the general solution.

We will not discuss at length general methods of obtaining particular integrals of PDEs but merely note that some of those methods available for ordinary differential equations can be suitably extended.[§]

►*Find the general solution of*

$$y\frac{\partial u}{\partial x} - x\frac{\partial u}{\partial y} = 3x. \tag{20.18}$$

Hence find the most general particular solution (i) *which satisfies* $u(x, 0) = x^2$ *and* (ii) *which has the value* $u(x, y) = 2$ *at the point* $(1, 0)$.

This equation is inhomogeneous, and so let us first find the general solution of (20.18) without regard for any boundary conditions. We begin by looking for the solution of the corresponding homogeneous equation ((20.18) but with the RHS equal to zero) of the form $u(x, y) = f(p)$. Following the same procedure as that used in the solution of (20.13) we find that $u(x, y)$ will be constant along lines of (x, y) that satisfy

$$\frac{dx}{y} = \frac{dy}{-x} \quad \Rightarrow \quad \frac{x^2}{2} + \frac{y^2}{2} = c.$$

Identifying the constant of integration c with $p/2$, we find that the general solution of the

[§] See for example H. T. H. Piaggio, *An Elementary Treatise on Differential Equations and their Applications* (London: G. Bell and Sons, Ltd, 1954), pp. 175 ff.

homogeneous equation is $u(x, y) = f(x^2 + y^2)$ for arbitrary function f. Now by inspection a particular integral of (20.18) is $u(x, y) = -3y$, and so the general solution to (20.18) is

$$u(x, y) = f(x^2 + y^2) - 3y.$$

Boundary condition (i) requires $u(x, 0) = f(x^2) = x^2$, i.e. $f(z) = z$, and so the particular solution in this case is

$$u(x, y) = x^2 + y^2 - 3y.$$

Similarly, boundary condition (ii) requires $u(1, 0) = f(1) = 2$. One possibility is $f(z) = 2z$, and if we make this choice, then one way of writing the most general particular solution is

$$u(x, y) = 2x^2 + 2y^2 - 3y + g(x^2 + y^2),$$

where g is any arbitrary function for which $g(1) = 0$. Alternatively, a simpler choice would be $f(z) = 2$, leading to

$$u(x, y) = 2 - 3y + g(x^2 + y^2). \blacktriangleleft$$

Although we have discussed the solution of inhomogeneous problems only for first-order equations, the general considerations hold true for linear PDEs of higher order.

20.3.3 Second-order equations

As noted in section 20.1, second-order linear PDEs are of great importance in describing the behaviour of many physical systems. As in our discussion of first-order equations, for the moment we shall restrict our discussion to equations with just two independent variables; extensions to a greater number of independent variables are straightforward.

The most general second-order linear PDE (containing two independent variables) has the form

$$A\frac{\partial^2 u}{\partial x^2} + B\frac{\partial^2 u}{\partial x \partial y} + C\frac{\partial^2 u}{\partial y^2} + D\frac{\partial u}{\partial x} + E\frac{\partial u}{\partial y} + Fu = R(x, y), \tag{20.19}$$

where $A, B, \ldots, F$ and $R(x, y)$ are given functions of x and y. Because of the nature of the solutions to such equations, they are usually divided into three classes, a division of which we will make further use in subsection 20.6.2. The equation (20.19) is called *hyperbolic* if $B^2 > 4AC$, *parabolic* if $B^2 = 4AC$ and *elliptic* if $B^2 < 4AC$. Clearly, if A, B and C are functions of x and y (rather than just constants) then the equation might be of different types in different parts of the xy-plane.

Equation (20.19) obviously represents a very large class of PDEs, and it is usually impossible to find closed-form solutions to most of these equations. Therefore, for the moment we shall consider only homogeneous equations, with $R(x, y) = 0$, and make the further (greatly simplifying) restriction that, throughout the remainder of this section, $A, B, \ldots, F$ are not functions of x and y but merely constants.

We now tackle the problem of solving some types of second-order PDE with constant coefficients by seeking solutions that are arbitrary functions of particular combinations of independent variables, just as we did for first-order equations.

Following the discussion of the previous section, we can hope to find such solutions only if all the terms of the equation involve the same total number of differentiations, i.e. all terms are of the same order, although the number of differentiations with respect to the individual independent variables may be different. This means that in (20.19) we require the constants D, E and F to be identically zero (we have, of course, already assumed that $R(x, y)$ is zero), so that we are now considering only equations of the form

$$A\frac{\partial^2 u}{\partial x^2} + B\frac{\partial^2 u}{\partial x \partial y} + C\frac{\partial^2 u}{\partial y^2} = 0, \tag{20.20}$$

where A, B and C are constants. We note that both the one-dimensional wave equation,

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0,$$

and the two-dimensional Laplace equation,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

are of this form, but that the diffusion equation,

$$\kappa\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t} = 0,$$

is not, since it contains a first-order derivative.

Since all the terms in (20.20) involve two differentiations, by assuming a solution of the form $u(x, y) = f(p)$, where p is some unknown function of x and y (or t), we may be able to obtain a common factor $d^2f(p)/dp^2$ as the only appearance of f on the LHS. Then, because of the zero RHS, all reference to the form of f can be cancelled out.

We can gain some guidance on suitable forms for the combination $p = p(x, y)$ by considering $\partial u/\partial x$ when u is given by $u(x, y) = f(p)$, for then

$$\frac{\partial u}{\partial x} = \frac{df(p)}{dp}\frac{\partial p}{\partial x}.$$

Clearly differentiation of this equation with respect to x (or y) will not lead to a single term on the RHS, containing f only as $d^2f(p)/dp^2$, unless the factor $\partial p/\partial x$ is a constant so that $\partial^2 p/\partial x^2$ and $\partial^2 p/\partial x\partial y$ are necessarily zero. This shows that p must be a linear function of x. In an exactly similar way p must also be a linear function of y, i.e. $p = ax + by$.

If we assume a solution of (20.20) of the form $u(x, y) = f(ax+by)$, and evaluate the terms ready for substitution into (20.20), we obtain

$$\frac{\partial u}{\partial x} = a\frac{df(p)}{dp}, \qquad \frac{\partial u}{\partial y} = b\frac{df(p)}{dp},$$

$$\frac{\partial^2 u}{\partial x^2} = a^2\frac{d^2f(p)}{dp^2}, \qquad \frac{\partial^2 u}{\partial x \partial y} = ab\frac{d^2f(p)}{dp^2}, \qquad \frac{\partial^2 u}{\partial y^2} = b^2\frac{d^2f(p)}{dp^2},$$

which on substitution give

$$\left(Aa^2 + Bab + Cb^2\right)\frac{d^2f(p)}{dp^2} = 0. \tag{20.21}$$

This is the form we have been seeking, since now a solution independent of the form of f can be obtained if we require that a and b satisfy

$$Aa^2 + Bab + Cb^2 = 0.$$

From this quadratic, two values for the ratio of the two constants a and b are obtained,

$$b/a = [-B \pm (B^2 - 4AC)^{1/2}]/2C.$$

If we denote these two ratios by λ_1 and λ_2 then *any* functions of the two variables

$$p_1 = x + \lambda_1 y, \qquad p_2 = x + \lambda_2 y$$

will be solutions of the original equation (20.20). The omission of the constant factor a from p_1 and p_2 is of no consequence since this can always be absorbed into the particular form of any chosen function; only the *relative* weighting of x and y in p is important.

Since p_1 and p_2 are in general different, we can thus write the general solution of (20.20) as

$$u(x, y) = f(x + \lambda_1 y) + g(x + \lambda_2 y), \tag{20.22}$$

where f and g are arbitrary functions.

Finally, we note that the alternative solution $d^2f(p)/dp^2 = 0$ to (20.21) leads only to the trivial solution $u(x, y) = kx + ly + m$, for which all second derivatives are individually zero.

► *Find the general solution of the one-dimensional wave equation*

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0.$$

This equation is (20.20) with $A = 1$, $B = 0$ and $C = -1/c^2$, and so the values of λ_1 and λ_2 are the solutions of

$$1 - \frac{\lambda^2}{c^2} = 0,$$

namely $\lambda_1 = -c$ and $\lambda_2 = c$. This means that arbitrary functions of the quantities

$$p_1 = x - ct, \qquad p_2 = x + ct$$

will be satisfactory solutions of the equation and that the general solution will be

$$u(x,t) = f(x-ct) + g(x+ct), \tag{20.23}$$

where f and g are arbitrary functions. This solution is discussed further in section 20.4. ◄

The method used to obtain the general solution of the wave equation may also be applied straightforwardly to Laplace's equation.

► *Find the general solution of the two-dimensional Laplace equation*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{20.24}$$

Following the established procedure, we look for a solution that is a function $f(p)$ of $p = x + \lambda y$, where from (20.24) λ satisfies

$$1 + \lambda^2 = 0.$$

This requires that $\lambda = \pm i$, and satisfactory variables p are $p = x \pm iy$. The general solution required is therefore, in terms of arbitrary functions f and g,

$$u(x,y) = f(x+iy) + g(x-iy). \blacktriangleleft$$

It will be apparent from the last two examples that the nature of the appropriate linear combination of x and y depends upon whether $B^2 > 4AC$ or $B^2 < 4AC$. This is exactly the same criterion as determines whether the PDE is hyperbolic or elliptic. Hence as a general result, hyperbolic and elliptic equations of the form (20.20), given the restriction that the constants A, B and C are real, have as solutions functions whose arguments have the form $x+\alpha y$ and $x+i\beta y$ respectively, where α and β themselves are real.

The one case not covered by this result is that in which $B^2 = 4AC$, i.e. a parabolic equation. In this case λ_1 and λ_2 are not different and only one suitable combination of x and y results, namely

$$u(x,y) = f(x - (B/2C)y).$$

To find the second part of the general solution we try, in analogy with the corresponding situation for ordinary differential equations, a solution of the form

$$u(x,y) = h(x,y)g(x - (B/2C)y).$$

Substituting this into (20.20) and using $A = B^2/4C$ results in

$$\left(A\frac{\partial^2 h}{\partial x^2} + B\frac{\partial^2 h}{\partial x \partial y} + C\frac{\partial^2 h}{\partial y^2}\right) g = 0.$$

Therefore we require $h(x,y)$ to be any solution of the original PDE. There are several simple solutions of this equation, but as only one is required we take the simplest non-trivial one, $h(x,y) = x$, to give the general solution of the parabolic equation

$$u(x,y) = f(x - (B/2C)y) + xg(x - (B/2C)y). \tag{20.25}$$

We could, of course, have taken $h(x, y) = y$, but this only leads to a solution that is already contained in (20.25).

►*Solve*

$$\frac{\partial^2 u}{\partial x^2} + 2\frac{\partial^2 u}{\partial x \partial y} + \frac{\partial^2 u}{\partial y^2} = 0,$$

subject to the boundary conditions $u(0, y) = 0$ *and* $u(x, 1) = x^2$.

From our general result, functions of $p = x + \lambda y$ will be solutions provided

$$1 + 2\lambda + \lambda^2 = 0,$$

i.e. $\lambda = -1$ and the equation is parabolic. The general solution is therefore

$$u(x, y) = f(x - y) + xg(x - y).$$

The boundary condition $u(0, y) = 0$ implies $f(p) \equiv 0$, whilst $u(x, 1) = x^2$ yields

$$xg(x - 1) = x^2,$$

which gives $g(p) = p + 1$, Therefore the particular solution required is

$$u(x, y) = x(p + 1) = x(x - y + 1). \blacktriangleleft$$

To reinforce the material discussed above we will now give alternative derivations of the general solutions (20.22) and (20.25) by expressing the original PDE in terms of new variables before solving it. The actual solution will then become almost trivial; but, of course, it will be recognised that suitable new variables could hardly have been guessed if it were not for the work already done. This does not detract from the validity of the derivation to be described, only from the likelihood that it would be discovered by inspection.

We start again with (20.20) and change to new variables

$$\zeta = x + \lambda_1 y, \qquad \eta = x + \lambda_2 y.$$

With this change of variables, we have from the chain rule that

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial \zeta} + \frac{\partial}{\partial \eta},$$
$$\frac{\partial}{\partial y} = \lambda_1 \frac{\partial}{\partial \zeta} + \lambda_2 \frac{\partial}{\partial \eta}.$$

Using these and the fact that

$$A + B\lambda_i + C\lambda_i^2 = 0 \quad \text{for } i = 1, 2,$$

equation (20.20) becomes

$$[2A + B(\lambda_1 + \lambda_2) + 2C\lambda_1\lambda_2]\frac{\partial^2 u}{\partial \zeta \partial \eta} = 0.$$

Then, providing the factor in brackets does not vanish, for which the required condition is easily shown to be $B^2 \neq 4AC$, we obtain

$$\frac{\partial^2 u}{\partial \zeta \partial \eta} = 0,$$

which has the successive integrals

$$\frac{\partial u}{\partial \eta} = F(\eta), \quad u(\zeta, \eta) = f(\eta) + g(\zeta).$$

This solution is just the same as (20.22),

$$u(x, y) = f(x + \lambda_2 y) + g(x + \lambda_1 y).$$

If the equation is parabolic (i.e. $B^2 = 4AC$), we instead use the new variables

$$\zeta = x + \lambda y, \qquad \eta = x,$$

and recalling that $\lambda = -(B/2C)$ we can reduce (20.20) to

$$A\frac{\partial^2 u}{\partial \eta^2} = 0.$$

Two straightforward integrations give as the general solution

$$u(\zeta, \eta) = \eta g(\zeta) + f(\zeta),$$

which in terms of x and y has exactly the form of (20.25),

$$u(x, y) = xg(x + \lambda y) + f(x + \lambda y).$$

Finally, as hinted at in subsection 20.3.2 with reference to first-order linear PDEs, some of the methods used to find particular integrals of linear ODEs can be suitably modified to find particular integrals of PDEs of higher order. In simple cases, however, an appropriate solution may often be found by inspection.

►*Find the general solution of*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 6(x + y).$$

Following our previous methods and results, the complementary function is

$$u(x, y) = f(x + iy) + g(x - iy),$$

and only a particular integral remains to be found. By inspection a particular integral of the equation is $u(x, y) = x^3 + y^3$, and so the general solution can be written

$$u(x, y) = f(x + iy) + g(x - iy) + x^3 + y^3. \blacktriangleleft$$

20.4 The wave equation

We have already found that the general solution of the one-dimensional wave equation is

$$u(x,t) = f(x - ct) + g(x + ct), \tag{20.26}$$

where f and g are arbitrary functions. However, the equation is of such general importance that further discussion will not be out of place.

Let us imagine that $u(x,t) = f(x-ct)$ represents the displacement of a string at time t and position x. It is clear that all positions x and times t for which $x-ct =$ constant will have the same instantaneous displacement. But $x - ct =$ constant is exactly the relation between the time and position of an observer travelling with speed c along the positive x-direction. Consequently this moving observer sees a constant displacement of the string, whereas to a stationary observer, the initial profile $u(x,0)$ moves with speed c along the x-axis as if it were a rigid system. Thus $f(x - ct)$ represents a wave form of constant shape travelling along the positive x-axis with speed c, the actual form of the wave depending upon the function f. Similarly, the term $g(x + ct)$ is a constant wave form travelling with speed c in the negative x-direction. The general solution (20.23) represents a superposition of these.

If the functions f and g are the same then the complete solution (20.23) represents identical progressive waves going in opposite directions. This may result in a wave pattern whose profile does not progress, described as a *standing wave*. As a simple example, suppose both $f(p)$ and $g(p)$ have the form[§]

$$f(p) = g(p) = A\cos(kp + \epsilon).$$

Then (20.23) can be written as

$$\begin{aligned} u(x,t) &= A[\cos(kx - kct + \epsilon) + \cos(kx + kct + \epsilon)] \\ &= 2A\cos(kct)\cos(kx + \epsilon). \end{aligned}$$

The important thing to notice is that the shape of the wave pattern, given by the factor in x, is the same at all times but that its amplitude $2A\cos(kct)$ depends upon time. At some points x that satisfy

$$\cos(kx + \epsilon) = 0$$

there is no displacement at any time; such points are called *nodes*.

So far we have not imposed any boundary conditions on the solution (20.26). The problem of finding a solution to the wave equation that satisfies given boundary conditions is normally treated using the method of separation of variables

[§] In the usual notation, k is the wave number ($= 2\pi$/wavelength) and $kc = \omega$, the angular frequency of the wave.

discussed in the next chapter. Nevertheless, we now consider *D'Alembert's solution* $u(x,t)$ of the wave equation subject to initial conditions (boundary conditions) in the following general form:

$$\text{initial displacement, } u(x,0) = \phi(x); \quad \text{initial velocity, } \frac{\partial u(x,0)}{\partial t} = \psi(x).$$

The functions $\phi(x)$ and $\psi(x)$ are given and describe the displacement and velocity of each part of the string at the (arbitrary) time $t = 0$.

It is clear that what we need are the particular forms of the functions f and g in (20.26) that lead to the required values at $t = 0$. This means that

$$\phi(x) = u(x,0) = f(x-0) + g(x+0), \tag{20.27}$$

$$\psi(x) = \frac{\partial u(x,0)}{\partial t} = -cf'(x-0) + cg'(x+0), \tag{20.28}$$

where it should be noted that $f'(x-0)$ stands for $df(p)/dp$ evaluated, after the differentiation, at $p = x - c \times 0$; likewise for $g'(x+0)$.

Looking on the above two left-hand sides as functions of $p = x \pm ct$, but everywhere evaluated at $t = 0$, we may integrate (20.28) between an arbitrary (and irrelevant) lower limit p_0 and an indefinite upper limit p to obtain

$$\frac{1}{c}\int_{p_0}^{p} \psi(q)\,dq + K = -f(p) + g(p),$$

the constant of integration K depending on p_0. Comparing this equation with (20.27), with x replaced by p, we can establish the forms of the functions f and g as

$$f(p) = \frac{\phi(p)}{2} - \frac{1}{2c}\int_{p_0}^{p} \psi(q)\,dq - \frac{K}{2}, \tag{20.29}$$

$$g(p) = \frac{\phi(p)}{2} + \frac{1}{2c}\int_{p_0}^{p} \psi(q)\,dq + \frac{K}{2}. \tag{20.30}$$

Adding (20.29) with $p = x - ct$ to (20.30) with $p = x + ct$ gives as the solution to the original problem

$$u(x,t) = \frac{1}{2}\left[\phi(x-ct) + \phi(x+ct)\right] + \frac{1}{2c}\int_{x-ct}^{x+ct} \psi(q)\,dq, \tag{20.31}$$

in which we notice that all dependence on p_0 has disappeared.

Each of the terms in (20.31) has a fairly straightforward physical interpretation. In each case the factor $1/2$ represents the fact that only half a displacement profile that starts at any particular point on the string travels towards any other position x, the other half travelling away from it. The first term $\frac{1}{2}\phi(x-ct)$ arises from the initial displacement at a distance ct to the left of x; this travels forward arriving at x at time t. Similarly, the second contribution is due to the initial displacement at a distance ct to the right of x. The interpretation of the final term is a little less obvious. It can be viewed as representing the accumulated transverse displacement at position x due to the passage past x of all parts of the initial motion whose effects can reach x within a time t, both backward and forward travelling.

The extension to the three-dimensional wave equation of solutions of the type we have so far encountered presents no serious difficulty. In Cartesian coordinates the three-dimensional wave equation is

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0. \tag{20.32}$$

In close analogy with the one-dimensional case we try solutions that are functions of linear combinations of all four variables,

$$p = lx + my + nz + \mu t.$$

It is clear that a solution $u(x, y, z, t) = f(p)$ will be acceptable provided that

$$\left(l^2 + m^2 + n^2 - \frac{\mu^2}{c^2}\right)\frac{d^2 f(p)}{dp^2} = 0.$$

Thus, as in the one-dimensional case, f can be arbitrary provided that

$$l^2 + m^2 + n^2 = \mu^2/c^2.$$

Using an obvious normalisation, we take $\mu = \pm c$ and l, m, n as three numbers such that

$$l^2 + m^2 + n^2 = 1.$$

In other words (l, m, n) are the Cartesian components of a unit vector $\hat{\mathbf{n}}$ that points along the direction of propagation of the wave. The quantity p can be written in terms of vectors as the scalar expression $p = \hat{\mathbf{n}} \cdot \mathbf{r} \pm ct$, and the general solution of (20.32) is then

$$u(x, y, z, t) = u(\mathbf{r}, t) = f(\hat{\mathbf{n}} \cdot \mathbf{r} - ct) + g(\hat{\mathbf{n}} \cdot \mathbf{r} + ct), \tag{20.33}$$

where $\hat{\mathbf{n}}$ is *any* unit vector. It would perhaps be more transparent to write $\hat{\mathbf{n}}$ explicitly as one of the arguments of u.

20.5 The diffusion equation

One important class of second-order PDEs, which we have not yet considered in detail, is that in which the second derivative with respect to one variable appears, but only the first derivative with respect to another (usually time). This is exemplified by the one-dimensional diffusion equation

$$\kappa\frac{\partial^2 u(x, t)}{\partial x^2} = \frac{\partial u}{\partial t}, \tag{20.34}$$

in which κ is a constant with the dimensions length^2 $\times$ time^{-1}. The physical constants that go to make up κ in a particular case depend upon the nature of the process (e.g. solute diffusion, heat flow, etc.) and the material being described.

With (20.34) we cannot hope to repeat successfully the method of subsection 20.3.3, since now $u(x,t)$ is differentiated a different number of times on the two sides of the equation; any attempted solution in the form $u(x,t) = f(p)$ with $p = ax + bt$ will lead only to an equation in which the form of f cannot be cancelled out. Clearly we must try other methods.

Solutions may be obtained by using the standard method of separation of variables discussed in the next chapter. Alternatively, a simple solution is also given if both sides of (20.34), as it stands, are separately set equal to a constant α (say), so that

$$\frac{\partial^2 u}{\partial x^2} = \frac{\alpha}{\kappa}, \qquad \frac{\partial u}{\partial t} = \alpha.$$

These equations have the general solutions

$$u(x,t) = \frac{\alpha}{2\kappa}x^2 + xg(t) + h(t) \quad \text{and} \quad u(x,t) = \alpha t + m(x)$$

respectively and may be made compatible with each other if $g(t)$ is taken as constant, $g(t) = g$ (where g could be zero), $h(t) = \alpha t$ and $m(x) = (\alpha/2\kappa)x^2 + gx$. An acceptable solution is thus

$$u(x,t) = \frac{\alpha}{2\kappa}x^2 + gx + \alpha t + \text{constant}. \tag{20.35}$$

Let us now return to seeking solutions of equations by combining the independent variables in particular ways. Having seen that a linear combination of x and t will be of no value, we must search for other possible combinations. It has been noted already that κ has the dimensions length^2 $\times$ time^{-1} and so the combination of variables

$$\eta = \frac{x^2}{\kappa t}$$

will be dimensionless. Let us see if we can satisfy (20.34) with a solution of the form $u(x,t) = f(\eta)$. Evaluating the necessary derivatives we have

$$\begin{aligned}
\frac{\partial u}{\partial x} &= \frac{df(\eta)}{d\eta}\frac{\partial \eta}{\partial x} = \frac{2x}{\kappa t}\frac{df(\eta)}{d\eta}, \\
\frac{\partial^2 u}{\partial x^2} &= \frac{2}{\kappa t}\frac{df(\eta)}{d\eta} + \left(\frac{2x}{\kappa t}\right)^2 \frac{d^2 f(\eta)}{d\eta^2}, \\
\frac{\partial u}{\partial t} &= -\frac{x^2}{\kappa t^2}\frac{df(\eta)}{d\eta}.
\end{aligned}$$

Substituting these expressions into (20.34) we find that the new equation can be written entirely in terms of η,

$$4\eta \frac{d^2 f(\eta)}{d\eta^2} + (2+\eta)\frac{df(\eta)}{d\eta} = 0.$$

This is a straightforward ODE, which can be solved as follows. Writing $f'(\eta) = df(\eta)/d\eta$, etc., we have

$$\begin{aligned} \frac{f''(\eta)}{f'(\eta)} &= -\frac{1}{2\eta} - \frac{1}{4} \\ \Rightarrow \quad \ln[\eta^{1/2} f'(\eta)] &= -\frac{\eta}{4} + c \\ \Rightarrow \quad f'(\eta) &= \frac{A}{\eta^{1/2}} \exp\left(\frac{-\eta}{4}\right) \\ \Rightarrow \quad f(\eta) &= A \int_{\eta_0}^{\eta} \mu^{-1/2} \exp\left(\frac{-\mu}{4}\right) d\mu. \end{aligned}$$

If we now write this in terms of a slightly different variable

$$\zeta = \frac{\eta^{1/2}}{2} = \frac{x}{2(\kappa t)^{1/2}},$$

then $d\zeta = \frac{1}{4}\eta^{-1/2}\,d\eta$, and the solution to (20.34) is given by

$$u(x,t) = f(\eta) = g(\zeta) = B \int_{\zeta_0}^{\zeta} \exp(-v^2)\,dv. \tag{20.36}$$

Here B is a constant and it should be noticed that x and t appear on the RHS only in the indefinite upper limit ζ, and then only in the combination $xt^{-1/2}$. If ζ_0 is chosen as zero then $u(x,t)$ is, to within a constant factor,[§] the error function $\operatorname{erf}[x/2(\kappa t)^{1/2}]$, which is tabulated in many reference books. Only non-negative values of x and t are to be considered here, so that $\zeta \geq \zeta_0$.

Let us try to determine what kind of (say) temperature distribution and flow this represents. For definiteness we take $\zeta_0 = 0$. Firstly, since $u(x,t)$ in (20.36) depends only upon the product $xt^{-1/2}$, it is clear that all points x at times t such that $xt^{-1/2}$ has the same value have the same temperature. Put another way, at any specific time t the region having a particular temperature has moved along the positive x-axis a distance proportional to the square root of t. This is a typical *diffusion* process.

Notice that, on the one hand, at $t = 0$ the variable $\zeta \to \infty$ and u becomes quite independent of x (except perhaps at $x = 0$); the solution then represents a uniform spatial temperature distribution. On the other hand, at $x = 0$ we have that $u(x,t)$ is identically zero for all t.

[§] Take $B = 2\pi^{-1/2}$ to give the usual error function normalised in such a way that $\operatorname{erf}(\infty) = 1$. See the Appendix.

►*An infrared laser delivers a pulse of (heat) energy E to a point P on a large insulated sheet of thickness b, thermal conductivity k, specific heat s and density ρ. The sheet is initially at a uniform temperature. If $u(r,t)$ is the excess temperature a time t later, at a point that is a distance r ($\gg b$) from P, then show that a suitable expression for u is*

$$u(r,t) = \frac{\alpha}{t}\exp\left(-\frac{r^2}{2\beta t}\right), \tag{20.37}$$

where α and β are constants. (Note that we use r instead of ρ to denote the radial coordinate in plane polars so as to avoid confusion with the density.)

Further, (i) *show that $\beta = 2k/(s\rho)$;* (ii) *demonstrate that the excess heat energy in the sheet is independent of t, and hence evaluate α; and* (iii) *prove that the total heat flow past any circle of radius r is E.*

The equation to be solved is the heat diffusion equation

$$k\nabla^2 u(\mathbf{r},t) = s\rho\frac{\partial u(\mathbf{r},t)}{\partial t}.$$

Since we only require the solution for $r \gg b$ we can treat the problem as two-dimensional with obvious circular symmetry. Thus only the r-derivative term in the expression for $\nabla^2 u$ is non-zero, giving

$$\frac{k}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) = s\rho\frac{\partial u}{\partial t}, \tag{20.38}$$

where now $u(\mathbf{r},t) = u(r,t)$.

(i) Substituting the given expression (20.37) into (20.38) we obtain

$$\frac{2k\alpha}{\beta t^2}\left(\frac{r^2}{2\beta t} - 1\right)\exp\left(-\frac{r^2}{2\beta t}\right) = \frac{s\rho\alpha}{t^2}\left(\frac{r^2}{2\beta t} - 1\right)\exp\left(-\frac{r^2}{2\beta t}\right),$$

from which we find that (20.37) is a solution, provided $\beta = 2k/(s\rho)$.

(ii) The excess heat in the system at any time t is

$$\begin{aligned} b\rho s\int_0^\infty u(r,t)2\pi r\,dr &= 2\pi b\rho s\alpha\int_0^\infty \frac{r}{t}\exp\left(-\frac{r^2}{2\beta t}\right)dr \\ &= 2\pi b\rho s\alpha\beta. \end{aligned}$$

The excess heat is therefore independent of t and so must be equal to the total heat input E, implying that

$$\alpha = \frac{E}{2\pi b\rho s\beta} = \frac{E}{4\pi bk}.$$

(iii) The total heat flow past a circle of radius r is

$$\begin{aligned} -2\pi rbk\int_0^\infty \frac{\partial u(r,t)}{\partial r}dt &= -2\pi rbk\int_0^\infty \frac{E}{4\pi bkt}\left(\frac{-r}{\beta t}\right)\exp\left(-\frac{r^2}{2\beta t}\right)dt \\ &= E\left[\exp\left(-\frac{r^2}{2\beta t}\right)\right]_0^\infty = E \qquad \text{for all } r. \end{aligned}$$

As we would expect, all the heat energy E deposited by the laser will eventually flow past a circle of any given radius r. ◄

20.6 Characteristics and the existence of solutions

So far in this chapter we have discussed how to find general solutions to various types of first- and second-order linear PDE. Moreover, given a set of boundary conditions we have shown how to find the particular solution (or class of solutions) that satisfies them. For first-order equations, for example, we found that if the value of $u(x, y)$ is specified along some curve in the xy-plane then the solution to the PDE is in general unique, but that if $u(x, y)$ is specified at only a single point then the solution is not unique: there exists a class of particular solutions all of which satisfy the boundary condition. In this section and the next we make more rigorous the notion of the respective types of boundary condition that cause a PDE to have a unique solution, a class of solutions, or no solution at all.

20.6.1 First-order equations

Let us consider the general first-order PDE (20.9) but now write it as

$$A(x, y)\frac{\partial u}{\partial x} + B(x, y)\frac{\partial u}{\partial y} = F(x, y, u). \tag{20.39}$$

Suppose we wish to solve this PDE subject to the boundary condition that $u(x, y) = \phi(s)$ is specified along some curve C in the xy-plane that is described parametrically by the equations $x = x(s)$ and $y = y(s)$, where s is the arc length along C. The variation of u along C is therefore given by

$$\frac{du}{ds} = \frac{\partial u}{\partial x}\frac{dx}{ds} + \frac{\partial u}{\partial y}\frac{dy}{ds} = \frac{d\phi}{ds}. \tag{20.40}$$

We may then solve the two (inhomogeneous) simultaneous linear equations (20.39) and (20.40) for $\partial u/\partial x$ and $\partial u/\partial y$, *unless* the determinant of the coefficients vanishes (see section 8.18), i.e. unless

$$\begin{vmatrix} dx/ds & dy/ds \\ A & B \end{vmatrix} = 0.$$

At each point in the xy-plane this equation determines a set of curves called *characteristic curves* (or just *characteristics*), which thus satisfy

$$B\frac{dx}{ds} - A\frac{dy}{ds} = 0,$$

or, multiplying through by ds/dx and dividing through by A,

$$\frac{dy}{dx} = \frac{B(x, y)}{A(x, y)}. \tag{20.41}$$

However, we have already met (20.41) in subsection 20.3.1 on first-order PDEs, where solutions of the form $u(x, y) = f(p)$, where p is some combination of x and y, were discussed. Comparing (20.41) with (20.12) we see that the characteristics are merely those curves along which p is constant.

Since the partial derivatives $\partial u/\partial x$ and $\partial u/\partial y$ may be evaluated provided the boundary curve C does *not* lie along a characteristic, defining $u(x, y) = \phi(s)$ along C is sufficient to specify the solution to the original problem (equation plus boundary conditions) near the curve C, in terms of a Taylor expansion about C. Therefore the characteristics can be considered as the curves along which information about the solution $u(x, y)$ 'propagates'. This is best understood by using an example.

►*Find the general solution of*

$$x\frac{\partial u}{\partial x} - 2y\frac{\partial u}{\partial y} = 0 \tag{20.42}$$

that takes the value $2y + 1$ *on the line* $x = 1$ between $y = 0$ and $y = 1$.

We solved this problem in subsection 20.3.1 for the case where $u(x, y)$ takes the value $2y + 1$ along the *entire* line $x = 1$. We found then that the general solution to the equation (ignoring boundary conditions) is of the form

$$u(x, y) = f(p) = f(x^2y),$$

for some arbitrary function f. Hence the characteristics of (20.42) are given by $x^2y = c$ where c is a constant; some of these curves are plotted in figure 20.2 for various values of c. Furthermore, we found that the particular solution for which $u(1, y) = 2y + 1$ *for all* y was given by

$$u(x, y) = 2x^2y + 1.$$

In the present case the value of x^2y is fixed by the boundary conditions only between $y = 0$ and $y = 1$. However, since the characteristics are curves along which x^2y, and hence $f(x^2y)$, remains constant, the solution is determined everywhere along any characteristic that intersects the line segment denoting the boundary conditions. Thus $u(x, y) = 2x^2y + 1$ is the particular solution that holds in the shaded region in figure 20.2 (corresponding to $0 \le c \le 1$).

Outside this region, however, the solution is not precisely specified, and any function of the form

$$u(x, y) = 2x^2y + 1 + g(x^2y)$$

will satisfy both the equation and the boundary condition, provided $g(p) = 0$ for $0 \le p \le 1$. ◄

In the above example the boundary curve was not itself a characteristic and furthermore it crossed each characteristic *once only*. For a general boundary curve C this may not be the case. Firstly, if C is itself a characteristic (or is just a single point) then information about the solution cannot 'propagate' away from C, and so the solution remains unspecified everywhere except on C.

The second possibility is that C (although not a characteristic itself) crosses some characteristics more than once, as in figure 20.3. In this case specifying the value of $u(x, y)$ along the curve PQ determines the solution along all the characteristics that intersect it. Therefore, also specifying $u(x, y)$ along QR can *overdetermine* the problem solution and generally results in there being no solution.

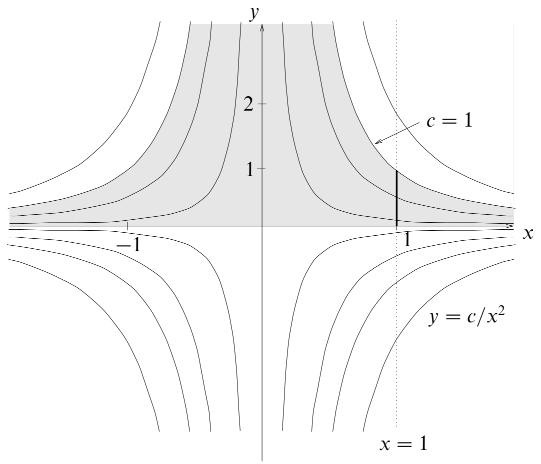


Figure 20.2 The characteristics of equation (20.42). The shaded region shows where the solution to the equation is defined, given the imposed boundary condition at $x = 1$ between $y = 0$ and $y = 1$, shown as a bold vertical line.

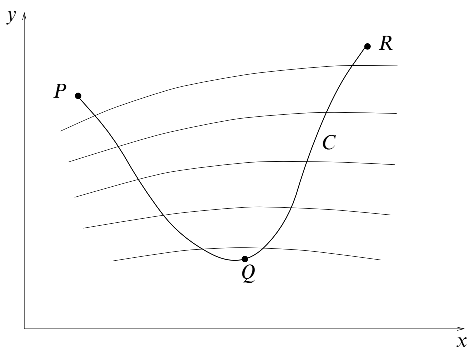


Figure 20.3 A boundary curve C that crosses characteristics more than once.

20.6.2 Second-order equations

The concept of characteristics can be extended naturally to second- (and higher-) order equations. In this case let us write the general second-order linear PDE (20.19) as

$$A(x,y)\frac{\partial^2 u}{\partial x^2} + B(x,y)\frac{\partial^2 u}{\partial x \partial y} + C(x,y)\frac{\partial^2 u}{\partial y^2} = F\left(x, y, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right). \qquad (20.43)$$

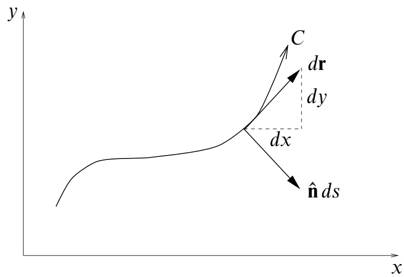


Figure 20.4 A boundary curve C and its tangent and unit normal at a given point.

For second-order equations we might expect that relevant boundary conditions would involve specifying u, or some of its first derivatives, or both, along a suitable set of boundaries bordering or enclosing the region over which a solution is sought. Three common types of boundary condition occur and are associated with the names of Dirichlet, Neumann and Cauchy. They are as follows.

(i) *Dirichlet*: The value of u is specified at each point of the boundary.

(ii) *Neumann*: The value of $\partial u/\partial n$, the *normal derivative* of u, is specified at each point of the boundary. Note that $\partial u/\partial n = \nabla u \cdot \hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the normal to the boundary at each point.

(iii) *Cauchy*: Both u and $\partial u/\partial n$ are specified at each point of the boundary.

Let us consider for the moment the solution of (20.43) subject to the Cauchy boundary conditions, i.e. u and $\partial u/\partial n$ are specified along some boundary curve C in the xy-plane defined by the parametric equations $x = x(s)$, $y = y(s)$, s being the arc length along C (see figure 20.4). Let us suppose that along C we have $u(x, y) = \phi(s)$ and $\partial u/\partial n = \psi(s)$. At any point on C the vector $d\mathbf{r} = dx\,\mathbf{i} + dy\,\mathbf{j}$ is a tangent to the curve and $\hat{\mathbf{n}}\,ds = dy\,\mathbf{i} - dx\,\mathbf{j}$ is a vector normal to the curve. Thus on C we have

$$\frac{\partial u}{\partial s} \equiv \nabla u \cdot \frac{d\mathbf{r}}{ds} = \frac{\partial u}{\partial x}\frac{dx}{ds} + \frac{\partial u}{\partial y}\frac{dy}{ds} = \frac{d\phi(s)}{ds},$$

$$\frac{\partial u}{\partial n} \equiv \nabla u \cdot \hat{\mathbf{n}} = \frac{\partial u}{\partial x}\frac{dy}{ds} - \frac{\partial u}{\partial y}\frac{dx}{ds} = \psi(s).$$

These two equations may then be solved straightforwardly for the first partial derivatives $\partial u/\partial x$ and $\partial u/\partial y$ along C. Using the chain rule to write

$$\frac{d}{ds} = \frac{dx}{ds}\frac{\partial}{\partial x} + \frac{dy}{ds}\frac{\partial}{\partial y},$$

we may differentiate the two first derivatives $\partial u/\partial x$ and $\partial u/\partial y$ along the boundary to obtain the pair of equations

$$\frac{d}{ds}\left(\frac{\partial u}{\partial x}\right) = \frac{dx}{ds}\frac{\partial^2 u}{\partial x^2} + \frac{dy}{ds}\frac{\partial^2 u}{\partial x \partial y},$$
$$\frac{d}{ds}\left(\frac{\partial u}{\partial y}\right) = \frac{dx}{ds}\frac{\partial^2 u}{\partial x \partial y} + \frac{dy}{ds}\frac{\partial^2 u}{\partial y^2}.$$

We may now solve these two equations, together with the original PDE (20.43), for the second partial derivatives of u, *except* where the determinant of their coefficients equals zero,

$$\begin{vmatrix} A & B & C \\ \dfrac{dx}{ds} & \dfrac{dy}{ds} & 0 \\ 0 & \dfrac{dx}{ds} & \dfrac{dy}{ds} \end{vmatrix} = 0.$$

Expanding out the determinant,

$$A\left(\frac{dy}{ds}\right)^2 - B\left(\frac{dx}{ds}\right)\left(\frac{dy}{ds}\right) + C\left(\frac{dx}{ds}\right)^2 = 0.$$

Multiplying through by $(ds/dx)^2$ we obtain

$$A\left(\frac{dy}{dx}\right)^2 - B\frac{dy}{dx} + C = 0, \qquad (20.44)$$

which is the ODE for the curves in the xy-plane along which the second partial derivatives of u *cannot* be found.

As for the first-order case, the curves satisfying (20.44) are called characteristics of the original PDE. These characteristics have tangents at each point given by (when $A \neq 0$)

$$\frac{dy}{dx} = \frac{B \pm \sqrt{B^2 - 4AC}}{2A}. \qquad (20.45)$$

Clearly, when the original PDE is hyperbolic ($B^2 > 4AC$), equation (20.45) defines two families of real curves in the xy-plane; when the equation is parabolic ($B^2 = 4AC$) it defines one family of real curves; and when the equation is elliptic ($B^2 < 4AC$) it defines two families of complex curves. Furthermore, when A, B and C are constants, rather than functions of x and y, the equations of the characteristics will be of the form $x + \lambda y =$ constant, which is reminiscent of the form of solution discussed in subsection 20.3.3.

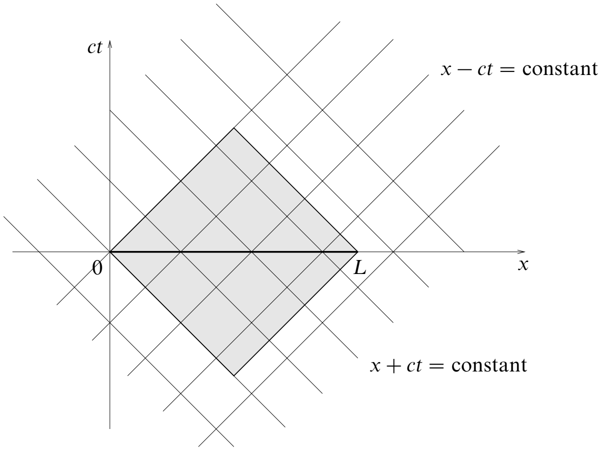


Figure 20.5 The characteristics for the one-dimensional wave equation. The shaded region indicates the region over which the solution is determined by specifying Cauchy boundary conditions at $t = 0$ on the line segment $x = 0$ to $x = L$.

►*Find the characteristics of the one-dimensional wave equation*

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0.$$

This is a hyperbolic equation with $A = 1$, $B = 0$ and $C = -1/c^2$. Therefore from (20.44) the characteristics are given by

$$\left(\frac{dx}{dt}\right)^2 = c^2,$$

and so the characteristics are the straight lines $x - ct =$ constant and $x + ct =$ constant. ◄

The characteristics of second-order PDEs can be considered as the curves along which *partial* information about the solution $u(x, y)$ 'propagates'. Consider a point in the space that has the independent variables as its coordinates; unless both of the two characteristics that pass through the point intersect the curve along which the boundary conditions are specified, the solution will not be determined at that point. In particular, if the equation is hyperbolic, so that we obtain two families of real characteristics in the xy-plane, then Cauchy boundary conditions propagate partial information concerning the solution along the characteristics, belonging to each family, that intersect the boundary curve C. The solution u is then specified in the region common to these two families of characteristics. For instance, the characteristics of the hyperbolic one-dimensional wave equation in the last example are shown in figure 20.5. By specifying Cauchy boundary

Equation type	Boundary	Conditions
hyperbolic	open	Cauchy
parabolic	open	Dirichlet or Neumann
elliptic	closed	Dirichlet or Neumann

Table 20.1 The appropriate boundary conditions for different types of partial differential equation.

conditions u and $\partial u/\partial t$ on the line segment $t = 0$, $x = 0$ to L, the solution is specified in the shaded region.

As in the case of first-order PDEs, however, problems can arise. For example, if for a hyperbolic equation the boundary curve intersects any characteristic more than once then Cauchy conditions along C can overdetermine the problem, resulting in there being no solution. In this case either the boundary curve C must be altered, or the boundary conditions on the offending parts of C must be relaxed to Dirichlet or Neumann conditions.

The general considerations involved in deciding which boundary conditions are appropriate for a particular problem are complex, and we do not discuss them any further here.[§] We merely note that whether the various types of boundary condition are appropriate (in that they give a solution that is unique, sometimes to within a constant, and is well defined) depends upon the type of second-order equation under consideration and on whether the region of solution is bounded by a closed or an open curve (or a surface if there are more than two independent variables). Note that part of a closed boundary may be at infinity if conditions are imposed on u or $\partial u/\partial n$ there.

It may be shown that the appropriate boundary-condition and equation-type pairings are as given in table 20.1.

For example, Laplace's equation $\nabla^2 u = 0$ is elliptic and thus requires either Dirichlet or Neumann boundary conditions on a closed boundary which, as we have already noted, may be at infinity if the behaviour of u is specified there (most often u or $\partial u/\partial n \to 0$ at infinity).

20.7 Uniqueness of solutions

Although we have merely stated the appropriate boundary types and conditions for which, in the general case, a PDE has a unique, well-defined solution, sometimes to within an additive constant, it is often important to be able to prove that a unique solution is obtained.

[§] For a discussion the reader is referred, for example, to P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, Part I* (New York: McGraw-Hill, 1953), chap. 6.

As an important example let us consider Poisson's equation in three dimensions,

$$\nabla^2 u(\mathbf{r}) = \rho(\mathbf{r}), \tag{20.46}$$

with either Dirichlet or Neumann conditions on a closed boundary appropriate to such an elliptic equation; for brevity, in (20.46), we have absorbed any physical constants into ρ. We aim to show that, to within an unimportant constant, the solution of (20.46) is *unique* if either the potential u or its normal derivative $\partial u/\partial n$ is specified on all surfaces bounding a given region of space (including, if necessary, a hypothetical spherical surface of indefinitely large radius on which u or $\partial u/\partial n$ is prescribed to have an arbitrarily small value). Stated more formally this is as follows.

Uniqueness theorem. *If u is real and its first and second partial derivatives are continuous in a region V and on its boundary S, and $\nabla^2 u = \rho$ in V and either $u = f$ or $\partial u/\partial n = g$ on S, where ρ, f and g are prescribed functions, then u is unique (at least to within an additive constant).*

►*Prove the uniqueness theorem for Poisson's equation.*

Let us suppose on the contrary that two solutions $u_1(\mathbf{r})$ and $u_2(\mathbf{r})$ both satisfy the conditions given above, and denote their difference by the function $w = u_1 - u_2$. We then have

$$\nabla^2 w = \nabla^2 u_1 - \nabla^2 u_2 = \rho - \rho = 0,$$

so that w satisfies Laplace's equation in V. Furthermore, since either $u_1 = f = u_2$ or $\partial u_1/\partial n = g = \partial u_2/\partial n$ on S, we must have either $w = 0$ or $\partial w/\partial n = 0$ on S.

If we now use Green's first theorem, (11.19), for the case where both scalar functions are taken as w we have

$$\int_V \left[w\nabla^2 w + (\nabla w)\cdot(\nabla w) \right] dV = \int_S w\frac{\partial w}{\partial n}\, dS.$$

However, either condition, $w = 0$ or $\partial w/\partial n = 0$, makes the RHS vanish whilst the first term on the LHS vanishes since $\nabla^2 w = 0$ in V. Thus we are left with

$$\int_V |\nabla w|^2 \, dV = 0.$$

Since $|\nabla w|^2$ can never be negative, this can only be satisfied if

$$\nabla w = \mathbf{0},$$

i.e. if w, and hence $u_1 - u_2$, is a constant in V.

If Dirichlet conditions are given then $u_1 \equiv u_2$ on (some part of) S and hence $u_1 = u_2$ everywhere in V. For Neumann conditions, however, u_1 and u_2 can differ throughout V by an arbitrary (but unimportant) constant. ◄

The importance of this uniqueness theorem lies in the fact that if a solution to Poisson's (or Laplace's) equation that fits the given set of Dirichlet or Neumann conditions can be found by any means whatever, then that solution is the correct one, since only one exists. This result is the mathematical justification for the *method of images*, which is discussed more fully in the next chapter.

We also note that often the same general method, used in the above example for proving the uniqueness theorem for Poisson's equation, can be employed to prove the uniqueness (or otherwise) of solutions to other equations and boundary conditions.

20.8 Exercises

20.1 Determine whether the following can be written as functions of $p = x^2 + 2y$ only, and hence whether they are solutions of (20.8):

(a) $x^2(x^2 - 4) + 4y(x^2 - 2) + 4(y^2 - 1)$;
(b) $x^4 + 2x^2y + y^2$;
(c) $[x^4 + 4x^2y + 4y^2 + 4]/[2x^4 + x^2(8y + 1) + 8y^2 + 2y]$.

20.2 Find partial differential equations satisfied by the following functions $u(x, y)$ for all arbitrary functions f and all arbitrary constants a and b:

(a) $u(x, y) = f(x^2 - y^2)$;
(b) $u(x, y) = (x - a)^2 + (y - b)^2$;
(c) $u(x, y) = y^n f(y/x)$;
(d) $u(x, y) = f(x + ay)$.

20.3 Solve the following partial differential equations for $u(x, y)$ with the boundary conditions given:

(a) $x\dfrac{\partial u}{\partial x} + xy = u, \quad u = 2y$ on the line $x = 1$;

(b) $1 + x\dfrac{\partial u}{\partial y} = xu, \quad u(x, 0) = x.$

20.4 Find the most general solutions $u(x, y)$ of the following equations, consistent with the boundary conditions stated:

(a) $y\dfrac{\partial u}{\partial x} - x\dfrac{\partial u}{\partial y} = 0, \quad u(x, 0) = 1 + \sin x$;

(b) $i\dfrac{\partial u}{\partial x} = 3\dfrac{\partial u}{\partial y}, \quad u = (4 + 3i)x^2$ on the line $x = y$;

(c) $\sin x \sin y\dfrac{\partial u}{\partial x} + \cos x \cos y\dfrac{\partial u}{\partial y} = 0, \quad u = \cos 2y$ on $x + y = \pi/2$;

(d) $\dfrac{\partial u}{\partial x} + 2x\dfrac{\partial u}{\partial y} = 0, \quad u = 2$ on the parabola $y = x^2$.

20.5 Find solutions of

$$\frac{1}{x}\frac{\partial u}{\partial x} + \frac{1}{y}\frac{\partial u}{\partial y} = 0$$

for which (a) $u(0, y) = y$ and (b) $u(1, 1) = 1$.

20.6 Find the most general solutions $u(x, y)$ of the following equations consistent with the boundary conditions stated:

(a) $y\dfrac{\partial u}{\partial x} - x\dfrac{\partial u}{\partial y} = 3x, \quad u = x^2$ on the line $y = 0$;

(b) $y\dfrac{\partial u}{\partial x} - x\dfrac{\partial u}{\partial y} = 3x, \quad u(1,0) = 2;$

(c) $y^2\dfrac{\partial u}{\partial x} + x^2\dfrac{\partial u}{\partial y} = x^2y^2(x^3+y^3),$ no boundary conditions.

20.7 Solve

$$\sin x\frac{\partial u}{\partial x} + \cos x\frac{\partial u}{\partial y} = \cos x$$

subject to (a) $u(\pi/2, y) = 0$ and (b) $u(\pi/2, y) = y(y+1)$.

20.8 A function $u(x, y)$ satisfies

$$2\frac{\partial u}{\partial x} + 3\frac{\partial u}{\partial y} = 10,$$

and takes the value 3 on the line $y = 4x$. Evaluate $u(2, 4)$.

20.9 If $u(x, y)$ satisfies

$$\frac{\partial^2 u}{\partial x^2} - 3\frac{\partial^2 u}{\partial x\partial y} + 2\frac{\partial^2 u}{\partial y^2} = 0$$

and $u = -x^2$ and $\partial u/\partial y = 0$ for $y = 0$ and all x, find the value of $u(0, 1)$.

20.10 Consider the partial differential equation

$$\frac{\partial^2 u}{\partial x^2} - 3\frac{\partial^2 u}{\partial x\partial y} + 2\frac{\partial^2 u}{\partial y^2} = 0. \qquad (*)$$

(a) Find the function $u(x, y)$ that satisfies (*) and the boundary condition $u = \partial u/\partial y = 1$ when $y = 0$ for all x. Evaluate $u(0, 1)$.

(b) In which region of the xy-plane would u be determined if the boundary condition were $u = \partial u/\partial y = 1$ when $y = 0$ for all $x > 0$?

20.11 In those cases in which it is possible to do so, evaluate $u(2, 2)$, where $u(x, y)$ is the solution of

$$2y\frac{\partial u}{\partial x} - x\frac{\partial u}{\partial y} = xy(2y^2 - x^2)$$

that satisfies the (separate) boundary conditions given below.

(a) $u(x, 1) = x^2$ for all x.
(b) $u(x, 1) = x^2$ for $x \geq 0$.
(c) $u(x, 1) = x^2$ for $0 \leq x \leq 3$.
(d) $u(x, 0) = x$ for $x \geq 0$.
(e) $u(x, 0) = x$ for all x.
(f) $u(1, \sqrt{10}) = 5$.
(g) $u(\sqrt{10}, 1) = 5$.

20.12 Solve

$$6\frac{\partial^2 u}{\partial x^2} - 5\frac{\partial^2 u}{\partial x\partial y} + \frac{\partial^2 u}{\partial y^2} = 14,$$

subject to $u = 2x + 1$ and $\partial u/\partial y = 4 - 6x$, both on the line $y = 0$.

20.13 By changing the independent variables in the previous exercise to

$$\xi = x + 2y \qquad \text{and} \qquad \eta = x + 3y,$$

show that it must be possible to write $14(x^2 + 5xy + 6y^2)$ in the form

$$f_1(x + 2y) + f_2(x + 3y) - (x^2 + y^2),$$

and determine the forms of $f_1(z)$ and $f_2(z)$.

20.14 Solve

$$\frac{\partial^2 u}{\partial x \partial y} + 3\frac{\partial^2 u}{\partial y^2} = x(2y + 3x).$$

20.15 Find the most general solution of $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = x^2 y^2$.

20.16 An infinitely long string on which waves travel at speed c has an initial displacement

$$y(x) = \begin{cases} \sin(\pi x/a), & -a \le x \le a, \\ 0, & |x| > a. \end{cases}$$

It is released from rest at time $t = 0$, and its subsequent displacement is described by $y(x, t)$.

By expressing the initial displacement as one explicit function incorporating Heaviside step functions, find an expression for $y(x, t)$ at a general time $t > 0$. In particular, determine the displacement as a function of time (a) at $x = 0$, (b) at $x = a$, and (c) at $x = a/2$.

20.17 The non-relativistic Schrödinger equation (20.7) is similar to the diffusion equation in having different orders of derivatives in its various terms; this precludes solutions that are arbitrary functions of particular linear combinations of variables. However, since exponential functions do not change their forms under differentiation, solutions in the form of exponential functions of combinations of the variables may still be possible.

Consider the Schrödinger equation for the case of a constant potential, i.e. for a free particle, and show that it has solutions of the form $A \exp(lx + my + nz + \lambda t)$, where the only requirement is that

$$-\frac{\hbar^2}{2m}\left(l^2 + m^2 + n^2\right) = i\hbar\lambda.$$

In particular, identify the equation and wavefunction obtained by taking λ as $-iE/\hbar$, and l, m and n as $ip_x/\hbar, ip_y/\hbar$ and $ip_z/\hbar$, respectively, where E is the energy and $\mathbf{p}$ the momentum of the particle; these identifications are essentially the content of the de Broglie and Einstein relationships.

20.18 Like the Schrödinger equation of the previous exercise, the equation describing the transverse vibrations of a rod,

$$a^4 \frac{\partial^4 u}{\partial x^4} + \frac{\partial^2 u}{\partial t^2} = 0,$$

has different orders of derivatives in its various terms. Show, however, that it has solutions of exponential form, $u(x, t) = A \exp(\lambda x + i\omega t)$, provided that the relation $a^4 \lambda^4 = \omega^2$ is satisfied.

Use a linear combination of such allowed solutions, expressed as the sum of sinusoids and hyperbolic sinusoids of λx, to describe the transverse vibrations of a rod of length L clamped at both ends. At a clamped point both u and $\partial u/\partial x$ must vanish; show that this implies that $\cos(\lambda L)\cosh(\lambda L) = 1$, thus determining the frequencies ω at which the rod can vibrate.

20.19 An incompressible fluid of density ρ and negligible viscosity flows with velocity v along a thin, straight, perfectly light and flexible tube, of cross-section A which is held under tension T. Assume that small transverse displacements u of the tube are governed by

$$\frac{\partial^2 u}{\partial t^2} + 2v\frac{\partial^2 u}{\partial x \partial t} + \left(v^2 - \frac{T}{\rho A}\right)\frac{\partial^2 u}{\partial x^2} = 0.$$

(a) Show that the general solution consists of a superposition of two waveforms travelling with different speeds.

(b) The tube initially has a small transverse displacement $u = a\cos kx$ and is suddenly released from rest. Find its subsequent motion.

20.20 A sheet of material of thickness w, specific heat capacity c and thermal conductivity k is isolated in a vacuum, but its two sides are exposed to fluxes of radiant heat of strengths J_1 and J_2. Ignoring short-term transients, show that the temperature difference between its two surfaces is steady at $(J_2 - J_1)w/2k$, whilst their average temperature increases at a rate $(J_2 + J_1)/cw$.

20.21 In an electrical cable of resistance R and capacitance C, each per unit length, voltage signals obey the equation $\partial^2 V/\partial x^2 = RC\partial V/\partial t$. This has solutions of the form given in (20.36) and also of the form $V = Ax + D$.

(a) Find a combination of these that represents the situation after a steady voltage V_0 is applied at $x = 0$ at time $t = 0$.
(b) Obtain a solution describing the propagation of the voltage signal resulting from the application of the signal $V = V_0$ for $0 < t < T$, $V = 0$ otherwise, to the end $x = 0$ of an infinite cable.
(c) Show that for $t \gg T$ the maximum signal occurs at a value of x proportional to $t^{1/2}$ and has a magnitude proportional to t^{-1}.

20.22 The daily and annual variations of temperature at the surface of the earth may be represented by sine-wave oscillations, with equal amplitudes and periods of 1 day and 365 days respectively. Assume that for (angular) frequency ω the temperature at depth x in the earth is given by $u(x,t) = A\sin(\omega t + \mu x)\exp(-\lambda x)$, where λ and μ are constants.

(a) Use the diffusion equation to find the values of λ and μ.
(b) Find the ratio of the depths below the surface at which the two amplitudes have dropped to 1/20 of their surface values.
(c) At what time of year is the soil coldest at the greater of these depths, assuming that the smoothed annual variation in temperature at the surface has a minimum on February 1st?

20.23 Consider each of the following situations in a qualitative way and determine the equation type, the nature of the boundary curve and the type of boundary conditions involved:

(a) a conducting bar given an initial temperature distribution and then thermally isolated;
(b) two long conducting concentric cylinders, on each of which the voltage distribution is specified;
(c) two long conducting concentric cylinders, on each of which the charge distribution is specified;
(d) a semi-infinite string, the end of which is made to move in a prescribed way.

20.24 *This example gives a formal demonstration that the type of a second-order PDE (elliptic, parabolic or hyperbolic) cannot be changed by a new choice of independent variable. The algebra is somewhat lengthy, but straightforward.*

If a change of variable $\xi = \xi(x,y)$, $\eta = \eta(x,y)$ is made in (20.19), so that it reads

$$A'\frac{\partial^2 u}{\partial \xi^2} + B'\frac{\partial^2 u}{\partial \xi \partial \eta} + C'\frac{\partial^2 u}{\partial \eta^2} + D'\frac{\partial u}{\partial \xi} + E'\frac{\partial u}{\partial \eta} + F'u = R'(\xi,\eta),$$

show that

$$B'^2 - 4A'C' = (B^2 - 4AC)\left[\frac{\partial(\xi,\eta)}{\partial(x,y)}\right]^2.$$

Hence deduce the conclusion stated above.

20.25 The Klein–Gordon equation (which is satisfied by the quantum-mechanical wavefunction $\Phi(\mathbf{r})$ of a relativistic spinless particle of non-zero mass m) is

$$\nabla^2\Phi - m^2\Phi = 0.$$

Show that the solution for the scalar field $\Phi(\mathbf{r})$ in any volume V bounded by a surface S is unique if either Dirichlet or Neumann boundary conditions are specified on S.

20.9 Hints and answers

20.1 (a) Yes, $p^2 - 4p - 4$; (b) no, $(p - y)^2$; (c) yes, $(p^2 + 4)/(2p^2 + p)$.

20.3 Each equation is effectively an ordinary differential equation, but with a function of the non-integrated variable as the constant of integration; (a) $u = xy(2 - \ln x)$; (b) $u = x^{-1}(1 - e^y) + xe^y$.

20.5 (a) $(y^2 - x^2)^{1/2}$; (b) $1 + f(y^2 - x^2)$, where $f(0) = 0$.

20.7 $u = y + f(y - \ln(\sin x))$; (a) $u = \ln(\sin x)$; (b) $u = y + [y - \ln(\sin x)]^2$.

20.9 General solution is $u(x, y) = f(x + y) + g(x + y/2)$. Show that $2p = -g'(p)/2$, and hence $g(p) = k - 2p^2$, whilst $f(p) = p^2 - k$, leading to $u(x, y) = -x^2 + y^2/2$; $u(0, 1) = 1/2$.

20.11 $p = x^2 + 2y^2$; $u(x, y) = f(p) + x^2y^2/2$.

(a) $u(x, y) = (x^2 + 2y^2 + x^2y^2 - 2)/2$; $u(2, 2) = 13$. The line $y = 1$ cuts each characteristic in zero or two distinct points, but this causes no difficulty with the given boundary conditions.
(b) As in (a).
(c) The solution is defined over the space between the ellipses $p = 2$ and $p = 11$; $(2, 2)$ lies on $p = 12$, and so $u(2, 2)$ is undetermined.
(d) $u(x, y) = (x^2 + 2y^2)^{1/2} + x^2y^2/2$; $u(2, 2) = 8 + \sqrt{12}$.
(e) The line $y = 0$ cuts each characteristic in two distinct points. No differentiable form of $f(p)$ gives $f(\pm a) = \pm a$ respectively, and so there is no solution.
(f) The solution is only specified on $p = 21$, and so $u(2, 2)$ is undetermined.
(g) The solution is specified on $p = 12$, and so $u(2, 2) = 5 + \frac{1}{2}(4)(4) = 13$.

20.13 The equation becomes $\partial^2 f/\partial\xi\partial\eta = -14$, with solution $f(\xi, \eta) = f(\xi) + g(\eta) - 14\xi\eta$, which can be compared with the answer from the previous question; $f_1(z) = 10z^2$ and $f_2(z) = 5z^2$.

20.15 $u(x, y) = f(x + iy) + g(x - iy) + (1/12)x^4(y^2 - (1/15)x^2)$. In the last term, x and y may be interchanged. There are (infinitely) many other possibilities for the specific PI, e.g. $[\,15x^2y^2(x^2 + y^2) - (x^6 + y^6)\,]/360$.

20.17 $E = p^2/(2m)$, the relationship between energy and momentum for a non-relativistic particle; $u(\mathbf{r}, t) = A\exp[i(\mathbf{p}\cdot\mathbf{r} - Et)/\hbar]$, a plane wave of wave number $\mathbf{k} = \mathbf{p}/\hbar$ and angular frequency $\omega = E/\hbar$ travelling in the direction $\mathbf{p}/p$.

20.19 (a) $c = v \pm \alpha$ where $\alpha^2 = T/\rho A$;
(b) $u(x, t) = a\cos[k(x - vt)]\cos(k\alpha t) - (va/\alpha)\sin[k(x - vt)]\sin(k\alpha t)$.

20.21 (a) $V_0\left[1 - (2/\sqrt{\pi})\int^{\frac{1}{2}x(CR/t)^{1/2}} \exp(-v^2)\,dv\right]$;
(b) consider the input as equivalent to V_0 applied at $t = 0$ and continued and $-V_0$ applied at $t = T$ and continued;

$$V(x, t) = \frac{2V_0}{\sqrt{\pi}}\int_{\frac{1}{2}x(CR/t)^{1/2}}^{\frac{1}{2}x[CR/(t-T)]^{1/2}} \exp\left(-v^2\right)\,dv;$$

(c) For $t \gg T$, maximum at $x = [2t/(CR)]^{1/2}$ with value $\dfrac{V_0T\exp(-\frac{1}{2})}{(2\pi)^{1/2}t}$.

20.23 (a) Parabolic, open, Dirichlet $u(x, 0)$ given, Neumann $\partial u/\partial x = 0$ at $x = \pm L/2$ for all t;
(b) elliptic, closed, Dirichlet;
(c) elliptic, closed, Neumann $\partial u/\partial n = \sigma/\epsilon_0$;
(d) hyperbolic, open, Cauchy.

20.25 Follow an argument similar to that in section 20.7 and argue that the additional term $\int m^2|w|^2\, dV$ must be zero, and hence that $w = 0$ everywhere.

21

Partial differential equations: separation of variables and other methods

In the previous chapter we demonstrated the methods by which general solutions of some partial differential equations (PDEs) may be obtained in terms of arbitrary functions. In particular, solutions containing the independent variables in definite combinations were sought, thus reducing the effective number of them.

In the present chapter we begin by taking the opposite approach, namely that of trying to keep the independent variables as separate as possible, using the method of separation of variables. We then consider integral transform methods by which one of the independent variables may be eliminated, at least from differential coefficients. Finally, we discuss the use of Green's functions in solving inhomogeneous problems.

21.1 Separation of variables: the general method

Suppose we seek a solution $u(x, y, z, t)$ to some PDE (expressed in Cartesian coordinates). Let us attempt to obtain one that has the product form[§]

$$u(x, y, z, t) = X(x)Y(y)Z(z)T(t). \tag{21.1}$$

A solution that has this form is said to be *separable* in x, y, z and t, and seeking solutions of this form is called the method of *separation of variables*.

As simple examples we may observe that, of the functions

$$\text{(i) } xyz^2 \sin bt, \qquad \text{(ii) } xy + zt, \qquad \text{(iii) } (x^2 + y^2)z \cos \omega t,$$

(i) is completely separable, (ii) is inseparable in that no single variable can be separated out from it and written as a multiplicative factor, whilst (iii) is separable in z and t but not in x and y.

[§] It should be noted that the conventional use here of upper-case (capital) letters to denote the functions of the corresponding lower-case variable is intended to enable an easy correspondence between a function and its argument to be made.

When seeking PDE solutions of the form (21.1), we are requiring not that there is no connection at all between the functions X, Y, Z and T (for example, certain parameters may appear in two or more of them), but only that X does not depend upon y, z, t, that Y does not depend on x, z, t, and so on.

For a general PDE it is likely that a separable solution is impossible, but certainly some common and important equations do have useful solutions of this form, and we will illustrate the method of solution by studying the three-dimensional wave equation

$$\nabla^2 u(\mathbf{r}) = \frac{1}{c^2}\frac{\partial^2 u(\mathbf{r})}{\partial t^2}. \tag{21.2}$$

We will work in Cartesian coordinates for the present and assume a solution of the form (21.1); the solutions in alternative coordinate systems, e.g. spherical or cylindrical polars, are considered in section 21.3. Expressed in Cartesian coordinates (21.2) takes the form

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}; \tag{21.3}$$

substituting (21.1) gives

$$\frac{d^2 X}{dx^2} YZT + X\frac{d^2 Y}{dy^2} ZT + XY\frac{d^2 Z}{dz^2} T = \frac{1}{c^2} XYZ\frac{d^2 T}{dt^2},$$

which can also be written as

$$X''YZT + XY''ZT + XYZ''T = \frac{1}{c^2} XYZT'', \tag{21.4}$$

where in each case the primes refer to the *ordinary* derivative with respect to the independent variable upon which the function depends. This emphasises the fact that each of the functions X, Y, Z and T has only one independent variable and thus its only derivative is its total derivative. For the same reason, in each term in (21.4) three of the four functions are unaltered by the partial differentiation and behave exactly as constant multipliers.

If we now divide (21.4) throughout by $u = XYZT$ we obtain

$$\frac{X''}{X} + \frac{Y''}{Y} + \frac{Z''}{Z} = \frac{1}{c^2}\frac{T''}{T}. \tag{21.5}$$

This form shows the particular characteristic that is the basis of the method of separation of variables, namely that of the four terms the first is a function of x only, the second of y only, the third of z only and the RHS a function of t only and yet there is an equation connecting them. This can only be so for all x, y, z and t if *each* of the terms does not in fact, despite appearances, depend upon the corresponding independent variable but *is equal to a constant*, the four constants being such that (21.5) is satisfied.

Since there is only one equation to be satisfied and four constants involved, there is considerable freedom in the values they may take. For the purposes of our illustrative example let us make the choice of $-l^2$, $-m^2$, $-n^2$, for the first three constants. The constant associated with $c^{-2}T''/T$ must then have the value $-\mu^2 = -(l^2 + m^2 + n^2)$.

Having recognised that each term of (21.5) is individually equal to a constant (or parameter), we can now replace (21.5) by four separate ordinary differential equations (ODEs):

$$\frac{X''}{X} = -l^2, \qquad \frac{Y''}{Y} = -m^2, \qquad \frac{Z''}{Z} = -n^2, \qquad \frac{1}{c^2}\frac{T''}{T} = -\mu^2. \tag{21.6}$$

The important point to notice is not the simplicity of the equations (21.6) (the corresponding ones for a general PDE are usually far from simple) but that, by the device of assuming a separable solution, a *partial* differential equation (21.3), containing derivatives with respect to the four independent variables all in one equation, has been reduced to four *separate ordinary* differential equations (21.6). The ordinary equations are connected through four constant parameters that satisfy an algebraic relation. These constants are called *separation constants*.

The general solutions of the equations (21.6) can be deduced straightforwardly and are

$$\begin{aligned} X(x) &= A\exp(ilx) + B\exp(-ilx), \\ Y(y) &= C\exp(imy) + D\exp(-imy), \\ Z(z) &= E\exp(inz) + F\exp(-inz), \\ T(t) &= G\exp(ic\mu t) + H\exp(-ic\mu t), \end{aligned} \tag{21.7}$$

where $A, B, \ldots, H$ are constants, which may be determined if boundary condtions are imposed on the solution. Depending on the geometry of the problem and any boundary conditions, it is sometimes more appropriate to write the solutions (21.7) in the alternative form

$$\begin{aligned} X(x) &= A'\cos lx + B'\sin lx, \\ Y(y) &= C'\cos my + D'\sin my, \\ Z(z) &= E'\cos nz + F'\sin nz, \\ T(t) &= G'\cos(c\mu t) + H'\sin(c\mu t), \end{aligned} \tag{21.8}$$

for some different set of constants $A', B', \ldots, H'$. Clearly the choice of how best to represent the solution depends on the problem being considered.

As an example, suppose that we take as particular solutions the four functions

$$X(x) = \exp(ilx), \qquad Y(y) = \exp(imy),$$
$$Z(z) = \exp(inz), \qquad T(t) = \exp(-ic\mu t).$$

This gives a particular solution of the original PDE (21.3)

$$\begin{aligned} u(x,y,z,t) &= \exp(ilx)\exp(imy)\exp(inz)\exp(-ic\mu t) \\ &= \exp[i(lx+my+nz-c\mu t)], \end{aligned}$$

which is a special case of the solution (20.33) obtained in the previous chapter and represents a plane wave of unit amplitude propagating in a direction given by the vector with components l, m, n in a Cartesian coordinate system. In the conventional notation of wave theory, l, m and n are the components of the wave-number vector $\mathbf{k}$, whose magnitude is given by $k = 2\pi/\lambda$, where λ is the wavelength of the wave; $c\mu$ is the angular frequency ω of the wave. This gives the equation in the form

$$\begin{aligned} u(x,y,z,t) &= \exp[i(k_x x + k_y y + k_z z - \omega t)] \\ &= \exp[i(\mathbf{k}\cdot\mathbf{r} - \omega t)], \end{aligned}$$

and makes the exponent dimensionless.

The method of separation of variables can be applied to many commonly occurring PDEs encountered in physical applications.

►*Use the method of separation of variables to obtain for the one-dimensional diffusion equation*

$$\kappa\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}, \qquad (21.9)$$

a solution that tends to zero as $t \to \infty$ for all x.

Here we have only two independent variables x and t and we therefore assume a solution of the form

$$u(x,t) = X(x)T(t).$$

Substituting this expression into (21.9) and dividing through by $u = XT$ (and also by κ) we obtain

$$\frac{X''}{X} = \frac{T'}{\kappa T}.$$

Now, arguing exactly as above that the LHS is a function of x only and the RHS is a function of t only, we conclude that each side must equal a constant, which, anticipating the result and noting the imposed boundary condition, we will take as $-\lambda^2$. This gives us two ordinary equations,

$$X'' + \lambda^2 X = 0, \qquad (21.10)$$

$$T' + \lambda^2 \kappa T = 0, \qquad (21.11)$$

which have the solutions

$$\begin{aligned} X(x) &= A\cos\lambda x + B\sin\lambda x, \\ T(t) &= C\exp(-\lambda^2\kappa t). \end{aligned}$$

Combining these to give the assumed solution $u = XT$ yields (absorbing the constant C into A and B)

$$u(x,t) = (A\cos\lambda x + B\sin\lambda x)\exp(-\lambda^2\kappa t). \qquad (21.12)$$

In order to satisfy the boundary condition $u \to 0$ as $t \to \infty$, $\lambda^2\kappa$ must be > 0. Since κ is real and > 0, this implies that λ is a real non-zero number and that the solution is sinusoidal in x and is not a disguised hyperbolic function; this was our reason for choosing the separation constant as $-\lambda^2$. ◄

As a final example we consider Laplace's equation in Cartesian coordinates; this may be treated in a similar manner.

►*Use the method of separation of variables to obtain a solution for the two-dimensional Laplace equation,*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{21.13}$$

If we assume a solution of the form $u(x, y) = X(x)Y(y)$ then, following the above method, and taking the separation constant as λ^2, we find

$$X'' = \lambda^2 X, \qquad Y'' = -\lambda^2 Y.$$

Taking λ^2 as > 0, the general solution becomes

$$u(x, y) = (A\cosh\lambda x + B\sinh\lambda x)(C\cos\lambda y + D\sin\lambda y). \tag{21.14}$$

An alternative form, in which the exponentials are written explicitly, may be useful for other geometries or boundary conditions:

$$u(x, y) = [A\exp\lambda x + B\exp(-\lambda x)](C\cos\lambda y + D\sin\lambda y), \tag{21.15}$$

with different constants A and B.

If $\lambda^2 < 0$ then the roles of x and y interchange. The particular combination of sinusoidal and hyperbolic functions and the values of λ allowed will be determined by the geometrical properties of any specific problem, together with any prescribed or necessary boundary conditions. ◄

We note here that a particular case of the solution (21.14) links up with the 'combination' result $u(x, y) = f(x + iy)$ of the previous chapter (equations (20.24) and following), namely that if $A = B$ and $D = iC$ then the solution is the same as $f(p) = AC\exp\lambda p$ with $p = x + iy$.

21.2 Superposition of separated solutions

It will be noticed in the previous two examples that there is considerable freedom in the values of the separation constant λ, the only essential requirement being that λ has the *same* value in both parts of the solution, i.e. the part depending on x and the part depending on y (or t). This is a general feature for solutions in separated form, which, if the original PDE has n independent variables, will contain $n - 1$ separation constants. All that is required in general is that we associate the correct function of one independent variable with the appropriate functions of the others, the correct function being the one with the same values of the separation constants.

If the original PDE is linear (as are the Laplace, Schrödinger, diffusion and wave equations) then mathematically acceptable solutions can be formed by superposing solutions corresponding to different allowed values of the separation constants. To take a two-variable example: if

$$u_{\lambda_1}(x, y) = X_{\lambda_1}(x)Y_{\lambda_1}(y)$$

is a solution of a linear PDE obtained by giving the separation constant the value λ_1, then the superposition

$$u(x, y) = a_1 X_{\lambda_1}(x)Y_{\lambda_1}(y) + a_2 X_{\lambda_2}(x)Y_{\lambda_2}(y) + \cdots = \sum_i a_i X_{\lambda_i}(x)Y_{\lambda_i}(y) \tag{21.16}$$

is also a solution for any constants a_i, provided that the λ_i are the allowed values of the separation constant λ given the imposed boundary conditions. Note that if the boundary conditions allow any of the separation constants to be zero then the form of the general solution is normally different and must be deduced by returning to the separated ordinary differential equations. We will encounter this behaviour in section 21.3.

The value of the superposition approach is that a boundary condition, say that $u(x, y)$ takes a particular form $f(x)$ when $y = 0$, might be met by choosing the constants a_i such that

$$f(x) = \sum_i a_i X_{\lambda_i}(x)Y_{\lambda_i}(0).$$

In general, this will be possible provided that the functions $X_{\lambda_i}(x)$ form a complete set – as do the sinusoidal functions of Fourier series or the spherical harmonics discussed in subsection 18.3.

►*A semi-infinite rectangular metal plate occupies the region $0 \le x \le \infty$ and $0 \le y \le b$ in the xy-plane. The temperature at the far end of the plate and along its two long sides is fixed at $0\,^\circ$C. If the temperature of the plate at $x = 0$ is also fixed and is given by $f(y)$, find the steady-state temperature distribution u(x,y) of the plate. Hence find the temperature distribution if $f(y) = u_0$, where u_0 is a constant.*

The physical situation is illustrated in figure 21.1. With the notation we have used several times before, the two-dimensional heat diffusion equation satisfied by the temperature $u(x, y, t)$ is

$$\kappa\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = \frac{\partial u}{\partial t},$$

with $\kappa = k/(s\rho)$. In this case, however, we are asked to find the steady-state temperature, which corresponds to $\partial u/\partial t = 0$, and so we are led to consider the (two-dimensional) Laplace equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

We saw that assuming a separable solution of the form $u(x, y) = X(x)Y(y)$ led to solutions such as (21.14) or (21.15), or equivalent forms with x and y interchanged. In the current problem we have to satisfy the boundary conditions $u(x, 0) = 0 = u(x, b)$ and so a solution that is sinusoidal in y seems appropriate. Furthermore, since we require $u(\infty, y) = 0$ it is best to write the x-dependence of the solution explicitly in terms of

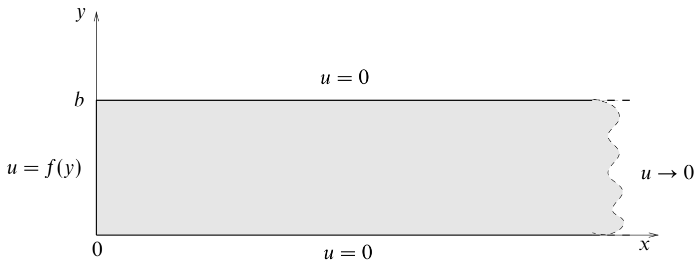


Figure 21.1 A semi-infinite metal plate whose edges are kept at fixed temperatures.

exponentials rather than of hyperbolic functions. We therefore write the separable solution in the form (21.15) as

$$u(x, y) = [A \exp \lambda x + B \exp(-\lambda x)](C \cos \lambda y + D \sin \lambda y).$$

Applying the boundary conditions, we see firstly that $u(\infty, y) = 0$ implies $A = 0$ if we take $\lambda > 0$. Secondly, since $u(x, 0) = 0$ we may set $C = 0$, which, if we absorb the constant D into B, leaves us with

$$u(x, y) = B \exp(-\lambda x) \sin \lambda y.$$

But, using the condition $u(x, b) = 0$, we require $\sin \lambda b = 0$ and so λ must be equal to $n\pi/b$, where n is any positive integer.

Using the principle of superposition (21.16), the general solution satisfying the given boundary conditions can therefore be written

$$u(x, y) = \sum_{n=1}^{\infty} B_n \exp(-n\pi x/b) \sin(n\pi y/b), \tag{21.17}$$

for some constants B_n. Notice that in the sum in (21.17) we have omitted negative values of n since they would lead to exponential terms that diverge as $x \to \infty$. The $n = 0$ term is also omitted since it is identically zero. Using the remaining boundary condition $u(0, y) = f(y)$ we see that the constants B_n must satisfy

$$f(y) = \sum_{n=1}^{\infty} B_n \sin(n\pi y/b). \tag{21.18}$$

This is clearly a Fourier sine series expansion of $f(y)$ (see chapter 12). For (21.18) to hold, however, the continuation of $f(y)$ outside the region $0 \le y \le b$ must be an odd periodic function with period $2b$ (see figure 21.2). We also see from figure 21.2 that if the original function $f(y)$ does not equal zero at either of $y = 0$ and $y = b$ then its continuation has a discontinuity at the corresponding point(s); nevertheless, as discussed in chapter 12, the Fourier series will converge to the mid-points of these jumps and hence tend to zero in this case. If, however, the top and bottom edges of the plate were held not at $0\,^\circ$C but at some other non-zero temperature, then, in general, the final solution would possess discontinuities at the corners $x = 0$, $y = 0$ and $x = 0$, $y = b$.

Bearing in mind these technicalities, the coefficients B_n in (21.18) are given by

$$B_n = \frac{2}{b} \int_0^b f(y) \sin\left(\frac{n\pi y}{b}\right) dy. \tag{21.19}$$

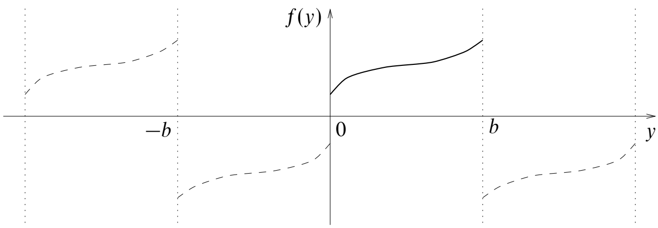


Figure 21.2 The continuation of $f(y)$ for a Fourier sine series.

Therefore, if $f(y) = u_0$ (i.e. the temperature of the side at $x = 0$ is constant along its length), (21.19) becomes

$$\begin{aligned} B_n &= \frac{2}{b}\int_0^b u_0 \sin\left(\frac{n\pi y}{b}\right) dy \\ &= \left[-\frac{2u_0}{b}\frac{b}{n\pi}\cos\left(\frac{n\pi y}{b}\right)\right]_0^b \\ &= -\frac{2u_0}{n\pi}[(-1)^n - 1] = \begin{cases} 4u_0/n\pi & \text{for } n \text{ odd,} \\ 0 & \text{for } n \text{ even.} \end{cases} \end{aligned}$$

Therefore the required solution is

$$u(x, y) = \sum_{n \text{ odd}} \frac{4u_0}{n\pi} \exp\left(-\frac{n\pi x}{b}\right) \sin\left(\frac{n\pi y}{b}\right). \blacktriangleleft$$

In the above example the boundary conditions meant that one term in each part of the separable solution could be immediately discarded, making the problem much easier to solve. Sometimes, however, a little ingenuity is required in writing the separable solution in such a way that certain parts can be neglected immediately.

►*Suppose that the semi-infinite rectangular metal plate in the previous example is replaced by one that in the x-direction has finite length a. The temperature of the right-hand edge is fixed at 0 °C and all other boundary conditions remain as before. Find the steady-state temperature in the plate.*

As in the previous example, the boundary conditions $u(x, 0) = 0 = u(x, b)$ suggest a solution that is sinusoidal in y. In this case, however, we require $u = 0$ on $x = a$ (rather than at infinity) and so a solution in which the x-dependence is written in terms of hyperbolic functions, such as (21.14), rather than exponentials is more appropriate. Moreover, since the constants in front of the hyperbolic functions are, at this stage, arbitrary, we may write the separable solution in the most convenient way that ensures that the condition $u(a, y) = 0$ is straightforwardly satisfied. We therefore write

$$u(x, y) = [A\cosh\lambda(a - x) + B\sinh\lambda(a - x)](C\cos\lambda y + D\sin\lambda y).$$

Now the condition $u(a, y) = 0$ is easily satisfied by setting $A = 0$. As before the conditions $u(x, 0) = 0 = u(x, b)$ imply $C = 0$ and $\lambda = n\pi/b$ for integer n. Superposing the solutions for different n we then obtain

$$u(x, y) = \sum_{n=1}^{\infty} B_n \sinh[n\pi(a - x)/b] \sin(n\pi y/b), \tag{21.20}$$

for some constants B_n. We have omitted negative values of n in the sum (21.20) since the relevant terms are already included in those obtained for positive n. Again the $n = 0$ term is identically zero. Using the final boundary condition $u(0, y) = f(y)$ as above we find that the constants B_n must satisfy

$$f(y) = \sum_{n=1}^{\infty} B_n \sinh(n\pi a/b) \sin(n\pi y/b),$$

and, remembering the caveats discussed in the previous example, the B_n are therefore given by

$$B_n = \frac{2}{b \sinh(n\pi a/b)} \int_0^b f(y) \sin(n\pi y/b)\, dy. \tag{21.21}$$

For the case where $f(y) = u_0$, following the working of the previous example gives (21.21) as

$$B_n = \frac{4u_0}{n\pi \sinh(n\pi a/b)} \quad \text{for } n \text{ odd}, \qquad B_n = 0 \quad \text{for } n \text{ even}. \tag{21.22}$$

The required solution is thus

$$u(x, y) = \sum_{n \text{ odd}} \frac{4u_0}{n\pi \sinh(n\pi a/b)} \sinh[n\pi(a - x)/b] \sin\left(n\pi y/b\right).$$

We note that, as required, in the limit $a \to \infty$ this solution tends to the solution of the previous example. ◄

Often the principle of superposition can be used to write the solution to problems with more complicated boundary conditions as the sum of solutions to problems that each satisfy only some part of the boundary condition but when added togther satisfy all the conditions.

►*Find the steady-state temperature in the (finite) rectangular plate of the previous example, subject to the boundary conditions $u(x, b) = 0$, $u(a, y) = 0$ and $u(0, y) = f(y)$ as before, but now, in addition, $u(x, 0) = g(x)$.*

Figure 21.3(c) shows the imposed boundary conditions for the metal plate. Although we could find a solution to this problem using the methods presented above, we can arrive at the answer almost immediately by using the principle of superposition and the result of the previous example.

Let us suppose the required solution $u(x, y)$ is made up of two parts:

$$u(x, y) = v(x, y) + w(x, y),$$

where $v(x, y)$ is the solution satisfying the boundary conditions shown in figure 21.3(a),

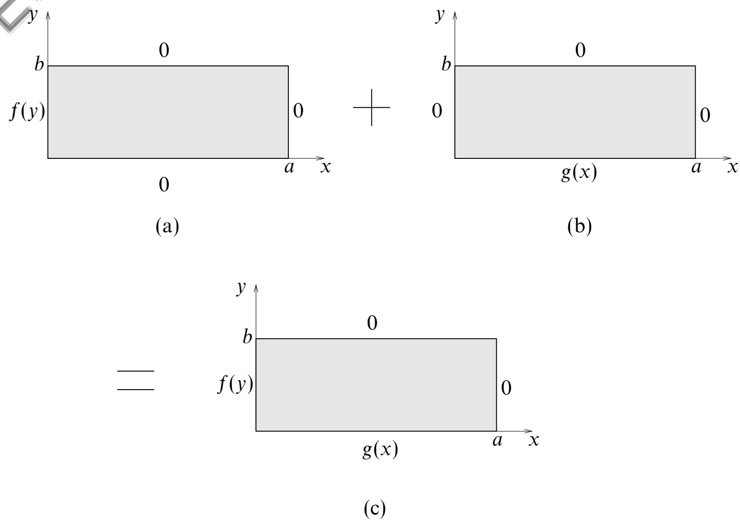


Figure 21.3 Superposition of boundary conditions for a metal plate.

whilst $w(x, y)$ is the solution satisfying the boundary conditions in figure 21.3(b). It is clear that $v(x, y)$ is simply given by the solution to the previous example,

$$v(x, y) = \sum_{n \text{ odd}} B_n \sinh\left[\frac{n\pi(a - x)}{b}\right] \sin\left(\frac{n\pi y}{b}\right),$$

where B_n is given by (21.21). Moreover, by symmetry, $w(x, y)$ must be of the same form as $v(x, y)$ but with x and a interchanged with y and b, respectively, and with $f(y)$ in (21.21) replaced by $g(x)$. Therefore the required solution can be written down immediately without further calculation as

$$u(x, y) = \sum_{n \text{ odd}} B_n \sinh\left[\frac{n\pi(a - x)}{b}\right] \sin\left(\frac{n\pi y}{b}\right) + \sum_{n \text{ odd}} C_n \sinh\left[\frac{n\pi(b - y)}{a}\right] \sin\left(\frac{n\pi x}{a}\right),$$

the B_n being given by (21.21) and the C_n by

$$C_n = \frac{2}{a \sinh(n\pi b/a)} \int_0^a g(x) \sin(n\pi x/a)\, dx.$$

Clearly, this method may be extended to cases in which three or four sides of the plate have non-zero boundary conditions. ◀

As a final example of the usefulness of the principle of superposition we now consider a problem that illustrates how to deal with inhomogeneous boundary conditions by a suitable change of variables.

►*A bar of length L is initially at a temperature of* $0\,°\mathrm{C}$. *One end of the bar* ($x = 0$) *is held at* $0\,°\mathrm{C}$ *and the other is supplied with heat at a constant rate per unit area of* H. *Find the temperature distribution within the bar after a time* t.

With our usual notation, the heat diffusion equation satisfied by the temperature $u(x,t)$ is

$$\kappa\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t},$$

with $\kappa = k/(s\rho)$, where k is the thermal conductivity of the bar, s is its specific heat capacity and ρ is its density.

The boundary conditions can be written as

$$u(x,0) = 0, \qquad u(0,t) = 0, \qquad \frac{\partial u(L,t)}{\partial x} = \frac{H}{k},$$

the last of which is inhomogeneous. In general, inhomogeneous boundary conditions can cause difficulties and it is usual to attempt a transformation of the problem into an equivalent homogeneous one. To this end, let us assume that the solution to our problem takes the form

$$u(x,t) = v(x,t) + w(x),$$

where the function $w(x)$ is to be suitably determined. In terms of v and w the problem becomes

$$\kappa\left(\frac{\partial^2 v}{\partial x^2} + \frac{d^2 w}{dx^2}\right) = \frac{\partial v}{\partial t},$$
$$v(x,0) + w(x) = 0,$$
$$v(0,t) + w(0) = 0,$$
$$\frac{\partial v(L,t)}{\partial x} + \frac{dw(L)}{dx} = \frac{H}{k}.$$

There are several ways of choosing $w(x)$ so as to make the new problem straightforward. Using some physical insight, however, it is clear that ultimately (at $t = \infty$), when all transients have died away, the end $x = L$ will attain a temperature u_0 such that $ku_0/L = H$ and there will be a constant temperature gradient $u(x,\infty) = u_0 x/L$. We therefore choose

$$w(x) = \frac{Hx}{k}.$$

Since the second derivative of $w(x)$ is zero, v satisfies the diffusion equation and the boundary conditions on v are now given by

$$v(x,0) = -\frac{Hx}{k}, \qquad v(0,t) = 0, \qquad \frac{\partial v(L,t)}{\partial x} = 0,$$

which are homogeneous in x.

From (21.12) a separated solution for the one-dimensional diffusion equation is

$$v(x,t) = (A\cos\lambda x + B\sin\lambda x)\exp(-\lambda^2\kappa t),$$

corresponding to a separation constant $-\lambda^2$. If we restrict λ to be real then all these solutions are transient ones decaying to zero as $t \to \infty$. These are just what is required to add to $w(x)$ to give the correct solution as $t \to \infty$. In order to satisfy $v(0,t) = 0$, however, we require $A = 0$. Furthermore, since

$$\frac{\partial v}{\partial x} = B\exp(-\lambda^2\kappa t)\lambda\cos\lambda x,$$

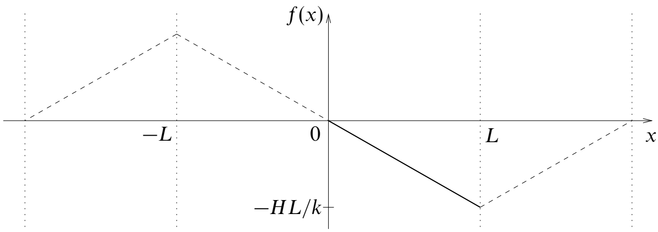


Figure 21.4 The appropriate continuation for a Fourier series containing only sine terms.

in order to satisfy $\partial v(L,t)/\partial x = 0$ we require $\cos \lambda L = 0$, and so λ is restricted to the values

$$\lambda = \frac{n\pi}{2L},$$

where n is an odd non-negative integer, i.e. $n = 1, 3, 5, \ldots$.

Thus, to satisfy the boundary condition $v(x,0) = -Hx/k$, we must have

$$\sum_{n \text{ odd}} B_n \sin\left(\frac{n\pi x}{2L}\right) = -\frac{Hx}{k},$$

in the range $x = 0$ to $x = L$. In this case we must be more careful about the continuation of the function $-Hx/k$, for which the Fourier sine series is required. We want a series that is odd in x (sine terms only) and continuous as $x = 0$ and $x = L$ (no discontinuities, since the series must converge at the end-points). This leads to a continuation of the function as shown in figure 21.4, with a period of $L' = 4L$. Following the discussion of section 12.3, since this continuation is odd about $x = 0$ and even about $x = L'/4 = L$ it can indeed be expressed as a Fourier sine series containing only odd-numbered terms.

The corresponding Fourier series coefficients are found to be

$$B_n = \frac{-8HL}{k\pi^2} \frac{(-1)^{(n-1)/2}}{n^2} \qquad \text{for } n \text{ odd},$$

and thus the final formula for $u(x,t)$ is

$$u(x,t) = \frac{Hx}{k} - \frac{8HL}{k\pi^2} \sum_{n \text{ odd}} \frac{(-1)^{(n-1)/2}}{n^2} \sin\left(\frac{n\pi x}{2L}\right) \exp\left(-\frac{kn^2\pi^2 t}{4L^2 s\rho}\right),$$

giving the temperature for all positions $0 \leq x \leq L$ and for all times $t \geq 0$. ◄

We note that in all the above examples the boundary conditions restricted the separation constant(s) to an infinite number of *discrete* values, usually integers. If, however, the boundary conditions allow the separation constant(s) λ to take a *continuum* of values then the summation in (21.16) is replaced by an integral over λ. This is discussed further in connection with integral transform methods in section 21.4.

21.3 Separation of variables in polar coordinates

So far we have considered the solution of PDEs only in Cartesian coordinates, but many systems in two and three dimensions are more naturally expressed in some form of polar coordinates, in which full advantage can be taken of any inherent symmetries. For example, the potential associated with an isolated point charge has a very simple expression, $q/(4\pi\epsilon_0 r)$, when polar coordinates are used, but involves all three coordinates and square roots when Cartesians are employed. For these reasons we now turn to the separation of variables in plane polar, cylindrical polar and spherical polar coordinates.

Most of the PDEs we have considered so far have involved the operator ∇^2, e.g. the wave equation, the diffusion equation, Schrödinger's equation and Poisson's equation (and of course Laplace's equation). It is therefore appropriate that we recall the expressions for ∇^2 when expressed in polar coordinate systems. From chapter 10, in plane polars, cylindrical polars and spherical polars, respectively, we have

$$\nabla^2 = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2}, \tag{21.23}$$

$$\nabla^2 = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2} + \frac{\partial^2}{\partial z^2}, \tag{21.24}$$

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}. \tag{21.25}$$

Of course the first of these may be obtained from the second by taking z to be identically zero.

21.3.1 Laplace's equation in polar coordinates

The simplest of the equations containing ∇^2 is Laplace's equation,

$$\nabla^2 u(\mathbf{r}) = 0. \tag{21.26}$$

Since it contains most of the essential features of the other more complicated equations, we will consider its solution first.

Laplace's equation in plane polars

Suppose that we need to find a solution of (21.26) that has a prescribed behaviour on the circle $\rho = a$ (e.g. if we are finding the shape taken up by a circular drumskin when its rim is slightly deformed from being planar). Then we may seek solutions of (21.26) that are separable in ρ and ϕ (measured from some arbitrary radius as $\phi = 0$) and hope to accommodate the boundary condition by examining the solution for $\rho = a$.

Thus, writing $u(\rho, \phi) = P(\rho)\Phi(\phi)$ and using the expression (21.23), Laplace's equation (21.26) becomes

$$\frac{\Phi}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial P}{\partial \rho}\right) + \frac{P}{\rho^2}\frac{\partial^2 \Phi}{\partial \phi^2} = 0.$$

Now, employing the same device as previously, that of dividing through by $u = P\Phi$ and multiplying through by ρ^2, results in the separated equation

$$\frac{\rho}{P}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial P}{\partial \rho}\right) + \frac{1}{\Phi}\frac{\partial^2 \Phi}{\partial \phi^2} = 0.$$

Following our earlier argument, since the first term on the RHS is a function of ρ only, whilst the second term depends only on ϕ, we obtain the two *ordinary* equations

$$\frac{\rho}{P}\frac{d}{d\rho}\left(\rho\frac{dP}{d\rho}\right) = n^2, \qquad (21.27)$$

$$\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = -n^2, \qquad (21.28)$$

where we have taken the separation constant to have the form n^2 for later convenience; for the present, n is a general (complex) number.

Let us first consider the case in which $n \neq 0$. The second equation, (21.28), then has the general solution

$$\Phi(\phi) = A\exp(in\phi) + B\exp(-in\phi). \qquad (21.29)$$

Equation (21.27), on the other hand, is the homogeneous equation

$$\rho^2 P'' + \rho P' - n^2 P = 0,$$

which must be solved either by trying a power solution in ρ or by making the substitution $\rho = \exp t$ as described in subsection 15.2.1 and so reducing it to an equation with constant coefficients. Carrying out this procedure we find

$$P(\rho) = C\rho^n + D\rho^{-n}. \qquad (21.30)$$

Returning to the solution (21.29) of the azimuthal equation (21.28), we can see that if Φ, and hence u, is to be single-valued and so not change when ϕ increases by 2π then n must be an integer. Mathematically, other values of n are permissible, but for the description of real physical situations it is clear that this limitation must be imposed. Having thus restricted the possible values of n in one part of the solution, the same limitations must be carried over into the radial part, (21.30). Thus we may write a particular solution of the two-dimensional Laplace equation as

$$u(\rho, \phi) = (A\cos n\phi + B\sin n\phi)(C\rho^n + D\rho^{-n}),$$

where A, B, C, D are arbitrary constants and n is any integer.

We have not yet, however, considered the solution when $n = 0$. In this case, the solutions of the separated ordinary equations (21.28) and (21.27), respectively, are easily shown to be

$$\Phi(\phi) = A\phi + B,$$
$$P(\rho) = C \ln \rho + D.$$

But, in order that $u = P\Phi$ is single-valued, we require $A = 0$, and so the solution for $n = 0$ is simply (absorbing B into C and D)

$$u(\rho, \phi) = C \ln \rho + D.$$

Superposing the solutions for the different allowed values of n, we can write the general solution to Laplace's equation in plane polars as

$$u(\rho, \phi) = (C_0 \ln \rho + D_0) + \sum_{n=1}^{\infty} (A_n \cos n\phi + B_n \sin n\phi)(C_n \rho^n + D_n \rho^{-n}), \tag{21.31}$$

where n can take only integer values. Negative values of n have been omitted from the sum since they are already included in the terms obtained for positive n. We note that, since $\ln \rho$ is singular at $\rho = 0$, whenever we solve Laplace's equation in a region containing the origin, C_0 must be identically zero.

►*A circular drumskin has a supporting rim at $\rho = a$. If the rim is twisted so that it is displaced vertically by a small amount $\epsilon(\sin\phi + 2\sin 2\phi)$, where ϕ is the azimuthal angle with respect to a given radius, find the resulting displacement $u(\rho, \phi)$ over the entire drumskin.*

The transverse displacement of a circular drumskin is usually described by the two-dimensional wave equation. In this case, however, there is no time dependence and so $u(\rho, \phi)$ solves the two-dimensional Laplace equation, subject to the imposed boundary condition.

Referring to (21.31), since we wish to find a solution that is finite everywhere inside $\rho = a$, we require $C_0 = 0$ and $D_n = 0$ for all $n > 0$. Now the boundary condition at the rim requires

$$u(a, \phi) = D_0 + \sum_{n=1}^{\infty} C_n a^n (A_n \cos n\phi + B_n \sin n\phi) = \epsilon(\sin\phi + 2\sin 2\phi).$$

Firstly we see that we require $D_0 = 0$ and $A_n = 0$ for all n. Furthermore, we must have $C_1 B_1 a = \epsilon$, $C_2 B_2 a^2 = 2\epsilon$ and $B_n = 0$ for $n > 2$. Hence the appropriate shape for the drumskin (valid over the whole skin, not just the rim) is

$$u(\rho, \phi) = \frac{\epsilon\rho}{a}\sin\phi + \frac{2\epsilon\rho^2}{a^2}\sin 2\phi = \frac{\epsilon\rho}{a}\left(\sin\phi + \frac{2\rho}{a}\sin 2\phi\right). \blacktriangleleft$$

Laplace's equation in cylindrical polars

Passing to three dimensions, we now consider the solution of Laplace's equation in cylindrical polar coordinates,

$$\frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial u}{\partial \rho}\right)+\frac{1}{\rho^2}\frac{\partial^2 u}{\partial \phi^2}+\frac{\partial^2 u}{\partial z^2}=0. \tag{21.32}$$

We note here that, even when considering a cylindrical physical system, if there is no dependence of the physical variables on z (i.e. along the length of the cylinder) then the problem may be treated using two-dimensional plane polars, as discussed above.

For the more general case, however, we proceed as previously by trying a solution of the form

$$u(\rho,\phi,z)=P(\rho)\Phi(\phi)Z(z),$$

which, on substitution into (21.32) and division through by $u=P\Phi Z$, gives

$$\frac{1}{P\rho}\frac{d}{d\rho}\left(\rho\frac{dP}{d\rho}\right)+\frac{1}{\Phi\rho^2}\frac{d^2\Phi}{d\phi^2}+\frac{1}{Z}\frac{d^2Z}{dz^2}=0.$$

The last term depends only on z, and the first and second (taken together) depend only on ρ and ϕ. Taking the separation constant to be k^2, we find

$$\frac{1}{Z}\frac{d^2Z}{dz^2}=k^2,$$

$$\frac{1}{P\rho}\frac{d}{d\rho}\left(\rho\frac{dP}{d\rho}\right)+\frac{1}{\Phi\rho^2}\frac{d^2\Phi}{d\phi^2}+k^2=0.$$

The first of these equations has the straightforward solution

$$Z(z)=E\exp(-kz)+F\exp kz.$$

Multiplying the second equation through by ρ^2, we obtain

$$\frac{\rho}{P}\frac{d}{d\rho}\left(\rho\frac{dP}{d\rho}\right)+\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}+k^2\rho^2=0,$$

in which the second term depends only on Φ and the other terms depend only on ρ. Taking the second separation constant to be m^2, we find

$$\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}=-m^2, \tag{21.33}$$

$$\rho\frac{d}{d\rho}\left(\rho\frac{dP}{d\rho}\right)+(k^2\rho^2-m^2)P=0. \tag{21.34}$$

The equation in the azimuthal angle ϕ has the very familiar solution

$$\Phi(\phi)=C\cos m\phi+D\sin m\phi.$$

As in the two-dimensional case, single-valuedness of u requires that m is an integer. However, in the particular case $m = 0$ the solution is

$$\Phi(\phi) = C\phi + D.$$

This form is appropriate to a solution with axial symmetry ($C = 0$) or one that is multivalued, but manageably so, such as the magnetic scalar potential associated with a current I (in which case $C = I/(2\pi)$ and D is arbitrary).

Finally, the ρ-equation (21.34) may be transformed into Bessel's equation of order m by writing $\mu = k\rho$. This has the solution

$$P(\rho) = AJ_m(k\rho) + BY_m(k\rho).$$

The properties of these functions were investigated in chapter 16 and will not be pursued here. We merely note that $Y_m(k\rho)$ is singular at $\rho = 0$, and so, when seeking solutions to Laplace's equation in cylindrical coordinates within some region containing the $\rho = 0$ axis, we require $B = 0$.

The complete separated-variable solution in cylindrical polars of Laplace's equation $\nabla^2 u = 0$ is thus given by

$$u(\rho, \phi, z) = [AJ_m(k\rho) + BY_m(k\rho)][C\cos m\phi + D\sin m\phi][E\exp(-kz) + F\exp kz]. \tag{21.35}$$

Of course we may use the principle of superposition to build up more general solutions by adding together solutions of the form (21.35) for all allowed values of the separation constants k and m.

►*A semi-infinite solid cylinder of radius a has its curved surface held at $0\,°$C and its base held at a temperature T_0. Find the steady-state temperature distribution in the cylinder.*

The physical situation is shown in figure 21.5. The steady-state temperature distribution $u(\rho, \phi, z)$ must satisfy Laplace's equation subject to the imposed boundary conditions. Let us take the cylinder to have its base in the $z = 0$ plane and to extend along the positive z-axis. From (21.35), in order that u is finite everywhere in the cylinder we immediately require $B = 0$ and $F = 0$. Furthermore, since the boundary conditions, and hence the temperature distribution, are axially symmetric, we require $m = 0$, and so the general solution must be a superposition of solutions of the form $J_0(k\rho)\exp(-kz)$ for all allowed values of the separation constant k.

The boundary condition $u(a, \phi, z) = 0$ restricts the allowed values of k, since we must have $J_0(ka) = 0$. The zeros of Bessel functions are given in most books of mathematical tables, and we find that, to two decimal places,

$$J_0(x) = 0 \qquad \text{for } x = 2.40,\ 5.52,\ 8.65,\ \ldots.$$

Writing the allowed values of k as k_n for $n = 1, 2, 3, \ldots$ (so, for example, $k_1 = 2.40/a$), the required solution takes the form

$$u(\rho, \phi, z) = \sum_{n=1}^{\infty} A_n J_0(k_n\rho)\exp(-k_n z).$$

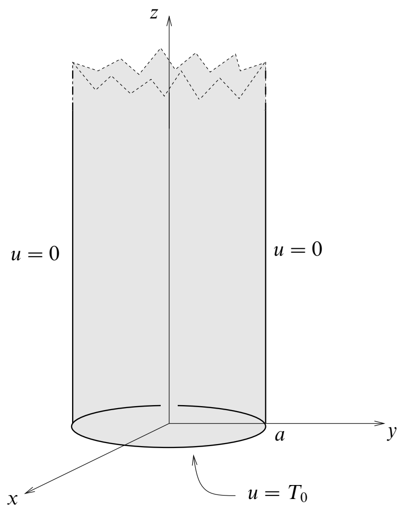


Figure 21.5 A uniform metal cylinder whose curved surface is kept at 0 °C and whose base is held at a temperature T_0.

By imposing the remaining boundary condition $u(\rho, \phi, 0) = T_0$, the coefficients A_n can be found in a similar way to Fourier coefficients but this time by exploiting the orthogonality of the Bessel functions, as discussed in chapter 16. From this boundary condition we require

$$u(\rho, \phi, 0) = \sum_{n=1}^{\infty} A_n J_0(k_n \rho) = T_0.$$

If we multiply this expression by $\rho J_0(k_r \rho)$ and integrate from $\rho = 0$ to $\rho = a$, and use the orthogonality of the Bessel functions $J_0(k_n \rho)$, then the coefficients are given by (18.91) as

$$A_n = \frac{2T_0}{a^2 J_1^2(k_n a)} \int_0^a J_0(k_n \rho) \rho \, d\rho. \tag{21.36}$$

The integral on the RHS can be evaluated using the recurrence relation (18.92) of chapter 16,

$$\frac{d}{dz}[z J_1(z)] = z J_0(z),$$

which on setting $z = k_n \rho$ yields

$$\frac{1}{k_n} \frac{d}{d\rho}[k_n \rho J_1(k_n \rho)] = k_n \rho J_0(k_n \rho).$$

Therefore the integral in (21.36) is given by

$$\int_0^a J_0(k_n \rho) \rho \, d\rho = \left[\frac{1}{k_n} \rho J_1(k_n \rho) \right]_0^a = \frac{1}{k_n} a J_1(k_n a),$$

and the coefficients A_n may be expressed as

$$A_n = \frac{2T_0}{a^2 J_1^2(k_n a)} \left[\frac{a J_1(k_n a)}{k_n} \right] = \frac{2T_0}{k_n a J_1(k_n a)}.$$

The steady-state temperature in the cylinder is then given by

$$u(\rho, \phi, z) = \sum_{n=1}^{\infty} \frac{2T_0}{k_n a J_1(k_n a)} J_0(k_n \rho) \exp(-k_n z). \blacktriangleleft$$

We note that if, in the above example, the base of the cylinder were not kept at a uniform temperature T_0, but instead had some fixed temperature distribution $T(\rho, \phi)$, then the solution of the problem would become more complicated. In such a case, the required temperature distribution $u(\rho, \phi, z)$ is in general *not* axially symmetric, and so the separation constant m is not restricted to be zero but may take any integer value. The solution will then take the form

$$u(\rho, \phi, z) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} J_m(k_{nm}\rho)(C_{nm} \cos m\phi + D_{nm} \sin m\phi) \exp(-k_{nm} z),$$

where the separation constants k_{nm} are such that $J_m(k_{nm}a) = 0$, i.e. $k_{nm}a$ is the nth zero of the mth-order Bessel function. At the base of the cylinder we would then require

$$u(\rho, \phi, 0) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} J_m(k_{nm}\rho)(C_{nm} \cos m\phi + D_{nm} \sin m\phi) = T(\rho, \phi). \tag{21.37}$$

The coefficients C_{nm} could be found by multiplying (21.37) by $J_q(k_{rq}\rho) \cos q\phi$, integrating with respect to ρ and ϕ over the base of the cylinder and exploiting the orthogonality of the Bessel functions and of the trigonometric functions. The D_{nm} could be found in a similar way by multiplying (21.37) by $J_q(k_{rq}\rho) \sin q\phi$.

Laplace's equation in spherical polars

We now come to an equation that is very widely applicable in physical science, namely $\nabla^2 u = 0$ in spherical polar coordinates:

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial u}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial u}{\partial \theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 u}{\partial \phi^2} = 0. \tag{21.38}$$

Our method of procedure will be as before; we try a solution of the form

$$u(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi).$$

Substituting this in (21.38), dividing through by $u = R\Theta\Phi$ and multiplying by r^2, we obtain

$$\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{1}{\Theta \sin\theta} \frac{d}{d\theta} \left(\sin\theta \frac{d\Theta}{d\theta} \right) + \frac{1}{\Phi \sin^2\theta} \frac{d^2\Phi}{d\phi^2} = 0. \tag{21.39}$$

The first term depends only on r and the second and third terms (taken together) depend only on θ and ϕ. Thus (21.39) is equivalent to the two equations

$$\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) = \lambda, \tag{21.40}$$

$$\frac{1}{\Theta \sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) + \frac{1}{\Phi\sin^2\theta}\frac{d^2\Phi}{d\phi^2} = -\lambda. \tag{21.41}$$

Equation (21.40) is a homogeneous equation,

$$r^2\frac{d^2R}{dr^2} + 2r\frac{dR}{dr} - \lambda R = 0,$$

which can be reduced, by the substitution $r = \exp t$ (and writing $R(r) = S(t)$), to

$$\frac{d^2S}{dt^2} + \frac{dS}{dt} - \lambda S = 0.$$

This has the straightforward solution

$$S(t) = A\exp\lambda_1 t + B\exp\lambda_2 t,$$

and so the solution to the radial equation is

$$R(r) = Ar^{\lambda_1} + Br^{\lambda_2},$$

where $\lambda_1 + \lambda_2 = -1$ and $\lambda_1\lambda_2 = -\lambda$. We can thus take λ_1 and λ_2 as given by ℓ and $-(\ell+1)$; λ then has the form $\ell(\ell+1)$. (It should be noted that at this stage nothing has been either assumed or proved about whether ℓ is an integer.)

Hence we have obtained some information about the first factor in the separated-variable solution, which will now have the form

$$u(r,\theta,\phi) = \left[Ar^{\ell} + Br^{-(\ell+1)}\right]\Theta(\theta)\Phi(\phi), \tag{21.42}$$

where Θ and Φ must satisfy (21.41) with $\lambda = \ell(\ell+1)$.

The next step is to take (21.41) further. Multiplying through by $\sin^2\theta$ and substituting for λ, it too takes a separated form:

$$\left[\frac{\sin\theta}{\Theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) + \ell(\ell+1)\sin^2\theta\right] + \frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = 0. \tag{21.43}$$

Taking the separation constant as m^2, the equation in the azimuthal angle ϕ has the same solution as in cylindrical polars, namely

$$\Phi(\phi) = C\cos m\phi + D\sin m\phi.$$

As before, single-valuedness of u requires that m is an integer; for $m = 0$ we again have $\Phi(\phi) = C\phi + D$.

Having settled the form of $\Phi(\phi)$, we are left only with the equation satisfied by $\Theta(\theta)$, which is

$$\frac{\sin\theta}{\Theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\ell(\ell+1)\sin^2\theta=m^2. \tag{21.44}$$

A change of independent variable from θ to $\mu=\cos\theta$ will reduce this to a form for which solutions are known, and of which some study has been made in chapter 16. Putting

$$\mu=\cos\theta, \qquad \frac{d\mu}{d\theta}=-\sin\theta, \qquad \frac{d}{d\theta}=-(1-\mu^2)^{1/2}\frac{d}{d\mu},$$

the equation for $M(\mu)\equiv\Theta(\theta)$ reads

$$\frac{d}{d\mu}\left[(1-\mu^2)\frac{dM}{d\mu}\right]+\left[\ell(\ell+1)-\frac{m^2}{1-\mu^2}\right]M=0. \tag{21.45}$$

This equation is the *associated Legendre equation*, which was mentioned in subsection 18.2 in the context of Sturm–Liouville equations.

We recall that for the case $m=0$, (21.45) reduces to Legendre's equation, which was studied at length in chapter 16, and has the solution

$$M(\mu)=EP_\ell(\mu)+FQ_\ell(\mu). \tag{21.46}$$

We have not solved (21.45) explicitly for general m, but the solutions were given in subsection 18.2 and are the associated Legendre functions $P_\ell^m(\mu)$ and $Q_\ell^m(\mu)$, where

$$P_\ell^m(\mu)=(1-\mu^2)^{|m|/2}\frac{d^{|m|}}{d\mu^{|m|}}P_\ell(\mu), \tag{21.47}$$

and similarly for $Q_\ell^m(\mu)$. We then have

$$M(\mu)=EP_\ell^m(\mu)+FQ_\ell^m(\mu); \tag{21.48}$$

here m must be an integer, $0\le|m|\le\ell$. We note that if we require solutions to Laplace's equation that are finite when $\mu=\cos\theta=\pm1$ (i.e. on the polar axis where $\theta=0,\pi$), then we must have $F=0$ in (21.46) and (21.48) since $Q_\ell^m(\mu)$ diverges at $\mu=\pm1$.

It will be remembered that one of the important conditions for obtaining finite polynomial solutions of Legendre's equation is that ℓ is an integer ≥ 0. This condition therefore applies also to the solutions (21.46) and (21.48) and is reflected back into the radial part of the general solution given in (21.42).

Now that the solutions of each of the three ordinary differential equations governing R, Θ and Φ have been obtained, we may assemble a complete separated-variable solution of Laplace's equation in spherical polars. It is

$$u(r,\theta,\phi) = (Ar^{\ell} + Br^{-(\ell+1)})(C\cos m\phi + D\sin m\phi)[EP_{\ell}^{m}(\cos\theta) + FQ_{\ell}^{m}(\cos\theta)], \tag{21.49}$$

where the three bracketted factors are connected only through the *integer* parameters ℓ and m, $0 \le |m| \le \ell$. As before, a general solution may be obtained by superposing solutions of this form for the allowed values of the separation constants ℓ and m. As mentioned above, if the solution is required to be finite on the polar axis then $F = 0$ for all ℓ and m.

►*An uncharged conducting sphere of radius a is placed at the origin in an initially uniform electrostatic field E. Show that it behaves as an electric dipole.*

The uniform field, taken in the direction of the polar axis, has an electrostatic potential

$$u = -Ez = -Er\cos\theta,$$

where u is arbitrarily taken as zero at $z = 0$. This satisfies Laplace's equation $\nabla^2 u = 0$, as must the potential v when the sphere is present; for large r the asymptotic form of v must still be $-Er\cos\theta$.

Since the problem is clearly axially symmetric, we have immediately that $m = 0$, and since we require v to be finite on the polar axis we must have $F = 0$ in (21.49). Therefore the solution must be of the form

$$v(r,\theta,\phi) = \sum_{\ell=0}^{\infty}(A_\ell r^\ell + B_\ell r^{-(\ell+1)})P_\ell(\cos\theta).$$

Now the $\cos\theta$-dependence of v for large r indicates that the (θ,ϕ)-dependence of $v(r,\theta,\phi)$ is given by $P_1^0(\cos\theta) = \cos\theta$. Thus the r-dependence of v must also correspond to an $\ell = 1$ solution, and the most general such solution (outside the sphere, i.e. for $r \ge a$) is

$$v(r,\theta,\phi) = (A_1 r + B_1 r^{-2})P_1(\cos\theta).$$

The asymptotic form of v for large r immediately gives $A_1 = -E$ and so yields the solution

$$v(r,\theta,\phi) = \left(-Er + \frac{B_1}{r^2}\right)\cos\theta.$$

Since the sphere is conducting, it is an equipotential region and so v must not depend on θ for $r = a$. This can only be the case if $B_1/a^2 = Ea$, thus fixing B_1. The final solution is therefore

$$v(r,\theta,\phi) = -Er\left(1 - \frac{a^3}{r^3}\right)\cos\theta.$$

Since a dipole of moment p gives rise to a potential $p/(4\pi\epsilon_0 r^2)$, this result shows that the sphere behaves as a dipole of moment $4\pi\epsilon_0 a^3 E$, because of the charge distribution induced on its surface; see figure 21.6. ◄

Often the boundary conditions are not so easily met, and it is necessary to use the mutual orthogonality of the associated Legendre functions (and the trigonometric functions) to obtain the coefficients in the general solution.

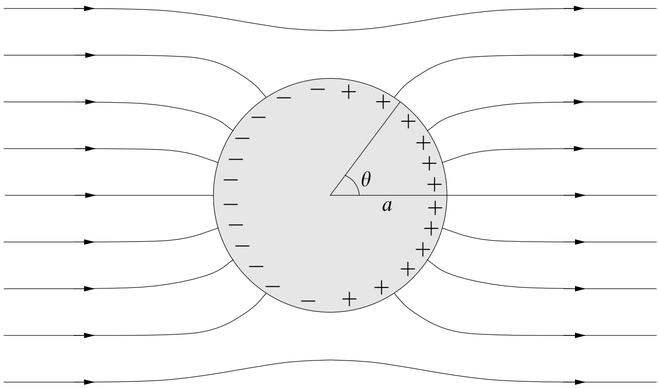


Figure 21.6 Induced charge and field lines associated with a conducting sphere placed in an initially uniform electrostatic field.

►*A hollow split conducting sphere of radius a is placed at the origin. If one half of its surface is charged to a potential v_0 and the other half is kept at zero potential, find the potential v inside and outside the sphere.*

Let us choose the top hemisphere to be charged to v_0 and the bottom hemisphere to be at zero potential, with the plane in which the two hemispheres meet perpendicular to the polar axis; this is shown in figure 21.7. The boundary condition then becomes

$$v(a,\theta,\phi) = \begin{cases} v_0 & \text{for } 0 < \theta < \pi/2 \quad (0 < \cos\theta < 1), \\ 0 & \text{for } \pi/2 < \theta < \pi \quad (-1 < \cos\theta < 0). \end{cases} \tag{21.50}$$

The problem is clearly axially symmetric and so we may set $m = 0$. Also, we require the solution to be finite on the polar axis and so it cannot contain $Q_\ell(\cos\theta)$. Therefore the general form of the solution to (21.38) is

$$v(r,\theta,\phi) = \sum_{\ell=0}^{\infty} (A_\ell r^\ell + B_\ell r^{-(\ell+1)}) P_\ell(\cos\theta). \tag{21.51}$$

Inside the sphere (for $r < a$) we require the solution to be finite at the origin and so $B_\ell = 0$ for all ℓ in (21.51). Imposing the boundary condition at $r = a$ we must then have

$$v(a,\theta,\phi) = \sum_{\ell=0}^{\infty} A_\ell a^\ell P_\ell(\cos\theta),$$

where $v(a,\theta,\phi)$ is also given by (21.50). Exploiting the mutual orthogonality of the Legendre polynomials, the coefficients in the Legendre polynomial expansion are given by (18.14) as (writing $\mu = \cos\theta$)

$$\begin{aligned} A_\ell a^\ell &= \frac{2\ell+1}{2} \int_{-1}^{1} v(a,\theta,\phi) P_\ell(\mu)\, d\mu \\ &= \frac{2\ell+1}{2} v_0 \int_0^1 P_\ell(\mu)\, d\mu, \end{aligned}$$

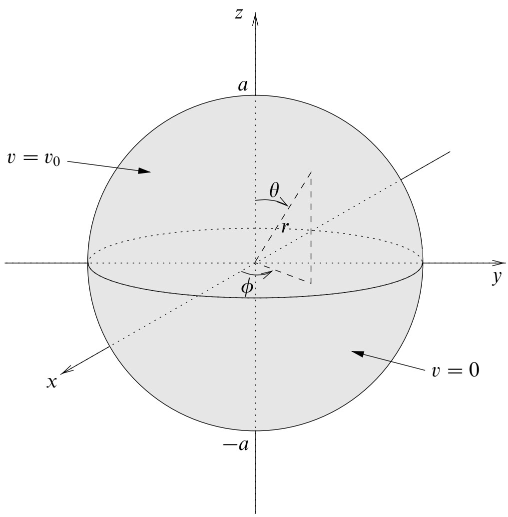


Figure 21.7 A hollow split conducting sphere with its top half charged to a potential v_0 and its bottom half at zero potential.

where in the last line we have used (21.50). The integrals of the Legendre polynomials are easily evaluated (see exercise 17.3) and we find

$$A_0 = \frac{v_0}{2}, \qquad A_1 = \frac{3v_0}{4a}, \qquad A_2 = 0, \qquad A_3 = -\frac{7v_0}{16a^3}, \qquad \cdots,$$

so that the required solution inside the sphere is

$$v(r,\theta,\phi) = \frac{v_0}{2}\left[1 + \frac{3r}{2a}P_1(\cos\theta) - \frac{7r^3}{8a^3}P_3(\cos\theta) + \cdots\right].$$

Outside the sphere (for $r > a$) we require the solution to be bounded as r tends to infinity and so in (21.51) we must have $A_\ell = 0$ for all ℓ. In this case, by imposing the boundary condition at $r = a$ we require

$$v(a,\theta,\phi) = \sum_{\ell=0}^{\infty} B_\ell a^{-(\ell+1)} P_\ell(\cos\theta),$$

where $v(a,\theta,\phi)$ is given by (21.50). Following the above argument the coefficients in the expansion are given by

$$B_\ell a^{-(\ell+1)} = \frac{2\ell+1}{2}v_0 \int_0^1 P_\ell(\mu)\,d\mu,$$

so that the required solution outside the sphere is

$$v(r,\theta,\phi) = \frac{v_0 a}{2r}\left[1 + \frac{3a}{2r}P_1(\cos\theta) - \frac{7a^3}{8r^3}P_3(\cos\theta) + \cdots\right]. ◀$$

In the above example, on the equator of the sphere (i.e. at $r = a$ and $\theta = \pi/2$) the potential is given by

$$v(a, \pi/2, \phi) = v_0/2,$$

i.e. mid-way between the potentials of the top and bottom hemispheres. This is so because a Legendre polynomial expansion of a function behaves in the same way as a Fourier series expansion, in that it converges to the average of the two values at any discontinuities present in the original function.

If the potential on the surface of the sphere had been given as a function of θ *and* ϕ, then we would have had to consider a double series summed over ℓ and m (for $-\ell \leq m \leq \ell$), since, in general, the solution would not have been axially symmetric.

Finally, we note in general that, when obtaining solutions of Laplace's equation in spherical polar coordinates, one finds that, for solutions that are finite on the polar axis, the angular part of the solution is given by

$$\Theta(\theta)\Phi(\phi) = P_\ell^m(\cos\theta)(C\cos m\phi + D\sin m\phi),$$

where ℓ and m are integers with $-\ell \leq m \leq \ell$. This general form is sufficiently common that particular functions of θ and ϕ called *spherical harmonics* are defined and tabulated (see section 18.3).

21.3.2 Other equations in polar coordinates

The development of the solutions of $\nabla^2 u = 0$ carried out in the previous subsection can be employed to solve other equations in which the ∇^2 operator appears. Since we have discussed the general method in some depth already, only an outline of the solutions will be given here.

Let us first consider the wave equation

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}, \tag{21.52}$$

and look for a separated solution of the form $u = F(\mathbf{r})T(t)$, so that initially we are separating only the spatial and time dependences. Substituting this form into (21.52) and taking the separation constant as k^2 we obtain

$$\nabla^2 F + k^2 F = 0, \qquad \frac{d^2 T}{dt^2} + k^2 c^2 T = 0. \tag{21.53}$$

The second equation has the simple solution

$$T(t) = A\exp(i\omega t) + B\exp(-i\omega t), \tag{21.54}$$

where $\omega = kc$; this may also be expressed in terms of sines and cosines, of course. The first equation in (21.53) is referred to as *Helmholtz's equation*; we discuss it below.

We may treat the diffusion equation

$$\kappa \nabla^2 u = \frac{\partial u}{\partial t}$$

in a similar way. Separating the spatial and time dependences by assuming a solution of the form $u = F(\mathbf{r})T(t)$, and taking the separation constant as k^2, we find

$$\nabla^2 F + k^2 F = 0, \qquad \frac{dT}{dt} + k^2 \kappa T = 0.$$

Just as in the case of the wave equation, the spatial part of the solution satisfies Helmholtz's equation. It only remains to consider the time dependence, which has the simple solution

$$T(t) = A \exp(-k^2 \kappa t).$$

Helmholtz's equation is clearly of central importance in the solutions of the wave and diffusion equations. It can be solved in polar coordinates in much the same way as Laplace's equation, and indeed reduces to Laplace's equation when $k = 0$. Therefore, we will merely sketch the method of its solution in each of the three polar coordinate systems.

Helmholtz's equation in plane polars

In two-dimensional plane polar coordinates, Helmholtz's equation takes the form

$$\frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial F}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 F}{\partial \phi^2} + k^2 F = 0.$$

If we try a separated solution of the form $F(\mathbf{r}) = P(\rho)\Phi(\phi)$, and take the separation constant as m^2, we find

$$\frac{d^2\Phi}{d\phi^2} + m^2 \phi = 0,$$

$$\frac{d^2 P}{d\rho^2} + \frac{1}{\rho}\frac{dP}{d\rho} + \left(k^2 - \frac{m^2}{\rho^2}\right) P = 0.$$

As for Laplace's equation, the angular part has the familiar solution (if $m \neq 0$)

$$\Phi(\phi) = A \cos m\phi + B \sin m\phi,$$

or an equivalent form in terms of complex exponentials. The radial equation differs from that found in the solution of Laplace's equation, but by making the substitution $\mu = k\rho$ it is easily transformed into Bessel's equation of order m (discussed in chapter 16), and has the solution

$$P(\rho) = C J_m(k\rho) + D Y_m(k\rho),$$

where Y_m is a Bessel function of the second kind, which is infinite at the origin and is not to be confused with a spherical harmonic (these are written with a superscript as well as a subscript).

Putting the two parts of the solution together we have

$$F(\rho,\phi) = [A\cos m\phi + B\sin m\phi][CJ_m(k\rho) + DY_m(k\rho)]. \tag{21.55}$$

Clearly, for solutions of Helmholtz's equation that are required to be finite at the origin, we must set $D = 0$.

►*Find the four lowest frequency modes of oscillation of a circular drumskin of radius a whose circumference is held fixed in a plane.*

The transverse displacement $u(\mathbf{r}, t)$ of the drumskin satisfies the two-dimensional wave equation

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2},$$

with $c^2 = T/\sigma$, where T is the tension of the drumskin and σ is its mass per unit area. From (21.54) and (21.55) a separated solution of this equation, in plane polar coordinates, that is finite at the origin is

$$u(\rho,\phi,t) = J_m(k\rho)(A\cos m\phi + B\sin m\phi)\exp(\pm i\omega t),$$

where $\omega = kc$. Since we require the solution to be single-valued we must have m as an integer. Furthermore, if the drumskin is clamped at its outer edge $\rho = a$ then we also require $u(a,\phi,t) = 0$. Thus we need

$$J_m(ka) = 0,$$

which in turn restricts the allowed values of k. The zeros of Bessel functions can be obtained from most books of tables; the first few are

$$\begin{aligned}
J_0(x) &= 0 \quad \text{for } x \approx 2.40,\ 5.52,\ 8.65, \ldots,\\
J_1(x) &= 0 \quad \text{for } x \approx 3.83,\ 7.02,\ 10.17, \ldots,\\
J_2(x) &= 0 \quad \text{for } x \approx 5.14,\ 8.42,\ 11.62 \ldots .
\end{aligned}$$

The smallest value of x for which any of the Bessel functions is zero is $x \approx 2.40$, which occurs for $J_0(x)$. Thus the lowest-frequency mode has $k = 2.40/a$ and angular frequency $\omega = 2.40c/a$. Since $m = 0$ for this mode, the shape of the drumskin is

$$u \propto J_0\left(2.40\frac{\rho}{a}\right);$$

this is illustrated in figure 21.8.

Continuing in the same way, the next three modes are given by

$$\begin{aligned}
\omega &= 3.83\frac{c}{a}, &\quad u &\propto J_1\left(3.83\frac{\rho}{a}\right)\cos\phi, &\quad J_1\left(3.83\frac{\rho}{a}\right)\sin\phi;\\
\omega &= 5.14\frac{c}{a}, &\quad u &\propto J_2\left(5.14\frac{\rho}{a}\right)\cos 2\phi, &\quad J_2\left(5.14\frac{\rho}{a}\right)\sin 2\phi;\\
\omega &= 5.52\frac{c}{a}, &\quad u &\propto J_0\left(5.52\frac{\rho}{a}\right).
\end{aligned}$$

These modes are also shown in figure 21.8. We note that the second and third frequencies have *two* corresponding modes of oscillation; these frequencies are therefore two-fold degenerate. ◄

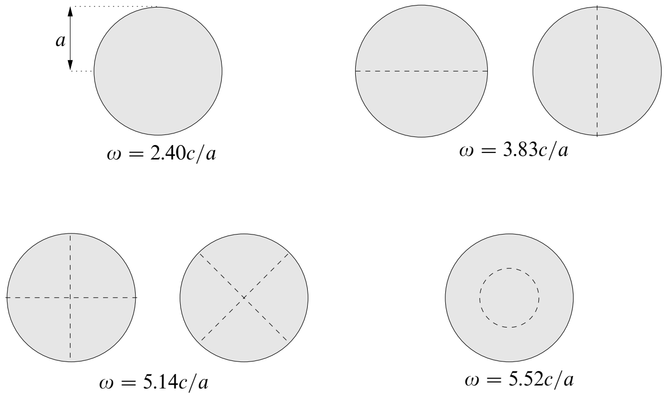


Figure 21.8 The modes of oscillation with the four lowest frequencies for a circular drumskin of radius a. The dashed lines indicate the nodes, where the displacement of the drumskin is always zero.

Helmholtz's equation in cylindrical polars

Generalising the above method to three-dimensional cylindrical polars is straightforward, and following a similar procedure to that used for Laplace's equation we find the separated solution of Helmholtz's equation takes the form

$$F(\rho, \phi, z) = \left[AJ_m\left(\sqrt{k^2-\alpha^2}\,\rho\right) + BY_m\left(\sqrt{k^2-\alpha^2}\,\rho\right)\right] \\ \times (C\cos m\phi + D\sin m\phi)[E\exp(i\alpha z) + F\exp(-i\alpha z)],$$

where α and m are separation constants. We note that the angular part of the solution is the same as for Laplace's equation in cylindrical polars.

Helmholtz's equation in spherical polars

In spherical polars, we find again that the angular parts of the solution $\Theta(\theta)\Phi(\phi)$ are identical to those of Laplace's equation in this coordinate system, i.e. they are the spherical harmonics $Y_\ell^m(\theta, \phi)$, and so we shall not discuss them further.

The radial equation in this case is given by

$$r^2R'' + 2rR' + [k^2r^2 - \ell(\ell+1)]R = 0, \tag{21.56}$$

which has an additional term k^2r^2R compared with the radial equation for the Laplace solution. The equation (21.56) looks very much like Bessel's equation. In fact, by writing $R(r) = r^{-1/2}S(r)$ and making the change of variable $\mu = kr$, it can be reduced to Bessel's equation of order $\ell + \frac{1}{2}$, which has as its solutions $S(\mu) = J_{\ell+1/2}(\mu)$ and $Y_{\ell+1/2}(\mu)$ (see section 18.6). The separated solution to

Helmholtz's equation in spherical polars is thus

$$F(r,\theta,\phi) = r^{-1/2}[AJ_{\ell+1/2}(kr) + BY_{\ell+1/2}(kr)](C\cos m\phi + D\sin m\phi) \\ \times[EP_\ell^m(\cos\theta) + FQ_\ell^m(\cos\theta)]. \qquad (21.57)$$

For solutions that are finite at the origin we require $B = 0$, and for solutions that are finite on the polar axis we require $F = 0$. It is worth mentioning that the solutions proportional to $r^{-1/2}J_{\ell+1/2}(kr)$ and $r^{-1/2}Y_{\ell+1/2}(kr)$, when suitably normalised, are called *spherical Bessel functions* of the first and second kind, respectively, and are denoted by $j_\ell(kr)$ and $n_\ell(\mu)$ (see section 18.6).

As mentioned at the beginning of this subsection, the separated solution of the wave equation in spherical polars is the product of a time-dependent part (21.54) and a spatial part (21.57). It will be noticed that, although this solution corresponds to a solution of definite frequency $\omega = kc$, the zeros of the radial function $j_\ell(kr)$ are not equally spaced in r, except for the case $\ell = 0$ involving $j_0(kr)$, and so there is no precise wavelength associated with the solution.

To conclude this subsection, let us mention briefly the Schrödinger equation for the electron in a hydrogen atom, the nucleus of which is taken at the origin and is assumed massive compared with the electron. Under these circumstances the Schrödinger equation is

$$-\frac{\hbar^2}{2m}\nabla^2 u - \frac{e^2}{4\pi\epsilon_0}\frac{u}{r} = i\hbar\frac{\partial u}{\partial t}.$$

For a 'stationary-state' solution, for which the energy is a constant E and the time-dependent factor T in u is given by $T(t) = A\exp(-iEt/\hbar)$, the above equation is similar to, but not quite the same as, the Helmholtz equation.§ However, as with the wave equation, the angular parts of the solution are identical to those for Laplace's equation and are expressed in terms of spherical harmonics.

The important point to note is that for *any* equation involving ∇^2, provided θ and ϕ do not appear in the equation other than as part of ∇^2, a separated-variable solution in spherical polars will always lead to spherical harmonic solutions. This is the case for the Schrödinger equation describing an atomic electron in a central potential $V(r)$.

21.3.3 Solution by expansion

It is sometimes possible to use the uniqueness theorem discussed in the previous chapter, together with the results of the last few subsections, in which Laplace's equation (and other equations) were considered in polar coordinates, to obtain solutions of such equations appropriate to particular physical situations.

§ For the solution by series of the r-equation in this case the reader may consult, for example, L. Schiff, *Quantum Mechanics* (New York: McGraw-Hill, 1955), p. 82.

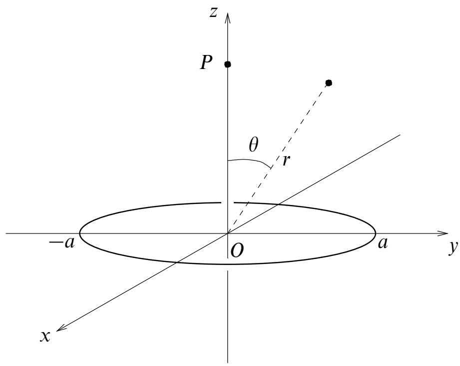


Figure 21.9 The polar axis Oz is taken as normal to the plane of the ring of matter and passing through its centre.

We will illustrate the method for Laplace's equation in spherical polars and first assume that the required solution of $\nabla^2 u = 0$ can be written as a superposition in the normal way:

$$u(r, \theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} (Ar^{\ell} + Br^{-(\ell+1)}) P_{\ell}^{m}(\cos\theta)(C\cos m\phi + D\sin m\phi). \tag{21.58}$$

Here, all the constants A, B, C, D may depend upon ℓ and m, and we have assumed that the required solution is finite on the polar axis. As usual, boundary conditions of a physical nature will then fix or eliminate some of the constants; for example, u finite at the origin implies all $B = 0$, or axial symmetry implies that only $m = 0$ terms are present.

The essence of the method is then to find the remaining constants by determining u at values of r, θ, ϕ for which it can be evaluated *by other means*, e.g. by direct calculation on an axis of symmetry. Once the remaining constants have been fixed by these special considerations to have particular values, the uniqueness theorem can be invoked to establish that they must have these values in general.

►*Calculate the gravitational potential at a general point in space due to a uniform ring of matter of radius a and total mass M.*

Everywhere except on the ring the potential $u(\mathbf{r})$ satisfies the Laplace equation, and so if we use polar coordinates with the normal to the ring as polar axis, as in figure 21.9, a solution of the form (21.58) can be assumed.

We expect the potential $u(r, \theta, \phi)$ to tend to zero as $r \to \infty$, and also to be finite at $r = 0$. At first sight this might seem to imply that all A and B, and hence u, must be identically zero, an unacceptable result. In fact, what it means is that different expressions must apply to different regions of space. On the ring itself we no longer have $\nabla^2 u = 0$ and so it is not surprising that the form of the expression for u changes there. Let us therefore take two separate regions.

In the region $r > a$

(i) we must have $u \to 0$ as $r \to \infty$, implying that all $A = 0$, *and*
(ii) the system is axially symmetric and so only $m = 0$ terms appear.

With these restrictions we can write as a trial form

$$u(r, \theta, \phi) = \sum_{\ell=0}^{\infty} B_\ell r^{-(\ell+1)} P_\ell^0(\cos\theta). \tag{21.59}$$

The constants B_ℓ are still to be determined; this we do by calculating *directly* the potential where this can be done simply – in this case, on the polar axis.

Considering a point P on the polar axis at a distance z ($> a$) from the plane of the ring (taken as $\theta = \pi/2$), all parts of the ring are at a distance $(z^2 + a^2)^{1/2}$ from it. The potential at P is thus straightforwardly

$$u(z, 0, \phi) = -\frac{GM}{(z^2 + a^2)^{1/2}}, \tag{21.60}$$

where G is the gravitational constant. This must be the same as (21.59) for the particular values $r = z$, $\theta = 0$, and ϕ undefined. Since $P_\ell^0(\cos\theta) = P_\ell(\cos\theta)$ with $P_\ell(1) = 1$, putting $r = z$ in (21.59) gives

$$u(z, 0, \phi) = \sum_{\ell=0}^{\infty} \frac{B_\ell}{z^{\ell+1}}. \tag{21.61}$$

However, expanding (21.60) for $z > a$ (as it applies to this region of space) we obtain

$$u(z, 0, \phi) = -\frac{GM}{z}\left[1 - \frac{1}{2}\left(\frac{a}{z}\right)^2 + \frac{3}{8}\left(\frac{a}{z}\right)^4 - \cdots\right],$$

which on comparison with (21.61) gives[§]

$$\begin{aligned} B_0 &= -GM, \\ B_{2\ell} &= -\frac{GMa^{2\ell}(-1)^\ell(2\ell-1)!!}{2^\ell \ell!} \qquad \text{for } \ell \geq 1, \\ B_{2\ell+1} &= 0. \end{aligned} \tag{21.62}$$

We now conclude the argument by saying that if a solution for a general point (r, θ, ϕ) exists at all, which of course we very much expect on physical grounds, then it must be (21.59) with the B_ℓ given by (21.62). This is so because thus defined it is a function with no arbitrary constants and which satisfies all the boundary conditions, and the uniqueness theorem states that there is only one such function. The expression for the potential in the region $r > a$ is therefore

$$u(r, \theta, \phi) = -\frac{GM}{r}\left[1 + \sum_{\ell=1}^{\infty} \frac{(-1)^\ell(2\ell-1)!!}{2^\ell \ell!}\left(\frac{a}{r}\right)^{2\ell} P_{2\ell}(\cos\theta)\right].$$

The expression for $r < a$ can be found in a similar way. The finiteness of u at $r = 0$ and the axial symmetry give

$$u(r, \theta, \phi) = \sum_{\ell=0}^{\infty} A_\ell r^\ell P_\ell^0(\cos\theta).$$

§ $(2\ell - 1)!! = 1 \times 3 \times \cdots \times (2\ell - 1)$.

Comparing this expression for $r = z$, $\theta = 0$ with the $z < a$ expansion of (21.60), which is valid for any z, establishes $A_{2\ell+1} = 0$, $A_0 = -GM/a$ and

$$A_{2\ell} = -\frac{GM}{a^{2\ell+1}} \frac{(-1)^\ell (2\ell-1)!!}{2^\ell \ell!},$$

so that the final expression valid, and convergent, for $r < a$ is thus

$$u(r,\theta,\phi) = -\frac{GM}{a}\left[1 + \sum_{\ell=1}^{\infty} \frac{(-1)^\ell (2\ell-1)!!}{2^\ell \ell!}\left(\frac{r}{a}\right)^{2\ell} P_{2\ell}(\cos\theta)\right].$$

It is easy to check that the solution obtained has the expected physical value for large r and for $r = 0$ and is continuous at $r = a$. ◄

21.3.4 Separation of variables for inhomogeneous equations

So far our discussion of the method of separation of variables has been limited to the solution of homogeneous equations such as the Laplace equation and the wave equation. The solutions of inhomogeneous PDEs are usually obtained using the Green's function methods to be discussed below in section 21.5. However, as a final illustration of the usefulness of the separation of variables, we now consider its application to the solution of inhomogeneous equations.

Because of the added complexity in dealing with inhomogeneous equations, we shall restrict our discussion to the solution of Poisson's equation,

$$\nabla^2 u = \rho(\mathbf{r}), \tag{21.63}$$

in spherical polar coordinates, although the general method can accommodate other coordinate systems and equations. In physical problems the RHS of (21.63) usually contains some multiplicative constant(s). If u is the electrostatic potential in some region of space in which ρ is the density of electric charge then $\nabla^2 u = -\rho(\mathbf{r})/\epsilon_0$. Alternatively, u might represent the gravitational potential in some region where the matter density is given by ρ, so that $\nabla^2 u = 4\pi G\rho(\mathbf{r})$.

We will simplify our discussion by assuming that the required solution u is finite on the polar axis and also that the system possesses axial symmetry about that axis – in which case ρ does not depend on the azimuthal angle ϕ. The key to the method is then to assume a separated form for both the solution u *and* the density term ρ.

From the discussion of Laplace's equation, for systems with axial symmetry only $m = 0$ terms appear, and so the angular part of the solution can be expressed in terms of Legendre polynomials $P_\ell(\cos\theta)$. Since these functions form an orthogonal set let us expand both u and ρ in terms of them:

$$u = \sum_{\ell=0}^{\infty} R_\ell(r) P_\ell(\cos\theta), \tag{21.64}$$

$$\rho = \sum_{\ell=0}^{\infty} F_\ell(r) P_\ell(\cos\theta), \tag{21.65}$$

where the coefficients $R_\ell(r)$ and $F_\ell(r)$ in the Legendre polynomial expansions are functions of r. Since in any particular problem ρ is given, we can find the coefficients $F_\ell(r)$ in the expansion in the usual way (see subsection 18.1.2). It then only remains to find the coefficients $R_\ell(r)$ in the expansion of the solution u.

Writing ∇^2 in spherical polars and substituting (21.64) and (21.65) into (21.63) we obtain

$$\sum_{\ell=0}^{\infty}\left[\frac{P_\ell(\cos\theta)}{r^2}\frac{d}{dr}\left(r^2\frac{dR_\ell}{dr}\right)+\frac{R_\ell}{r^2\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dP_\ell(\cos\theta)}{d\theta}\right)\right]=\sum_{\ell=0}^{\infty}F_\ell(r)P_\ell(\cos\theta). \tag{21.66}$$

However, if, in equation (21.44) of our discussion of the angular part of the solution to Laplace's equation, we set $m=0$ we conclude that

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dP_\ell(\cos\theta)}{d\theta}\right)=-\ell(\ell+1)P_\ell(\cos\theta).$$

Substituting this into (21.66), we find that the LHS is greatly simplified and we obtain

$$\sum_{\ell=0}^{\infty}\left[\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR_\ell}{dr}\right)-\frac{\ell(\ell+1)R_\ell}{r^2}\right]P_\ell(\cos\theta)=\sum_{\ell=0}^{\infty}F_\ell(r)P_\ell(\cos\theta).$$

This relation is most easily satisfied by equating terms on both sides for each value of ℓ separately, so that for $\ell=0,1,2,\ldots$ we have

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR_\ell}{dr}\right)-\frac{\ell(\ell+1)R_\ell}{r^2}=F_\ell(r). \tag{21.67}$$

This is an ODE in which $F_\ell(r)$ is given, and it can therefore be solved for $R_\ell(r)$. The solution to Poisson's equation, u, is then obtained by making the superposition (21.64).

►*In a certain system, the electric charge density ρ is distributed as follows:*

$$\rho=\begin{cases}Ar\cos\theta & \text{for } 0\le r<a,\\ 0 & \text{for } r\ge a.\end{cases}$$

Find the electrostatic potential inside and outside the charge distribution, given that both the potential and its radial derivative are continuous everywhere.

The electrostatic potential u satisfies

$$\nabla^2 u=\begin{cases}-(A/\epsilon_0)r\cos\theta & \text{for } 0\le r<a,\\ 0 & \text{for } r\ge a.\end{cases}$$

For $r<a$ the RHS can be written $-(A/\epsilon_0)rP_1(\cos\theta)$, and the coefficients in (21.65) are simply $F_1(r)=-(Ar/\epsilon_0)$ and $F_\ell(r)=0$ for $\ell\neq 1$. Therefore we need only calculate $R_1(r)$, which satisfies (21.67) for $\ell=1$:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR_1}{dr}\right)-\frac{2R_1}{r^2}=-\frac{Ar}{\epsilon_0}.$$

This can be rearranged to give

$$r^2 R_1'' + 2rR_1' - 2R_1 = -\frac{Ar^3}{\epsilon_0},$$

where the prime denotes differentiation with respect to r. The LHS is homogeneous and the equation can be reduced by the substitution $r = \exp t$, and writing $R_1(r) = S(t)$, to

$$\ddot{S} + \dot{S} - 2S = -\frac{A}{\epsilon_0}\exp 3t, \tag{21.68}$$

where the dots indicate differentiation with respect to t.

This is an inhomogeneous second-order ODE with constant coefficients and can be straightforwardly solved by the methods of subsection 15.2.1 to give

$$S(t) = c_1 \exp t + c_2 \exp(-2t) - \frac{A}{10\epsilon_0}\exp 3t.$$

Recalling that $r = \exp t$ we find

$$R_1(r) = c_1 r + c_2 r^{-2} - \frac{A}{10\epsilon_0} r^3.$$

Since we are interested in the region $r < a$ we must have $c_2 = 0$ for the solution to remain finite. Thus inside the charge distribution the electrostatic potential has the form

$$u_1(r,\theta,\phi) = \left(c_1 r - \frac{A}{10\epsilon_0} r^3\right) P_1(\cos\theta). \tag{21.69}$$

Outside the charge distribution (for $r \geq a$), however, the electrostatic potential obeys Laplace's equation, $\nabla^2 u = 0$, and so given the symmetry of the problem and the requirement that $u \to \infty$ as $r \to \infty$ the solution must take the form

$$u_2(r,\theta,\phi) = \sum_{\ell=0}^{\infty} \frac{B_\ell}{r^{\ell+1}} P_\ell(\cos\theta). \tag{21.70}$$

We can now use the boundary conditions at $r = a$ to fix the constants in (21.69) and (21.70). The requirement of continuity of the potential and its radial derivative at $r = a$ imply that

$$u_1(a,\theta,\phi) = u_2(a,\theta,\phi),$$
$$\frac{\partial u_1}{\partial r}(a,\theta,\phi) = \frac{\partial u_2}{\partial r}(a,\theta,\phi).$$

Clearly $B_\ell = 0$ for $\ell \neq 1$; carrying out the necessary differentiations and setting $r = a$ in (21.69) and (21.70) we obtain the simultaneous equations

$$c_1 a - \frac{A}{10\epsilon_0} a^3 = \frac{B_1}{a^2},$$
$$c_1 - \frac{3A}{10\epsilon_0} a^2 = -\frac{2B_1}{a^3},$$

which may be solved to give $c_1 = Aa^2/(6\epsilon_0)$ and $B_1 = Aa^5/(15\epsilon_0)$. Since $P_1(\cos\theta) = \cos\theta$, the electrostatic potentials inside and outside the charge distribution are given, respectively, by

$$u_1(r,\theta,\phi) = \frac{A}{\epsilon_0}\left(\frac{a^2 r}{6} - \frac{r^3}{10}\right)\cos\theta, \qquad u_2(r,\theta,\phi) = \frac{Aa^5}{15\epsilon_0}\frac{\cos\theta}{r^2}. \blacktriangleleft$$

21.4 Integral transform methods

In the method of separation of variables our aim was to keep the independent variables in a PDE as separate as possible. We now discuss the use of integral transforms in solving PDEs, a method by which one of the independent variables can be eliminated from the differential coefficients. It will be assumed that the reader is familiar with Laplace and Fourier transforms and their properties, as discussed in chapter 13.

The method consists simply of transforming the PDE into one containing derivatives with respect to a smaller number of variables. Thus, if the original equation has just two independent variables, it may be possible to reduce the PDE into a soluble ODE. The solution obtained can then (where possible) be transformed back to give the solution of the original PDE. As we shall see, boundary conditions can usually be incorporated in a natural way.

Which sort of transform to use, and the choice of the variable(s) with respect to which the transform is to be taken, is a matter of experience; we illustrate this in the example below. In practice, transforms can be taken with respect to each variable in turn, and the transformation that affords the greatest simplification can be pursued further.

►*A semi-infinite tube of constant cross-section contains initially pure water. At time $t = 0$, one end of the tube is put into contact with a salt solution and maintained at a concentration u_0. Find the total amount of salt that has diffused into the tube after time t, if the diffusion constant is κ.*

The concentration $u(x, t)$ at time t and distance x from the end of the tube satisfies the diffusion equation

$$\kappa \frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}, \tag{21.71}$$

which has to be solved subject to the boundary conditions $u(0, t) = u_0$ for all t and $u(x, 0) = 0$ for all $x > 0$.

Since we are interested only in $t > 0$, the use of the Laplace transform is suggested. Furthermore, it will be recalled from chapter 13 that one of the major virtues of Laplace transformations is the possibility they afford of replacing derivatives of functions by simple multiplication by a scalar. If the derivative with respect to time were so removed, equation (21.71) would contain only differentiation with respect to a single variable. Let us therefore take the Laplace transform of (21.71) with respect to t:

$$\int_0^\infty \kappa \frac{\partial^2 u}{\partial x^2} \exp(-st)\, dt = \int_0^\infty \frac{\partial u}{\partial t} \exp(-st)\, dt.$$

On the LHS the (double) differentiation is with respect to x, whereas the integration is with respect to the independent variable t. Therefore the derivative can be taken outside the integral. Denoting the Laplace transform of $u(x, t)$ by $\bar{u}(x, s)$ and using result (13.57) to rewrite the transform of the derivative on the RHS (or by integrating directly by parts), we obtain

$$\kappa \frac{\partial^2 \bar{u}}{\partial x^2} = s\bar{u}(x, s) - u(x, 0).$$

But from the boundary condition $u(x, 0) = 0$ the last term on the RHS vanishes, and the solution is immediate:

$$\bar{u}(x,s) = A\exp\left(\sqrt{\frac{s}{\kappa}}\,x\right) + B\exp\left(-\sqrt{\frac{s}{\kappa}}\,x\right),$$

where the constants A and B may depend on s.

We require $u(x,t) \to 0$ as $x \to \infty$ and so we must also have $\bar{u}(\infty, s) = 0$; consequently we require that $A = 0$. The value of B is determined by the need for $u(0,t) = u_0$ and hence that

$$\bar{u}(0,s) = \int_0^\infty u_0 \exp(-st)\,dt = \frac{u_0}{s}.$$

We thus conclude that the appropriate expression for the Laplace transform of $u(x,t)$ is

$$\bar{u}(x,s) = \frac{u_0}{s}\exp\left(-\sqrt{\frac{s}{\kappa}}\,x\right). \tag{21.72}$$

To obtain $u(x,t)$ from this result requires the inversion of this transform – a task that is generally difficult and requires a contour integration. This is discussed in chapter 24, but for completeness we note that the solution is

$$u(x,t) = u_0\left[1 - \operatorname{erf}\left(\frac{x}{\sqrt{4\kappa t}}\right)\right],$$

where $\operatorname{erf}(x)$ is the error function discussed in the Appendix. (The more complete sets of mathematical tables list this inverse Laplace transform.)

In the present problem, however, an alternative method is available. Let $w(t)$ be the amount of salt that has diffused into the tube in time t; then

$$w(t) = \int_0^\infty u(x,t)\,dx,$$

and its transform is given by

$$\begin{aligned}\bar{w}(s) &= \int_0^\infty dt\ \exp(-st)\int_0^\infty u(x,t)\,dx \\ &= \int_0^\infty dx \int_0^\infty u(x,t)\exp(-st)\,dt \\ &= \int_0^\infty \bar{u}(x,s)\,dx.\end{aligned}$$

Substituting for $\bar{u}(x,s)$ from (21.72) into the last integral and integrating, we obtain

$$\bar{w}(s) = u_0\kappa^{1/2}s^{-3/2}.$$

This expression is much simpler to invert, and referring to the table of standard Laplace transforms (table 13.1) we find

$$w(t) = 2(\kappa/\pi)^{1/2}u_0 t^{1/2},$$

which is thus the required expression for the amount of diffused salt at time t. ◄

The above example shows that in some circumstances the use of a Laplace transformation can greatly simplify the solution of a PDE. However, it will have been observed that (as with ODEs) the easy elimination of some derivatives is usually paid for by the introduction of a difficult inverse transformation. This problem, although still present, is less severe for Fourier transformations.

►*An infinite metal bar has an initial temperature distribution $f(x)$ along its length. Find the temperature distribution at a later time t.*

We are interested in values of x from $-\infty$ to ∞, which suggests Fourier transformation with respect to x. Assuming that the solution obeys the boundary conditions $u(x,t) \to 0$ and $\partial u/\partial x \to 0$ as $|x| \to \infty$, we may Fourier-transform the one-dimensional diffusion equation (21.71) to obtain

$$\frac{\kappa}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\partial^2 u(x,t)}{\partial x^2} \exp(-ikx)\, dx = \frac{1}{\sqrt{2\pi}} \frac{\partial}{\partial t} \int_{-\infty}^{\infty} u(x,t) \exp(-ikx)\, dx,$$

where on the RHS we have taken the partial derivative with respect to t outside the integral. Denoting the Fourier transform of $u(x,t)$ by $\widetilde{u}(k,t)$, and using equation (13.28) to rewrite the Fourier transform of the second derivative on the LHS, we then have

$$-\kappa k^2 \widetilde{u}(k,t) = \frac{\partial \widetilde{u}(k,t)}{\partial t}.$$

This first-order equation has the simple solution

$$\widetilde{u}(k,t) = \widetilde{u}(k,0) \exp(-\kappa k^2 t),$$

where the initial conditions give

$$\begin{aligned} \widetilde{u}(k,0) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(x,0) \exp(-ikx)\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \exp(-ikx)\, dx = \widetilde{f}(k). \end{aligned}$$

Thus we may write the Fourier transform of the solution as

$$\widetilde{u}(k,t) = \widetilde{f}(k) \exp(-\kappa k^2 t) = \sqrt{2\pi}\, \widetilde{f}(k) \widetilde{G}(k,t), \qquad (21.73)$$

where we have defined the function $\widetilde{G}(k,t) = (\sqrt{2\pi})^{-1} \exp(-\kappa k^2 t)$. Since $\widetilde{u}(k,t)$ can be written as the product of two Fourier transforms, we can use the convolution theorem, subsection 13.1.7, to write the solution as

$$u(x,t) = \int_{-\infty}^{\infty} G(x-x',t) f(x')\, dx',$$

where $G(x,t)$ is the Green's function for this problem (see subsection 15.2.5). This function is the inverse Fourier transform of $\widetilde{G}(k,t)$ and is thus given by

$$\begin{aligned} G(x,t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\kappa k^2 t) \exp(ikx)\, dk \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left[-\kappa t \left(k^2 - \frac{ix}{\kappa t} k\right)\right] dk. \end{aligned}$$

Completing the square in the integrand we find

$$\begin{aligned} G(x,t) &= \frac{1}{2\pi} \exp\left(-\frac{x^2}{4\kappa t}\right) \int_{-\infty}^{\infty} \exp\left[-\kappa t \left(k - \frac{ix}{2\kappa t}\right)^2\right] dk \\ &= \frac{1}{2\pi} \exp\left(-\frac{x^2}{4\kappa t}\right) \int_{-\infty}^{\infty} \exp\left(-\kappa t k'^2\right) dk' \\ &= \frac{1}{\sqrt{4\pi\kappa t}} \exp\left(-\frac{x^2}{4\kappa t}\right), \end{aligned}$$

where in the second line we have made the substitution $k' = k - ix/(2\kappa t)$, and in the last

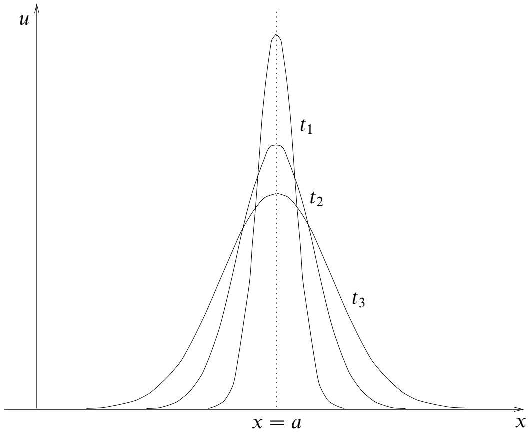


Figure 21.10 Diffusion of heat from a point source in a metal bar: the curves show the temperature u at position x for various times $t_1 < t_2 < t_3$. The area under the curves remains constant, since the total heat energy is conserved.

line we have used the standard result for the integral of a Gaussian, given in subsection 6.4.2. (Strictly speaking the change of variable from k to k' shifts the path of integration off the real axis, since k' is complex for real k, and so results in a complex integral, as will be discussed in chapter 24. Nevertheless, in this case the path of integration can be shifted back to the real axis without affecting the value of the integral.)

Thus the temperature in the bar at a later time t is given by

$$u(x,t) = \frac{1}{\sqrt{4\pi\kappa t}} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-x')^2}{4\kappa t}\right] f(x')\,dx', \qquad (21.74)$$

which may be evaluated (numerically if necessary) when the form of $f(x)$ is given. ◄

As we might expect from our discussion of Green's functions in chapter 15, we see from (21.74) that, if the initial temperature distribution is $f(x) = \delta(x-a)$, i.e. a 'point' source at $x = a$, then the temperature distribution at later times is simply given by

$$u(x,t) = G(x-a,t) = \frac{1}{\sqrt{4\pi\kappa t}} \exp\left[-\frac{(x-a)^2}{4\kappa t}\right].$$

The temperature at several later times is illustrated in figure 21.10, which shows that the heat diffuses out from its initial position; the width of the Gaussian increases as $\sqrt{t}$, a dependence on time which is characteristic of diffusion processes.

The reader may have noticed that in both examples using integral transforms the solutions have been obtained in closed form – albeit in one case in the form of an integral. This differs from the infinite series solutions usually obtained via the separation of variables. It should be noted that this behaviour is a result of the infinite range in x, rather than of the transform method itself. In fact the method of separation of variables would yield the same solutions, since in the infinite-range case the separation constant is not restricted to take on an infinite set of discrete values but may have any real value, with the result that the sum over λ becomes an integral, as mentioned at the end of section 21.2.

►*An infinite metal bar has an initial temperature distribution $f(x)$ along its length. Find the temperature distribution at a later time t using the method of separation of variables.*

This is the same problem as in the previous example, but we now seek a solution by separating variables. From (21.12) a separated solution for the one-dimensional diffusion equation is given by

$$u(x,t) = [A\exp(i\lambda x) + B\exp(-i\lambda x)]\exp(-\kappa\lambda^2 t),$$

where $-\lambda^2$ is the separation constant. Since the bar is infinite we do not require the solution to take a given form at any finite value of x (for instance at $x = 0$) and so there is no restriction on λ other than its being real. Therefore instead of the superposition of such solutions in the form of a sum over allowed values of λ we have an integral over all λ,

$$u(x,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} A(\lambda)\exp(-\kappa\lambda^2 t)\exp(i\lambda x)\,d\lambda, \qquad (21.75)$$

where in taking λ from $-\infty$ to ∞ we need include only one of the complex exponentials; we have taken a factor $1/\sqrt{2\pi}$ out of $A(\lambda)$ for convenience. We can see from (21.75) that the expression for $u(x,t)$ has the form of an inverse Fourier transform (where λ is the transform variable). Therefore, Fourier-transforming both sides and using the Fourier inversion theorem, we find

$$\widetilde{u}(\lambda,t) = A(\lambda)\exp(-\kappa\lambda^2 t).$$

Now, the initial boundary condition requires

$$u(x,0) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} A(\lambda)\exp(i\lambda x)\,d\lambda = f(x),$$

from which, using the Fourier inversion theorem once more, we see that $A(\lambda) = \widetilde{f}(\lambda)$. Therefore we have

$$\widetilde{u}(\lambda,t) = \widetilde{f}(\lambda)\exp(-\kappa\lambda^2 t),$$

which is identical to (21.73) in the previous example (but with k replaced by λ), and hence leads to the same result. ◄

21.5 Inhomogeneous problems – Green's functions

In chapters 15 and 17 we encountered Green's functions and found them a useful tool for solving inhomogeneous linear ODEs. We now discuss their usefulness in solving inhomogeneous linear PDEs.

For the sake of brevity we shall again denote a linear PDE by

$$\mathcal{L}u(\mathbf{r}) = \rho(\mathbf{r}), \qquad (21.76)$$

where $\mathcal{L}$ is a linear partial differential operator. For example, in Laplace's equation we have $\mathcal{L} = \nabla^2$, whereas for Helmholtz's equation $\mathcal{L} = \nabla^2 + k^2$. Note that we have not specified the dimensionality of the problem, and (21.76) may, for example, represent Poisson's equation in two or three (or more) dimensions. The reader will also notice that for the sake of simplicity we have not included any time dependence in (21.76). Nevertheless, the following discussion can be generalised to include it.

As we discussed in subsection 20.3.2, a problem is inhomogeneous if the fact that $u(\mathbf{r})$ is a solution does *not* imply that any constant multiple $\lambda u(\mathbf{r})$ is also a solution. This inhomogeneity may derive from either the PDE itself or from the boundary conditions imposed on the solution.

In our discussion of Green's function solutions of inhomogeneous ODEs (see subsection 15.2.5) we dealt with inhomogeneous boundary conditions by making a suitable change of variable such that in the new variable the boundary conditions were homogeneous. In an analogous way, as illustrated in the final example of section 21.2, it is usually possible to make a change of variables in PDEs to transform between inhomogeneity of the boundary conditions and inhomogeneity of the equation. Therefore let us assume for the moment that the boundary conditions imposed on the solution $u(\mathbf{r})$ of (21.76) are homogeneous. This most commonly means that if we seek a solution to (21.76) in some region V then on the surface S that bounds V the solution obeys the conditions $u(\mathbf{r}) = 0$ or $\partial u/\partial n = 0$, where $\partial u/\partial n$ is the normal derivative of u at the surface S.

We shall discuss the extension of the Green's function method to the direct solution of problems with inhomogeneous boundary conditions in subsection 21.5.2, but we first highlight how the Green's function approach to solving ODEs can be simply extended to PDEs for homogeneous boundary conditions.

21.5.1 Similarities to Green's functions for ODEs

As in the discussion of ODEs in chapter 15, we may consider the Green's function for a system described by a PDE as the response of the system to a 'unit impulse' or 'point source'. Thus if we seek a solution to (21.76) that satisfies some homogeneous boundary conditions on $u(\mathbf{r})$ then the Green's function $G(\mathbf{r}, \mathbf{r}_0)$ for the problem is a solution of

$$\mathcal{L}G(\mathbf{r}, \mathbf{r}_0) = \delta(\mathbf{r} - \mathbf{r}_0), \tag{21.77}$$

where $\mathbf{r}_0$ lies in V. The Green's function $G(\mathbf{r}, \mathbf{r}_0)$ must also satisfy the imposed (homogeneous) boundary conditions.

It is understood that in (21.77) the $\mathcal{L}$ operator expresses differentiation with respect to $\mathbf{r}$ as opposed to $\mathbf{r}_0$. Also, $\delta(\mathbf{r} - \mathbf{r}_0)$ is the Dirac delta function (see chapter 13) of dimension appropriate to the problem; it may be thought of as representing a unit-strength point source at $\mathbf{r} = \mathbf{r}_0$.

Following an analogous argument to that given in subsection 15.2.5 for ODEs, if the boundary conditions on $u(\mathbf{r})$ are homogeneous then a solution to (21.76) that satisfies the imposed boundary conditions is given by

$$u(\mathbf{r}) = \int G(\mathbf{r}, \mathbf{r}_0)\rho(\mathbf{r}_0)\, dV(\mathbf{r}_0), \tag{21.78}$$

where the integral on $\mathbf{r}_0$ is over some appropriate 'volume'. In two or more dimensions, however, the task of finding directly a solution to (21.77) that satisfies the imposed boundary conditions on S can be a difficult one, and we return to this in the next subsection.

An alternative approach is to follow a similar argument to that presented in chapter 17 for ODEs and so to construct the Green's function for (21.76) as a superposition of eigenfunctions of the operator $\mathcal{L}$, provided $\mathcal{L}$ is Hermitian. By analogy with an ordinary differential operator, a partial differential operator is Hermitian if it satisfies

$$\int_V v^*(\mathbf{r})\mathcal{L}w(\mathbf{r})\, dV = \left[\int_V w^*(\mathbf{r})\mathcal{L}v(\mathbf{r})\, dV\right]^*,$$

where the asterisk denotes complex conjugation and v and w are arbitrary functions obeying the imposed (homogeneous) boundary condition on the solution of $\mathcal{L}u(\mathbf{r}) = 0$.

The eigenfunctions $u_n(\mathbf{r})$, $n = 0, 1, 2, \ldots$, of $\mathcal{L}$ satisfy

$$\mathcal{L}u_n(\mathbf{r}) = \lambda_n u_n(\mathbf{r}),$$

where λ_n are the corresponding eigenvalues, which are all real for an Hermitian operator $\mathcal{L}$. Furthermore, each eigenfunction must obey any imposed (homogeneous) boundary conditions. Using an argument analogous to that given in chapter 17, the Green's function for the problem is given by

$$G(\mathbf{r}, \mathbf{r}_0) = \sum_{n=0}^{\infty} \frac{u_n(\mathbf{r})u_n^*(\mathbf{r}_0)}{\lambda_n}. \tag{21.79}$$

From (21.79) we see immediately that the Green's function (irrespective of how it is found) enjoys the property

$$G(\mathbf{r}, \mathbf{r}_0) = G^*(\mathbf{r}_0, \mathbf{r}).$$

Thus, if the Green's function is real then it is symmetric in its two arguments.

Once the Green's function has been obtained, the solution to (21.76) is again given by (21.78). For PDEs this approach can become very cumbersome, however, and so we shall not pursue it further here.

21.5.2 General boundary-value problems

As mentioned above, often inhomogeneous boundary conditions can be dealt with by making an appropriate change of variables, such that the boundary conditions in the new variables are homogeneous although the equation itself is generally inhomogeneous. In this section, however, we extend the use of Green's functions to problems with inhomogeneous boundary conditions (and equations). This provides a more consistent and intuitive approach to the solution of such *boundary-value problems.*

For definiteness we shall consider Poisson's equation

$$\nabla^2 u(\mathbf{r}) = \rho(\mathbf{r}), \tag{21.80}$$

but the material of this section may be extended to other linear PDEs of the form (21.76). Clearly, Poisson's equation reduces to Laplace's equation for $\rho(\mathbf{r}) = 0$ and so our discussion is equally applicable to this case.

We wish to solve (21.80) in some region V bounded by a surface S, which may consist of several disconnected parts. As stated above, we shall allow the possibility that the boundary conditions on the solution $u(\mathbf{r})$ may be inhomogeneous on S, although as we shall see this method reduces to those discussed above in the special case that the boundary conditions are in fact homogeneous.

The two common types of inhomogeneous boundary condition for Poisson's equation are (as discussed in subsection 20.6.2):

(i) Dirichlet conditions, in which $u(\mathbf{r})$ is specified on S, and
(ii) Neumann conditions, in which $\partial u/\partial n$ is specified on S.

In general, specifying *both* Dirichlet *and* Neumann conditions on S overdetermines the problem and leads to there being no solution.

The specification of the surface S requires some further comment, since S may have several disconnected parts. If we wish to solve Poisson's equation inside some closed surface S then the situation is straightforward and is shown in figure 21.11(a). If, however, we wish to solve Poisson's equation in the gap between two closed surfaces (for example in the gap between two concentric conducting cylinders) then the volume V is bounded by a surface S that has two disconnected parts S_1 and S_2, as shown in figure 21.11(b); the direction of the normal to the surface is always taken as pointing *out* of the volume V. A similar situation arises when we wish to solve Poisson's equation *outside* some closed surface S_1. In this case the volume V is infinite but is treated formally by taking the surface S_2 as a large sphere of radius R and letting R tend to infinity.

In order to solve (21.80) subject to either Dirichlet or Neumann boundary conditions on S, we will remind ourselves of Green's second theorem, equation (11.20), which states that, for two scalar functions $\phi(\mathbf{r})$ and $\psi(\mathbf{r})$ defined in some volume V bounded by a surface S,

$$\int_V (\phi\nabla^2\psi - \psi\nabla^2\phi)\, dV = \int_S (\phi\nabla\psi - \psi\nabla\phi)\cdot\hat{\mathbf{n}}\, dS, \tag{21.81}$$

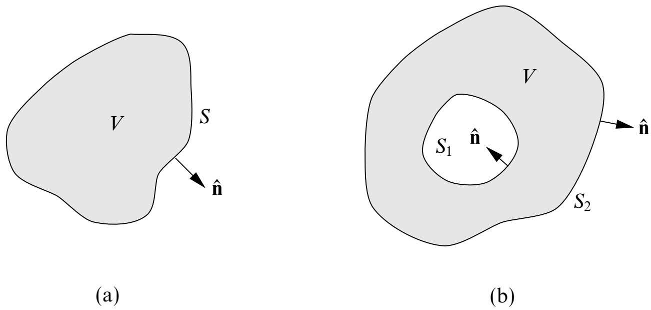


Figure 21.11 Surfaces used for solving Poisson's equation in different regions V.

where on the RHS it is common to write, for example, $\nabla\psi \cdot \hat{\mathbf{n}}\, dS$ as $(\partial\psi/\partial n)\, dS$. The expression $\partial\psi/\partial n$ stands for $\nabla\psi \cdot \hat{\mathbf{n}}$, the rate of change of ψ in the direction of the unit outward normal $\hat{\mathbf{n}}$ to the surface S.

The Green's function for Poisson's equation (21.80) must satisfy

$$\nabla^2 G(\mathbf{r}, \mathbf{r}_0) = \delta(\mathbf{r} - \mathbf{r}_0), \tag{21.82}$$

where $\mathbf{r}_0$ lies in V. (As mentioned above, we may think of $G(\mathbf{r}, \mathbf{r}_0)$ as the solution to Poisson's equation for a unit-strength point source located at $\mathbf{r} = \mathbf{r}_0$.) Let us for the moment impose no boundary conditions on $G(\mathbf{r}, \mathbf{r}_0)$.

If we now let $\phi = u(\mathbf{r})$ and $\psi = G(\mathbf{r}, \mathbf{r}_0)$ in Green's theorem (21.81) then we obtain

$$\int_V \left[u(\mathbf{r})\nabla^2 G(\mathbf{r}, \mathbf{r}_0) - G(\mathbf{r}, \mathbf{r}_0)\, \nabla^2 u(\mathbf{r})\right] dV(\mathbf{r}) = \int_S \left[u(\mathbf{r})\frac{\partial G(\mathbf{r}, \mathbf{r}_0)}{\partial n} - G(\mathbf{r}, \mathbf{r}_0)\frac{\partial u(\mathbf{r})}{\partial n}\right] dS(\mathbf{r}),$$

where we have made explicit that the volume and surface integrals are with respect to $\mathbf{r}$. Using (21.80) and (21.82) the LHS can be simplified to give

$$\int_V [u(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}_0) - G(\mathbf{r}, \mathbf{r}_0)\rho(\mathbf{r})]\, dV(\mathbf{r}) = \int_S \left[u(\mathbf{r})\frac{\partial G(\mathbf{r}, \mathbf{r}_0)}{\partial n} - G(\mathbf{r}, \mathbf{r}_0)\frac{\partial u(\mathbf{r})}{\partial n}\right] dS(\mathbf{r}). \tag{21.83}$$

Since $\mathbf{r}_0$ lies within the volume V,

$$\int_V u(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}_0)\, dV(\mathbf{r}) = u(\mathbf{r}_0),$$

and thus on rearranging (21.83) the solution to Poisson's equation (21.80) can be written as

$$u(\mathbf{r}_0) = \int_V G(\mathbf{r},\mathbf{r}_0)\rho(\mathbf{r})\,dV(\mathbf{r}) + \int_S \left[u(\mathbf{r})\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial n} - G(\mathbf{r},\mathbf{r}_0)\frac{\partial u(\mathbf{r})}{\partial n} \right] dS(\mathbf{r}). \tag{21.84}$$

Clearly, we can interchange the roles of $\mathbf{r}$ and $\mathbf{r}_0$ in (21.84) if we wish. (Remember also that, for a real Green's function, $G(\mathbf{r},\mathbf{r}_0) = G(\mathbf{r}_0,\mathbf{r})$.)

Equation (21.84) is *central* to the extension of the Green's function method to problems with inhomogeneous boundary conditions, and we next discuss its application to both Dirichlet and Neumann boundary-value problems. But, before doing so, we also note that if the boundary condition on S is in fact homogeneous, so that $u(\mathbf{r}) = 0$ or $\partial u(\mathbf{r})/\partial n = 0$ on S, then demanding that the Green's function $G(\mathbf{r},\mathbf{r}_0)$ also obeys the same boundary condition causes the surface integral in (21.84) to vanish, and we are left with the familiar form of solution given in (21.78). The extension of (21.84) to a PDE other than Poisson's equation is discussed in exercise 21.28.

21.5.3 Dirichlet problems

In a Dirichlet problem we require the solution $u(\mathbf{r})$ of Poisson's equation (21.80) to take specific values on some surface S that bounds V, i.e. we require that $u(\mathbf{r}) = f(\mathbf{r})$ on S where f is a given function.

If we seek a Green's function $G(\mathbf{r},\mathbf{r}_0)$ for this problem it must clearly satisfy (21.82), but we are free to choose the boundary conditions satisfied by $G(\mathbf{r},\mathbf{r}_0)$ in such a way as to make the solution (21.84) as simple as possible. From (21.84), we see that by choosing

$$G(\mathbf{r},\mathbf{r}_0) = 0 \qquad \text{for } \mathbf{r} \text{ on } S \tag{21.85}$$

the second term in the surface integral vanishes. Since $u(\mathbf{r}) = f(\mathbf{r})$ on S, (21.84) then becomes

$$u(\mathbf{r}_0) = \int_V G(\mathbf{r},\mathbf{r}_0)\rho(\mathbf{r})\,dV(\mathbf{r}) + \int_S f(\mathbf{r})\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial n}\,dS(\mathbf{r}). \tag{21.86}$$

Thus we wish to find the *Dirichlet Green's function* that

(i) satisfies (21.82) and hence is singular at $\mathbf{r} = \mathbf{r}_0$, and
(ii) obeys the boundary condition $G(\mathbf{r},\mathbf{r}_0) = 0$ for $\mathbf{r}$ on S.

In general, it is difficult to obtain this function directly, and so it is useful to separate these two requirements. We therefore look for a solution of the form

$$G(\mathbf{r},\mathbf{r}_0) = F(\mathbf{r},\mathbf{r}_0) + H(\mathbf{r},\mathbf{r}_0),$$

where $F(\mathbf{r},\mathbf{r}_0)$ satisfies (21.82) and has the required singular character at $\mathbf{r} = \mathbf{r}_0$ but does not necessarily obey the boundary condition on S, whilst $H(\mathbf{r},\mathbf{r}_0)$ satisfies the corresponding homogeneous equation (i.e. Laplace's equation) inside V but is adjusted in such a way that the sum $G(\mathbf{r},\mathbf{r}_0)$ equals zero on S. The Green's function $G(\mathbf{r},\mathbf{r}_0)$ is still a solution of (21.82) since

$$\nabla^2 G(\mathbf{r},\mathbf{r}_0) = \nabla^2 F(\mathbf{r},\mathbf{r}_0) + \nabla^2 H(\mathbf{r},\mathbf{r}_0) = \nabla^2 F(\mathbf{r},\mathbf{r}_0) + 0 = \delta(\mathbf{r}-\mathbf{r}_0).$$

The function $F(\mathbf{r},\mathbf{r}_0)$ is called the *fundamental solution* and will clearly take different forms depending on the dimensionality of the problem. Let us first consider the fundamental solution to (21.82) in three dimensions.

►*Find the fundamental solution to Poisson's equation in three dimensions that tends to zero as* $|\mathbf{r}| \to \infty$.

We wish to solve

$$\nabla^2 F(\mathbf{r},\mathbf{r}_0) = \delta(\mathbf{r}-\mathbf{r}_0) \qquad (21.87)$$

in three dimensions, subject to the boundary condition $F(\mathbf{r},\mathbf{r}_0) \to 0$ as $|\mathbf{r}| \to \infty$. Since the problem is spherically symmetric about $\mathbf{r}_0$, let us consider a large sphere S of radius R centred on $\mathbf{r}_0$, and integrate (21.87) over the enclosed volume V. We then obtain

$$\int_V \nabla^2 F(\mathbf{r},\mathbf{r}_0)\,dV = \int_V \delta(\mathbf{r}-\mathbf{r}_0)\,dV = 1, \qquad (21.88)$$

since V encloses the point $\mathbf{r}_0$. However, using the divergence theorem,

$$\int_V \nabla^2 F(\mathbf{r},\mathbf{r}_0)\,dV = \int_S \nabla F(\mathbf{r},\mathbf{r}_0)\cdot\hat{\mathbf{n}}\,dS, \qquad (21.89)$$

where $\hat{\mathbf{n}}$ is the unit normal to the large sphere S at any point.

Since the problem is spherically symmetric about $\mathbf{r}_0$, we expect that

$$F(\mathbf{r},\mathbf{r}_0) = F(|\mathbf{r}-\mathbf{r}_0|) = F(r),$$

i.e. that F has the same value everywhere on S. Thus, evaluating the surface integral in (21.89) and equating it to unity from (21.88), we have[§]

$$4\pi r^2 \frac{dF}{dr}\bigg|_{r=R} = 1.$$

Integrating this expression we obtain

$$F(r) = -\frac{1}{4\pi r} + \text{constant},$$

but, since we require $F(\mathbf{r},\mathbf{r}_0) \to 0$ as $|\mathbf{r}| \to \infty$, the constant must be zero. The fundamental solution in three dimensions is consequently given by

$$F(\mathbf{r},\mathbf{r}_0) = -\frac{1}{4\pi|\mathbf{r}-\mathbf{r}_0|}. \qquad (21.90)$$

This is clearly also the full Green's function for Poisson's equation subject to the boundary condition $u(\mathbf{r}) \to 0$ as $|\mathbf{r}| \to \infty$. ◄

Using (21.90) we can write down the solution of Poisson's equation to find,

§ A vertical bar to the right of an expression is a common alternative to enclosing the expression in square brackets; as usual, the subscript shows the value of the variable at which the expression is to be evaluated.

for example, the electrostatic potential $u(\mathbf{r})$ due to some distribution of electric charge $\rho(\mathbf{r})$. The electrostatic potential satisfies

$$\nabla^2 u(\mathbf{r}) = -\frac{\rho}{\epsilon_0},$$

where $u(\mathbf{r}) \to 0$ as $|\mathbf{r}| \to \infty$. Since the boundary condition on the surface at infinity is homogeneous the surface integral in (21.86) vanishes, and using (21.90) we recover the familiar solution

$$u(\mathbf{r}_0) = \int \frac{\rho(\mathbf{r})}{4\pi\epsilon_0|\mathbf{r} - \mathbf{r}_0|}\, dV(\mathbf{r}), \tag{21.91}$$

where the volume integral is over all space.

We can develop an analogous theory in two dimensions. As before the fundamental solution satisfies

$$\nabla^2 F(\mathbf{r}, \mathbf{r}_0) = \delta(\mathbf{r} - \mathbf{r}_0), \tag{21.92}$$

where $\delta(\mathbf{r}-\mathbf{r}_0)$ is now the two-dimensional delta function. Following an analogous method to that used in the previous example, we find the fundamental solution in two dimensions to be given by

$$F(\mathbf{r}, \mathbf{r}_0) = \frac{1}{2\pi} \ln|\mathbf{r} - \mathbf{r}_0| + \text{constant}. \tag{21.93}$$

From the form of the solution we see that in two dimensions we cannot apply the condition $F(\mathbf{r}, \mathbf{r}_0) \to 0$ as $|\mathbf{r}| \to \infty$, and in this case the constant does not necessarily vanish.

We now return to the task of constructing the full Dirichlet Green's function. To do so we wish to add to the fundamental solution a solution of the homogeneous equation (in this case Laplace's equation) such that $G(\mathbf{r}, \mathbf{r}_0) = 0$ on S, as required by (21.86) and its attendant conditions. The appropriate Green's function is constructed by adding to the fundamental solution 'copies' of itself that represent 'image' sources at different locations *outside* V. Hence this approach is called the *method of images.*

In summary, if we wish to solve Poisson's equation in some region V subject to Dirichlet boundary conditions on its surface S then the procedure and argument are as follows.

(i) To the single source $\delta(\mathbf{r} - \mathbf{r}_0)$ inside V add image sources *outside* V

$$\sum_{n=1}^{N} q_n \delta(\mathbf{r} - \mathbf{r}_n) \qquad \text{with } \mathbf{r}_n \text{ outside } V,$$

where the positions $\mathbf{r}_n$ and the strengths q_n of the image sources are to be determined as described in step (iii) below.

(ii) Since all the image sources lie outside V, the fundamental solution corresponding to each source satisfies Laplace's equation *inside* V. Thus we may add the fundamental solutions $F(\mathbf{r}, \mathbf{r}_n)$ corresponding to each image source to that corresponding to the single source inside V, obtaining the Green's function

$$G(\mathbf{r}, \mathbf{r}_0) = F(\mathbf{r}, \mathbf{r}_0) + \sum_{n=1}^{N} q_n F(\mathbf{r}, \mathbf{r}_n).$$

(iii) Now adjust the positions $\mathbf{r}_n$ and strengths q_n of the image sources so that the required boundary conditions are satisfied on S. For a Dirichlet Green's function we require $G(\mathbf{r}, \mathbf{r}_0) = 0$ for $\mathbf{r}$ on S.

(iv) The solution to Poisson's equation subject to the Dirichlet boundary condition $u(\mathbf{r}) = f(\mathbf{r})$ on S is then given by (21.86).

In general it is very difficult to find the correct positions and strengths for the images, i.e. to make them such that the boundary conditions on S are satisfied. Nevertheless, it is possible to do so for certain problems that have simple geometry. In particular, for problems in which the boundary S consists of straight lines (in two dimensions) or planes (in three dimensions), positions of the image points can be deduced simply by imagining the boundary lines or planes to be mirrors in which the single source in V (at $\mathbf{r}_0$) is reflected.

►*Solve Laplace's equation $\nabla^2 u = 0$ in three dimensions in the half-space $z > 0$, given that $u(\mathbf{r}) = f(\mathbf{r})$ on the plane $z = 0$.*

The surface S bounding V consists of the xy-plane and the surface at infinity. Therefore, the Dirichlet Green's function for this problem must satisfy $G(\mathbf{r}, \mathbf{r}_0) = 0$ on $z = 0$ and $G(\mathbf{r}, \mathbf{r}_0) \to 0$ as $|\mathbf{r}| \to \infty$. Thus it is clear in this case that we require one image source at a position $\mathbf{r}_1$ that is the reflection of $\mathbf{r}_0$ in the plane $z = 0$, as shown in figure 21.12 (so that $\mathbf{r}_1$ lies in $z < 0$, outside the region in which we wish to obtain a solution). It is also clear that the strength of this image should be -1.

Therefore by adding the fundamental solutions corresponding to the original source and its image we obtain the Green's function

$$G(\mathbf{r}, \mathbf{r}_0) = -\frac{1}{4\pi|\mathbf{r} - \mathbf{r}_0|} + \frac{1}{4\pi|\mathbf{r} - \mathbf{r}_1|}, \tag{21.94}$$

where $\mathbf{r}_1$ is the reflection of $\mathbf{r}_0$ in the plane $z = 0$, i.e. if $\mathbf{r}_0 = (x_0, y_0, z_0)$ then $\mathbf{r}_1 = (x_0, y_0, -z_0)$. Clearly $G(\mathbf{r}, \mathbf{r}_0) \to 0$ as $|\mathbf{r}| \to \infty$ as required. Also $G(\mathbf{r}, \mathbf{r}_0) = 0$ on $z = 0$, and so (21.94) is the desired Dirichlet Green's function.

The solution to Laplace's equation is then given by (21.86) with $\rho(\mathbf{r}) = 0$,

$$u(\mathbf{r}_0) = \int_S f(\mathbf{r}) \frac{\partial G(\mathbf{r}, \mathbf{r}_0)}{\partial n}\, dS(\mathbf{r}). \tag{21.95}$$

Clearly the surface at infinity makes no contribution to this integral. The outward-pointing unit vector normal to the xy-plane is simply $\hat{\mathbf{n}} = -\mathbf{k}$ (where $\mathbf{k}$ is the unit vector in the z-direction), and so

$$\frac{\partial G(\mathbf{r}, \mathbf{r}_0)}{\partial n} = -\frac{\partial G(\mathbf{r}, \mathbf{r}_0)}{\partial z} = -\mathbf{k} \cdot \nabla G(\mathbf{r}, \mathbf{r}_0).$$

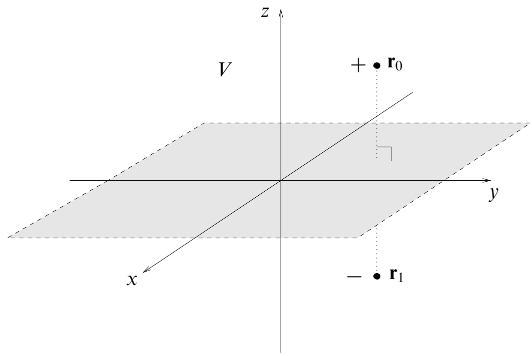


Figure 21.12 The arrangement of images for solving Laplace's equation in the half-space $z > 0$.

We may evaluate this normal derivative by writing the Green's function (21.94) explicitly in terms of x, y and z (and x_0, y_0 and z_0) and calculating the partial derivative with respect to z directly. It is usually quicker, however, to use the fact that[§]

$$\nabla|\mathbf{r}-\mathbf{r}_0| = \frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|}; \tag{21.96}$$

thus

$$\nabla G(\mathbf{r},\mathbf{r}_0) = \frac{\mathbf{r}-\mathbf{r}_0}{4\pi|\mathbf{r}-\mathbf{r}_0|^3} - \frac{\mathbf{r}-\mathbf{r}_1}{4\pi|\mathbf{r}-\mathbf{r}_1|^3}.$$

Since $\mathbf{r}_0 = (x_0, y_0, z_0)$ and $\mathbf{r}_1 = (x_0, y_0, -z_0)$ the normal derivative is given by

$$\begin{aligned} -\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial z} &= -\mathbf{k}\cdot\nabla G(\mathbf{r},\mathbf{r}_0) \\ &= -\frac{z-z_0}{4\pi|\mathbf{r}-\mathbf{r}_0|^3} + \frac{z+z_0}{4\pi|\mathbf{r}-\mathbf{r}_1|^3}. \end{aligned}$$

Therefore on the surface $z = 0$, and writing out the dependence on x, y and z explicitly, we have

$$-\left.\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial z}\right|_{z=0} = \frac{2z_0}{4\pi[(x-x_0)^2 + (y-y_0)^2 + z_0^2]^{3/2}}.$$

Inserting this expression into (21.95) we obtain the solution

$$u(x_0, y_0, z_0) = \frac{z_0}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f(x,y)}{[(x-x_0)^2 + (y-y_0)^2 + z_0^2]^{3/2}}\,dx\,dy. \blacktriangleleft$$

An analogous procedure may be applied in two-dimensional problems. For

[§] Since $|\mathbf{r}-\mathbf{r}_0|^2 = (\mathbf{r}-\mathbf{r}_0)\cdot(\mathbf{r}-\mathbf{r}_0)$ we have $\nabla|\mathbf{r}-\mathbf{r}_0|^2 = 2(\mathbf{r}-\mathbf{r}_0)$, from which we obtain

$$\nabla(|\mathbf{r}-\mathbf{r}_0|^2)^{1/2} = \frac{1}{2}\frac{2(\mathbf{r}-\mathbf{r}_0)}{(|\mathbf{r}-\mathbf{r}_0|^2)^{1/2}} = \frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|}.$$

Note that this result holds in two *and* three dimensions.

example, in solving Poisson's equation in two dimensions in the half-space $x > 0$ we again require just one image charge, of strength $q_1 = -1$, at a position $\mathbf{r}_1$ that is the reflection of $\mathbf{r}_0$ in the line $x = 0$. Since we require $G(\mathbf{r}, \mathbf{r}_0) = 0$ when $\mathbf{r}$ lies on $x = 0$, the constant in (21.93) must equal zero, and so the Dirichlet Green's function is

$$G(\mathbf{r}, \mathbf{r}_0) = \frac{1}{2\pi}\left(\ln|\mathbf{r} - \mathbf{r}_0| - \ln|\mathbf{r} - \mathbf{r}_1|\right).$$

Clearly $G(\mathbf{r}, \mathbf{r}_0)$ tends to zero as $|\mathbf{r}| \to \infty$. If, however, we wish to solve the two-dimensional Poisson equation in the quarter space $x > 0$, $y > 0$, then more image points are required.

►*A line charge in the z-direction of charge density λ is placed at some position $\mathbf{r}_0$ in the quarter-space $x > 0$, $y > 0$. Calculate the force per unit length on the line charge due to the presence of thin earthed plates along $x = 0$ and $y = 0$.*

Here we wish to solve Poisson's equation,

$$\nabla^2 u = -\frac{\lambda}{\epsilon_0}\delta(\mathbf{r} - \mathbf{r}_0),$$

in the quarter space $x > 0$, $y > 0$. It is clear that we require three image line charges with positions and strengths as shown in figure 21.13 (all of which lie outside the region in which we seek a solution). The boundary condition that the electrostatic potential u is zero on $x = 0$ and $y = 0$ (shown as the 'curve' C in figure 21.13) is then automatically satisfied, and so this system of image charges is directly equivalent to the original situation of a single line charge in the presence of the earthed plates along $x = 0$ and $y = 0$. Thus the electrostatic potential is simply equal to the Dirichlet Green's function

$$u(\mathbf{r}) = G(\mathbf{r}, \mathbf{r}_0) = -\frac{\lambda}{2\pi\epsilon_0}\left(\ln|\mathbf{r} - \mathbf{r}_0| - \ln|\mathbf{r} - \mathbf{r}_1| + \ln|\mathbf{r} - \mathbf{r}_2| - \ln|\mathbf{r} - \mathbf{r}_3|\right),$$

which equals zero on C and on the 'surface' at infinity.

The force on the line charge at $\mathbf{r}_0$, therefore, is simply that due to the three line charges at $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$. The elecrostatic potential due to a line charge at $\mathbf{r}_i$, $i = 1$, 2 or 3, is given by the fundamental solution

$$u_i(\mathbf{r}) = \mp\frac{\lambda}{2\pi\epsilon_0}\ln|\mathbf{r} - \mathbf{r}_i| + c,$$

the upper or lower sign being taken according to whether the line charge is positive or negative, respectively. Therefore the force per unit length on the line charge at $\mathbf{r}_0$, due to the one at $\mathbf{r}_i$, is given by

$$-\lambda\nabla u_i(\mathbf{r})\Big|_{\mathbf{r}=\mathbf{r}_0} = \pm\frac{\lambda^2}{2\pi\epsilon_0}\frac{\mathbf{r}_0 - \mathbf{r}_i}{|\mathbf{r}_0 - \mathbf{r}_i|^2}.$$

Adding the contributions from the three image charges shown in figure 21.13, the total force experienced by the line charge at $\mathbf{r}_0$ is given by

$$\mathbf{F} = \frac{\lambda^2}{2\pi\epsilon_0}\left(-\frac{\mathbf{r}_0 - \mathbf{r}_1}{|\mathbf{r}_0 - \mathbf{r}_1|^2} + \frac{\mathbf{r}_0 - \mathbf{r}_2}{|\mathbf{r}_0 - \mathbf{r}_2|^2} - \frac{\mathbf{r}_0 - \mathbf{r}_3}{|\mathbf{r}_0 - \mathbf{r}_3|^2}\right),$$

where, from the figure, $\mathbf{r}_0 - \mathbf{r}_1 = 2y_0\mathbf{j}$, $\mathbf{r}_0 - \mathbf{r}_2 = 2x_0\mathbf{i} + 2y_0\mathbf{j}$ and $\mathbf{r}_0 - \mathbf{r}_3 = 2x_0\mathbf{i}$. Thus, in

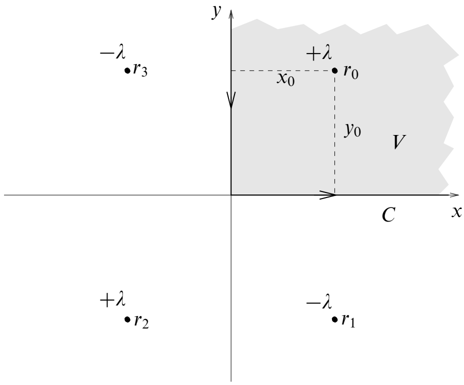


Figure 21.13 The arrangement of images for finding the force on a line charge situated in the (two-dimensional) quarter-space $x > 0$, $y > 0$, when the planes $x = 0$ and $y = 0$ are earthed.

terms of x_0 and y_0, the total force on the line charge due to the charge induced on the plates is given by

$$\mathbf{F} = \frac{\lambda^2}{2\pi\epsilon_0}\left(-\frac{1}{2y_0}\mathbf{j} + \frac{2x_0\,\mathbf{i} + 2y_0\,\mathbf{j}}{4x_0^2 + 4y_0^2} - \frac{1}{2x_0}\mathbf{i}\right)$$
$$= -\frac{\lambda^2}{4\pi\epsilon_0(x_0^2 + y_0^2)}\left(\frac{y_0^2}{x_0}\mathbf{i} + \frac{x_0^2}{y_0}\mathbf{j}\right). \blacktriangleleft$$

Further generalisations are possible. For instance, solving Poisson's equation in the two-dimensional strip $-\infty < x < \infty$, $0 < y < b$ requires an infinite series of image points.

So far we have considered problems in which the boundary S consists of straight lines (in two dimensions) or planes (in three dimensions), in which simple reflections of the source at $\mathbf{r}_0$ in these boundaries fix the positions of the image points. For more complicated (curved) boundaries this is no longer possible, and finding the appropriate position(s) and strength(s) of the image source(s) requires further work.

►*Use the method of images to find the Dirichlet Green's function for solving Poisson's equation outside a sphere of radius a centred at the origin.*

We need to find a solution of Poisson's equation valid outside the sphere of radius a. Since an image point $\mathbf{r}_1$ cannot lie in this region, it must be located within the sphere. The Green's function for this problem is therefore

$$G(\mathbf{r}, \mathbf{r}_0) = -\frac{1}{4\pi|\mathbf{r} - \mathbf{r}_0|} - \frac{q}{4\pi|\mathbf{r} - \mathbf{r}_1|},$$

where $|\mathbf{r}_0| > a$, $|\mathbf{r}_1| < a$ and q is the strength of the image which we have yet to determine. Clearly, $G(\mathbf{r}, \mathbf{r}_0) \to 0$ on the surface at infinity.

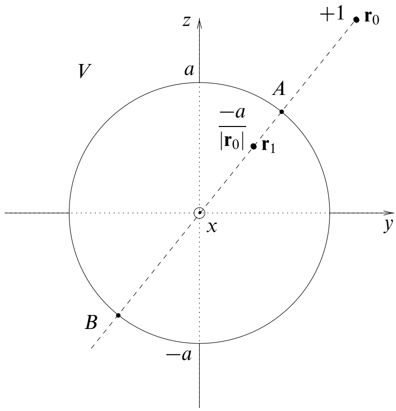


Figure 21.14 The arrangement of images for solving Poisson's equation outside a sphere of radius a centred at the origin. For a charge $+1$ at $\mathbf{r}_0$, the image point $\mathbf{r}_1$ is given by $(a/|\mathbf{r}_0|)^2\mathbf{r}_0$ and the strength of the image charge is $-a/|\mathbf{r}_0|$.

By symmetry we expect the image point $\mathbf{r}_1$ to lie on the same radial line as the original source, $\mathbf{r}_0$, as shown in figure 21.14, and so $\mathbf{r}_1 = k\mathbf{r}_0$ where $k < 1$. However, for a Dirichlet Green's function we require $G(\mathbf{r} - \mathbf{r}_0) = 0$ on $|\mathbf{r}| = a$, and the form of the Green's function suggests that we need

$$|\mathbf{r} - \mathbf{r}_0| \propto |\mathbf{r} - \mathbf{r}_1| \qquad \text{for all } |\mathbf{r}| = a. \tag{21.97}$$

Referring to figure 21.14, if this relationship is to hold over the whole surface of the sphere, then it must certainly hold for the points A and B. We thus require

$$\frac{|\mathbf{r}_0| - a}{a - |\mathbf{r}_1|} = \frac{|\mathbf{r}_0| + a}{a + |\mathbf{r}_1|},$$

which reduces to $|\mathbf{r}_1| = a^2/|\mathbf{r}_0|$. Therefore the image point must be located at the position

$$\mathbf{r}_1 = \frac{a^2}{|\mathbf{r}_0|^2}\mathbf{r}_0.$$

It may now be checked that, for this location of the image point, (21.97) is satisfied over the whole sphere. Using the geometrical result

$$\begin{aligned}|\mathbf{r} - \mathbf{r}_1|^2 &= |\mathbf{r}|^2 - \frac{2a^2}{|\mathbf{r}_0|^2}\mathbf{r}\cdot\mathbf{r}_0 + \frac{a^4}{|\mathbf{r}_0|^2} \\ &= \frac{a^2}{|\mathbf{r}_0|^2}\left(|\mathbf{r}_0|^2 - 2\mathbf{r}\cdot\mathbf{r}_0 + a^2\right) \qquad \text{for } |\mathbf{r}| = a,\end{aligned} \tag{21.98}$$

we see that, on the surface of the sphere,

$$|\mathbf{r} - \mathbf{r}_1| = \frac{a}{|\mathbf{r}_0|}|\mathbf{r} - \mathbf{r}_0| \qquad \text{for } |\mathbf{r}| = a. \tag{21.99}$$

Therefore, in order that $G = 0$ at $|\mathbf{r}| = a$, the strength of the image charge must be $-a/|\mathbf{r}_0|$. Consequently, the Dirichlet Green's function for the exterior of the sphere is

$$G(\mathbf{r}, \mathbf{r}_0) = -\frac{1}{4\pi|\mathbf{r} - \mathbf{r}_0|} + \frac{a/|\mathbf{r}_0|}{4\pi\,|\mathbf{r} - (a^2/|\mathbf{r}_0|^2)\mathbf{r}_0|}.$$

For a less formal treatment of the same problem see exercise 21.22. ◄

If we seek solutions to Poisson's equation in the *interior* of a sphere then the above analysis still holds, but $\mathbf{r}$ and $\mathbf{r}_0$ are now inside the sphere and the image $\mathbf{r}_1$ lies outside it.

For two-dimensional Dirichlet problems outside the circle $|\mathbf{r}| = a$, we are led by arguments similar to those employed previously to use the same image point as in the three-dimensional case, namely

$$\mathbf{r}_1 = \frac{a^2}{|\mathbf{r}_0|^2}\mathbf{r}_0. \tag{21.100}$$

As illustrated below, however, it is usually necessary to take the image strength as -1 in two-dimensional problems.

►*Solve Laplace's equation in the two-dimensional region $|\mathbf{r}| \leq a$, subject to the boundary condition $u = f(\phi)$ on $|\mathbf{r}| = a$.*

In this case we wish to find the Dirichlet Green's function in the interior of a disc of radius a, so the image charge must lie outside the disc. Taking the strength of the image to be -1, we have

$$G(\mathbf{r}, \mathbf{r}_0) = \frac{1}{2\pi}\ln|\mathbf{r} - \mathbf{r}_0| - \frac{1}{2\pi}\ln|\mathbf{r} - \mathbf{r}_1| + c,$$

where $\mathbf{r}_1 = (a^2/|\mathbf{r}_0|^2)\mathbf{r}_0$ lies outside the disc, and c is a constant that includes the strength of the image charge and does not necessarily equal zero.

Since we require $G(\mathbf{r}, \mathbf{r}_0) = 0$ when $|\mathbf{r}| = a$, the value of the constant c is determined, and the Dirichlet Green's function for this problem is given by

$$G(\mathbf{r}, \mathbf{r}_0) = \frac{1}{2\pi}\left(\ln|\mathbf{r} - \mathbf{r}_0| - \ln\left|\mathbf{r} - \frac{a^2}{|\mathbf{r}_0|^2}\mathbf{r}_0\right| - \ln\frac{|\mathbf{r}_0|}{a}\right). \tag{21.101}$$

Using plane polar coordinates, the solution to the boundary-value problem can be written as a line integral around the circle $\rho = a$:

$$\begin{aligned} u(\mathbf{r}_0) &= \int_C f(\mathbf{r})\frac{\partial G(\mathbf{r}, \mathbf{r}_0)}{\partial n}\,dl \\ &= \int_0^{2\pi} f(\mathbf{r})\left.\frac{\partial G(\mathbf{r}, \mathbf{r}_0)}{\partial \rho}\right|_{\rho=a} a\,d\phi. \end{aligned} \tag{21.102}$$

The normal derivative of the Green's function (21.101) is given by

$$\begin{aligned} \frac{\partial G(\mathbf{r}, \mathbf{r}_0)}{\partial \rho} &= \frac{\mathbf{r}}{|\mathbf{r}|}\cdot\nabla G(\mathbf{r}, \mathbf{r}_0) \\ &= \frac{\mathbf{r}}{2\pi|\mathbf{r}|}\cdot\left(\frac{\mathbf{r} - \mathbf{r}_0}{|\mathbf{r} - \mathbf{r}_0|^2} - \frac{\mathbf{r} - \mathbf{r}_1}{|\mathbf{r} - \mathbf{r}_1|^2}\right). \end{aligned} \tag{21.103}$$

Using the fact that $\mathbf{r}_1 = (a^2/|\mathbf{r}_0|^2)\mathbf{r}_0$ and the geometrical result (21.99), we find that

$$\left.\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial \rho}\right|_{\rho=a} = \frac{a^2 - |\mathbf{r}_0|^2}{2\pi a|\mathbf{r}-\mathbf{r}_0|^2}.$$

In plane polar coordinates, $\mathbf{r} = \rho\cos\phi\,\mathbf{i} + \rho\sin\phi\,\mathbf{j}$ and $\mathbf{r}_0 = \rho_0\cos\phi_0\,\mathbf{i} + \rho_0\sin\phi_0\,\mathbf{j}$, and so

$$\left.\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial \rho}\right|_{\rho=a} = \left(\frac{1}{2\pi a}\right)\frac{a^2-\rho_0^2}{a^2+\rho_0^2-2a\rho_0\cos(\phi-\phi_0)}.$$

On substituting into (21.102), we obtain

$$u(\rho_0,\phi_0) = \frac{1}{2\pi}\int_0^{2\pi}\frac{(a^2-\rho_0^2)f(\phi)\,d\phi}{a^2+\rho_0^2-2a\rho_0\cos(\phi-\phi_0)}, \qquad (21.104)$$

which is the solution to the problem. ◄

21.5.4 Neumann problems

In a Neumann problem we require the normal derivative of the solution of Poisson's equation to take on specific values on some surface S that bounds V, i.e. we require $\partial u(\mathbf{r})/\partial n = f(\mathbf{r})$ on S, where f is a given function. As we shall see, much of our discussion of Dirichlet problems can be immediately taken over into the solution of Neumann problems.

As we proved in section 20.7 of the previous chapter, specifying Neumann boundary conditions determines the relevant solution of Poisson's equation to within an (unimportant) additive constant. Unlike Dirichlet conditions, Neumann conditions impose a self-consistency requirement. In order for a solution u to exist, it is necessary that the following consistency condition holds:

$$\int_S f\,dS = \int_S \nabla u\cdot\hat{\mathbf{n}}\,dS = \int_V \nabla^2 u\,dV = \int_V \rho\,dV, \qquad (21.105)$$

where we have used the divergence theorem to convert the surface integral into a volume integral. As a physical example, the integral of the normal component of an electric field over a surface bounding a given volume cannot be chosen arbitrarily when the charge inside the volume has already been specified (Gauss's theorem).

Let us again consider (21.84), which is central to our discussion of Green's functions in inhomogeneous problems. It reads

$$u(\mathbf{r}_0) = \int_V G(\mathbf{r},\mathbf{r}_0)\rho(\mathbf{r})\,dV(\mathbf{r}) + \int_S\left[u(\mathbf{r})\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial n} - G(\mathbf{r},\mathbf{r}_0)\frac{\partial u(\mathbf{r})}{\partial n}\right]dS(\mathbf{r}).$$

As always, the Green's function must obey

$$\nabla^2 G(\mathbf{r},\mathbf{r}_0) = \delta(\mathbf{r}-\mathbf{r}_0),$$

where $\mathbf{r}_0$ lies in V. In the solution of Dirichlet problems in the previous subsection, we chose the Green's function to obey the boundary condition $G(\mathbf{r},\mathbf{r}_0) = 0$ on S

and, in a similar way, we might wish to choose $\partial G(\mathbf{r},\mathbf{r}_0)/\partial n = 0$ in the solution of Neumann problems. However, in general this is *not* permitted since the Green's function must obey the consistency condition

$$\int_S \frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial n}\,dS = \int_S \nabla G(\mathbf{r},\mathbf{r}_0)\cdot\hat{\mathbf{n}}\,dS = \int_V \nabla^2 G(\mathbf{r},\mathbf{r}_0)\,dV = 1.$$

The simplest permitted boundary condition is therefore

$$\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial n} = \frac{1}{A} \qquad \text{for } \mathbf{r} \text{ on } S,$$

where A is the area of the surface S; this defines a *Neumann Green's function.*

If we require $\partial u(\mathbf{r})/\partial n = f(\mathbf{r})$ on S, the solution to Poisson's equation is given by

$$\begin{aligned} u(\mathbf{r}_0) &= \int_V G(\mathbf{r},\mathbf{r}_0)\rho(\mathbf{r})\,dV(\mathbf{r}) + \frac{1}{A}\int_S u(\mathbf{r})\,dS(\mathbf{r}) - \int_S G(\mathbf{r},\mathbf{r}_0)f(\mathbf{r})\,dS(\mathbf{r}) \\ &= \int_V G(\mathbf{r},\mathbf{r}_0)\rho(\mathbf{r})\,dV(\mathbf{r}) + \langle u(\mathbf{r})\rangle_S - \int_S G(\mathbf{r},\mathbf{r}_0)f(\mathbf{r})\,dS(\mathbf{r}), \end{aligned} \tag{21.106}$$

where $\langle u(\mathbf{r})\rangle_S$ is the average of u over the surface S and is a freely specifiable constant. For Neumann problems in which the volume V is bounded by a surface S at infinity, we do not need the $\langle u(\mathbf{r})\rangle_S$ term. For example, if we wish to solve a Neumann problem outside the unit sphere centred at the origin then $r > a$ is the region V throughout which we require the solution; this region may be considered as being bounded by two disconnected surfaces, the surface of the sphere and a surface at infinity. By requiring that $u(\mathbf{r}) \to 0$ as $|\mathbf{r}| \to \infty$, the term $\langle u(\mathbf{r})\rangle_S$ becomes zero.

As mentioned above, much of our discussion of Dirichlet problems can be taken over into the solution of Neumann problems. In particular, we may use the method of images to find the appropriate Neumann Green's function.

►*Solve Laplace's equation in the two-dimensional region $|\mathbf{r}| \le a$ subject to the boundary condition $\partial u/\partial n = f(\phi)$ on $|\mathbf{r}| = a$, with $\int_0^{2\pi} f(\phi)\,d\phi = 0$ as required by the consistency condition (21.105).*

Let us assume, as in Dirichlet problems with this geometry, that a single image charge is placed outside the circle at

$$\mathbf{r}_1 = \frac{a^2}{|\mathbf{r}_0|^2}\mathbf{r}_0,$$

where $\mathbf{r}_0$ is the position of the source inside the circle (see equation (21.100)). Then, from (21.99), we have the useful geometrical result

$$|\mathbf{r}-\mathbf{r}_1| = \frac{a}{|\mathbf{r}_0|}|\mathbf{r}-\mathbf{r}_0| \qquad \text{for } |\mathbf{r}| = a. \tag{21.107}$$

Leaving the strength q of the image as a parameter, the Green's function has the form

$$G(\mathbf{r},\mathbf{r}_0) = \frac{1}{2\pi}\left(\ln|\mathbf{r}-\mathbf{r}_0| + q\ln|\mathbf{r}-\mathbf{r}_1| + c\right). \tag{21.108}$$

Using plane polar coordinates, the radial (i.e. normal) derivative of this function is given by

$$\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial \rho} = \frac{\mathbf{r}}{|\mathbf{r}|}\cdot \nabla G(\mathbf{r},\mathbf{r}_0)$$
$$= \frac{\mathbf{r}}{2\pi|\mathbf{r}|}\cdot\left[\frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|^2}+\frac{q(\mathbf{r}-\mathbf{r}_1)}{|\mathbf{r}-\mathbf{r}_1|^2}\right].$$

Using (21.107), on the perimeter of the circle $\rho = a$ the radial derivative takes the form

$$\left.\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial \rho}\right|_{\rho=a} = \frac{1}{2\pi|\mathbf{r}|}\left[\frac{|\mathbf{r}|^2-\mathbf{r}\cdot\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|^2}+\frac{q|\mathbf{r}|^2-q(a^2/|\mathbf{r}_0|^2)\mathbf{r}\cdot\mathbf{r}_0}{(a^2/|\mathbf{r}_0|^2)|\mathbf{r}-\mathbf{r}_0|^2}\right]$$
$$= \frac{1}{2\pi a}\frac{1}{|\mathbf{r}-\mathbf{r}_0|^2}\left[|\mathbf{r}|^2+q|\mathbf{r}_0|^2-(1+q)\mathbf{r}\cdot\mathbf{r}_0\right],$$

where we have set $|\mathbf{r}|^2 = a^2$ in the second term on the RHS, but not in the first. If we take $q = 1$, the radial derivative simplifies to

$$\left.\frac{\partial G(\mathbf{r},\mathbf{r}_0)}{\partial \rho}\right|_{\rho=a} = \frac{1}{2\pi a},$$

or $1/L$, where L is the circumference, and so (21.108) with $q = 1$ is the required Neumann Green's function.

Since $\rho(\mathbf{r}) = 0$, the solution to our boundary-value problem is now given by (21.106) as

$$u(\mathbf{r}_0) = \langle u(\mathbf{r})\rangle_C - \int_C G(\mathbf{r},\mathbf{r}_0)f(\mathbf{r})\,dl(\mathbf{r}),$$

where the integral is around the circumference of the circle C. In plane polar coordinates $\mathbf{r} = \rho\cos\phi\,\mathbf{i}+\rho\sin\phi\,\mathbf{j}$ and $\mathbf{r}_0 = \rho_0\cos\phi_0\,\mathbf{i}+\rho_0\sin\phi_0\,\mathbf{j}$, and again using (21.107) we find that on C the Green's function is given by

$$G(\mathbf{r},\mathbf{r}_0)|_{\rho=a} = \frac{1}{2\pi}\left[\ln|\mathbf{r}-\mathbf{r}_0| + \ln\left(\frac{a}{|\mathbf{r}_0|}|\mathbf{r}-\mathbf{r}_0|\right)+c\right]$$
$$= \frac{1}{2\pi}\left(\ln|\mathbf{r}-\mathbf{r}_0|^2 + \ln\frac{a}{|\mathbf{r}_0|}+c\right)$$
$$= \frac{1}{2\pi}\left\{\ln\left[a^2+\rho_0^2-2a\rho_0\cos(\phi-\phi_0)\right]+\ln\frac{a}{\rho_0}+c\right\}. \qquad (21.109)$$

Since $dl = a\,d\phi$ on C, the solution to the problem is given by

$$u(\rho_0,\phi_0) = \langle u\rangle_C - \frac{a}{2\pi}\int_0^{2\pi} f(\phi)\ln[a^2+\rho_0^2-2a\rho_0\cos(\phi-\phi_0)]\,d\phi.$$

The contributions of the final two terms terms in the Green's function (21.109) vanish because $\int_0^{2\pi} f(\phi)\,d\phi = 0$. The average value of u around the circumference, $\langle u\rangle_C$, is a freely specifiable constant as we would expect for a Neumann problem. This result should be compared with the result (21.104) for the corresponding Dirichlet problem, but it should be remembered that in the one case $f(\phi)$ is a potential, and in the other the gradient of a potential. ◄

21.6 Exercises

21.1 Solve the following first-order partial differential equations by separating the variables:

$$\text{(a)}\ \frac{\partial u}{\partial x} - x\frac{\partial u}{\partial y} = 0; \qquad \text{(b)}\ x\frac{\partial u}{\partial x} - 2y\frac{\partial u}{\partial y} = 0.$$

21.2 A cube, made of material whose conductivity is k, has as its six faces the planes $x = \pm a$, $y = \pm a$ and $z = \pm a$, and contains no internal heat sources. Verify that the temperature distribution

$$u(x, y, z, t) = A \cos \frac{\pi x}{a} \sin \frac{\pi z}{a} \exp\left(-\frac{2\kappa\pi^2 t}{a^2}\right)$$

obeys the appropriate diffusion equation. Across which faces is there heat flow? What is the direction and rate of heat flow at the point $(3a/4, a/4, a)$ at time $t = a^2/(\kappa\pi^2)$?

21.3 The wave equation describing the transverse vibrations of a stretched membrane under tension T and having a uniform surface density ρ is

$$T\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = \rho\frac{\partial^2 u}{\partial t^2}.$$

Find a separable solution appropriate to a membrane stretched on a frame of length a and width b, showing that the natural angular frequencies of such a membrane are given by

$$\omega^2 = \frac{\pi^2 T}{\rho}\left(\frac{n^2}{a^2} + \frac{m^2}{b^2}\right),$$

where n and m are any positive integers.

21.4 Schrödinger's equation for a non-relativistic particle in a constant potential region can be taken as

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) = i\hbar\frac{\partial u}{\partial t}.$$

(a) Find a solution, separable in the four independent variables, that can be written in the form of a plane wave,

$$\psi(x, y, z, t) = A \exp[i(\mathbf{k}\cdot\mathbf{r} - \omega t)].$$

Using the relationships associated with de Broglie ($\mathbf{p} = \hbar\mathbf{k}$) and Einstein ($E = \hbar\omega$), show that the separation constants must be such that

$$p_x^2 + p_y^2 + p_z^2 = 2mE.$$

(b) Obtain a different separable solution describing a particle confined to a box of side a (ψ must vanish at the walls of the box). Show that the energy of the particle can only take the quantised values

$$E = \frac{\hbar^2\pi^2}{2ma^2}(n_x^2 + n_y^2 + n_z^2),$$

where n_x, n_y and n_z are integers.

21.5 Denoting the three terms of ∇^2 in spherical polars by ∇_r^2, ∇_θ^2, ∇_ϕ^2 in an obvious way, evaluate $\nabla_r^2 u$, etc. for the two functions given below and verify that, in each case, although the individual terms are not necessarily zero their sum $\nabla^2 u$ is zero. Identify the corresponding values of ℓ and m.

(a) $u(r, \theta, \phi) = \left(Ar^2 + \dfrac{B}{r^3}\right)\dfrac{3\cos^2\theta - 1}{2}.$

(b) $u(r, \theta, \phi) = \left(Ar + \dfrac{B}{r^2}\right)\sin\theta \exp i\phi.$

21.6 Prove that the expression given in equation (21.47) for the associated Legendre function $P_\ell^m(\mu)$ satisfies the appropriate equation, (21.45), as follows.

(a) Evaluate $dP_\ell^m(\mu)/d\mu$ and $d^2P_\ell^m(\mu)/d\mu^2$, using the forms given in (21.47), and substitute them into (21.45).
(b) Differentiate Legendre's equation m times using Leibnitz' theorem.
(c) Show that the equations obtained in (a) and (b) are multiples of each other, and hence that the validity of (b) implies that of (a).

21.7 Continue the analysis of exercise 10.20, concerned with the flow of a very viscous fluid past a sphere, to find the full expression for the stream function $\psi(r,\theta)$. At the surface of the sphere $r = a$, the velocity field $\mathbf{u} = \mathbf{0}$, whilst far from the sphere $\psi \simeq (Ur^2 \sin^2\theta)/2$.

Show that $f(r)$ can be expressed as a superposition of powers of r, and determine which powers give acceptable solutions. Hence show that

$$\psi(r,\theta) = \frac{U}{4}\left(2r^2 - 3ar + \frac{a^3}{r}\right)\sin^2\theta.$$

21.8 The motion of a very viscous fluid in the two-dimensional (wedge) region $-\alpha < \phi < \alpha$ can be described, in (ρ,ϕ) coordinates, by the (biharmonic) equation

$$\nabla^2\nabla^2\psi \equiv \nabla^4\psi = 0,$$

together with the boundary conditions $\partial\psi/\partial\phi = 0$ at $\phi = \pm\alpha$, which represent the fact that there is no radial fluid velocity close to either of the bounding walls because of the viscosity, and $\partial\psi/\partial\rho = \pm\rho$ at $\phi = \pm\alpha$, which impose the condition that azimuthal flow increases linearly with r along any radial line. Assuming a solution in separated-variable form, show that the full expression for ψ is

$$\psi(\rho,\phi) = \frac{\rho^2}{2}\,\frac{\sin 2\phi - 2\phi\cos 2\alpha}{\sin 2\alpha - 2\alpha\cos 2\alpha}.$$

21.9 A circular disc of radius a is heated in such a way that its perimeter $\rho = a$ has a steady temperature distribution $A + B\cos^2\phi$, where ρ and ϕ are plane polar coordinates and A and B are constants. Find the temperature $T(\rho,\phi)$ everywhere in the region $\rho < a$.

21.10 Consider possible solutions of Laplace's equation inside a circular domain as follows.

(a) Find the solution in plane polar coordinates ρ, ϕ, that takes the value $+1$ for $0 < \phi < \pi$ and the value -1 for $-\pi < \phi < 0$, when $\rho = a$.
(b) For a point (x,y) on or inside the circle $x^2 + y^2 = a^2$, identify the angles α and β defined by

$$\alpha = \tan^{-1}\frac{y}{a+x} \quad \text{and} \quad \beta = \tan^{-1}\frac{y}{a-x}.$$

Show that $u(x,y) = (2/\pi)(\alpha+\beta)$ is a solution of Laplace's equation that satisfies the boundary conditions given in (a).
(c) Deduce a Fourier series expansion for the function

$$\tan^{-1}\frac{\sin\phi}{1+\cos\phi} + \tan^{-1}\frac{\sin\phi}{1-\cos\phi}.$$

21.11 The free transverse vibrations of a thick rod satisfy the equation

$$a^4\frac{\partial^4 u}{\partial x^4} + \frac{\partial^2 u}{\partial t^2} = 0.$$

Obtain a solution in separated-variable form and, for a rod clamped at one end,

$x = 0$, and free at the other, $x = L$, show that the angular frequency of vibration ω satisfies

$$\cosh\left(\frac{\omega^{1/2}L}{a}\right) = -\sec\left(\frac{\omega^{1/2}L}{a}\right).$$

[At a clamped end both u and $\partial u/\partial x$ vanish, whilst at a free end, where there is no bending moment, $\partial^2 u/\partial x^2$ and $\partial^3 u/\partial x^3$ are both zero.]

21.12 A membrane is stretched between two concentric rings of radii a and b ($b > a$). If the smaller ring is transversely distorted from the planar configuration by an amount $c|\phi|$, $-\pi \le \phi \le \pi$, show that the membrane then has a shape given by

$$u(\rho, \phi) = \frac{c\pi}{2}\frac{\ln(b/\rho)}{\ln(b/a)} - \frac{4c}{\pi}\sum_{m \text{ odd}} \frac{a^m}{m^2(b^{2m} - a^{2m})}\left(\frac{b^{2m}}{\rho^m} - \rho^m\right)\cos m\phi.$$

21.13 A string of length L, fixed at its two ends, is plucked at its mid-point by an amount A and then released. Prove that the subsequent displacement is given by

$$u(x,t) = \sum_{n=0}^{\infty} \frac{8A\,(-1)^n}{\pi^2(2n+1)^2}\sin\left[\frac{(2n+1)\pi x}{L}\right]\cos\left[\frac{(2n+1)\pi ct}{L}\right],$$

where, in the usual notation, $c^2 = T/\rho$.

Find the total kinetic energy of the string when it passes through its unplucked position, by calculating it in each mode (each n) and summing, using the result

$$\sum_0^{\infty} \frac{1}{(2n+1)^2} = \frac{\pi^2}{8}.$$

Confirm that the total energy is equal to the work done in plucking the string initially.

21.14 Prove that the potential for $\rho < a$ associated with a vertical split cylinder of radius a, the two halves of which ($\cos\phi > 0$ and $\cos\phi < 0$) are maintained at equal and opposite potentials $\pm V$, is given by

$$u(\rho, \phi) = \frac{4V}{\pi}\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}\left(\frac{\rho}{a}\right)^{2n+1}\cos(2n+1)\phi.$$

21.15 A conducting spherical shell of radius a is cut round its equator and the two halves connected to voltages of $+V$ and $-V$. Show that an expression for the potential at the point (r, θ, ϕ) anywhere inside the two hemispheres is

$$u(r, \theta, \phi) = V\sum_{n=0}^{\infty} \frac{(-1)^n(2n)!(4n+3)}{2^{2n+1}n!(n+1)!}\left(\frac{r}{a}\right)^{2n+1} P_{2n+1}(\cos\theta).$$

[This is the spherical polar analogue of the previous question.]

21.16 A slice of biological material of thickness L is placed into a solution of a radioactive isotope of constant concentration C_0 at time $t = 0$. For a later time t find the concentration of radioactive ions at a depth x inside one of its surfaces if the diffusion constant is κ.

21.17 Two identical copper bars are each of length a. Initially, one is at 0 °C and the other is at 100 °C; they are then joined together end to end and thermally isolated. Obtain in the form of a Fourier series an expression $u(x, t)$ for the temperature at any point a distance x from the join at a later time t. Bear in mind the heat flow conditions at the free ends of the bars.

Taking $a = 0.5$ m estimate the time it takes for one of the free ends to attain a temperature of 55 °C. The thermal conductivity of copper is 3.8×10^2 J m^{-1} K^{-1} s^{-1}, and its specific heat capacity is 3.4×10^6 J m^{-3} K^{-1}.

21.18 A sphere of radius a and thermal conductivity k_1 is surrounded by an infinite medium of conductivity k_2 in which far away the temperature tends to T_∞. A distribution of heat sources $q(\theta)$ embedded in the sphere's surface establish steady temperature fields $T_1(r, \theta)$ inside the sphere and $T_2(r, \theta)$ outside it. It can be shown, by considering the heat flow through a small volume that includes part of the sphere's surface, that

$$k_1 \frac{\partial T_1}{\partial r} - k_2 \frac{\partial T_2}{\partial r} = q(\theta) \quad \text{on} \quad r = a.$$

Given that

$$q(\theta) = \frac{1}{a} \sum_{n=0}^{\infty} q_n P_n(\cos\theta),$$

find complete expressions for $T_1(r, \theta)$ and $T_2(r, \theta)$. What is the temperature at the centre of the sphere?

21.19 Using result (21.74) from the worked example in the text, find the general expression for the temperature $u(x, t)$ in the bar, given that the temperature distribution at time $t = 0$ is $u(x, 0) = \exp(-x^2/a^2)$.

21.20 Working in *spherical* polar coordinates $\mathbf{r} = (r, \theta, \phi)$, but for a system that has azimuthal symmetry around the polar axis, consider the following gravitational problem.

(a) Show that the gravitational potential due to a uniform disc of radius a and mass M, centred at the origin, is given for $r < a$ by

$$\frac{2GM}{a}\left[1 - \frac{r}{a}P_1(\cos\theta) + \frac{1}{2}\left(\frac{r}{a}\right)^2 P_2(\cos\theta) - \frac{1}{8}\left(\frac{r}{a}\right)^4 P_4(\cos\theta) + \cdots\right],$$

and for $r > a$ by

$$\frac{GM}{r}\left[1 - \frac{1}{4}\left(\frac{a}{r}\right)^2 P_2(\cos\theta) + \frac{1}{8}\left(\frac{a}{r}\right)^4 P_4(\cos\theta) - \cdots\right],$$

where the polar axis is normal to the plane of the disc.

(b) Reconcile the presence of a term $P_1(\cos\theta)$, which is odd under $\theta \to \pi - \theta$, with the symmetry with respect to the plane of the disc of the physical system.

(c) Deduce that the gravitational field near an infinite sheet of matter of constant density ρ per unit area is $2\pi G\rho$.

21.21 In the region $-\infty < x,\ y < \infty$ and $-t \le z \le t$, a charge-density wave $\rho(\mathbf{r}) = A\cos qx$, in the x-direction, is represented by

$$\rho(\mathbf{r}) = \frac{e^{iqx}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{\rho}(\alpha) e^{i\alpha z}\, d\alpha.$$

The resulting potential is represented by

$$V(\mathbf{r}) = \frac{e^{iqx}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{V}(\alpha) e^{i\alpha z}\, d\alpha.$$

Determine the relationship between $\tilde{V}(\alpha)$ and $\tilde{\rho}(\alpha)$, and hence show that the potential at the point $(0, 0, 0)$ is

$$\frac{A}{\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{\sin kt}{k(k^2 + q^2)}\, dk.$$

21.22 Point charges q and $-qa/b$ (with $a < b$) are placed, respectively, at a point P, a distance b from the origin O, and a point Q between O and P, a distance a^2/b from O. Show, by considering similar triangles QOS and SOP, where S is any point on the surface of the sphere centred at O and of radius a, that the net potential anywhere on the sphere due to the two charges is zero.

Use this result (backed up by the uniqueness theorem) to find the force with which a point charge q placed a distance b from the centre of a spherical conductor of radius a ($< b$) is attracted to the sphere (i) if the sphere is earthed, and (ii) if the sphere is uncharged and insulated.

21.23 Find the Green's function $G(\mathbf{r}, \mathbf{r}_0)$ in the half-space $z > 0$ for the solution of $\nabla^2\Phi = 0$ with Φ specified in cylindrical polar coordinates (ρ, ϕ, z) on the plane $z = 0$ by

$$\Phi(\rho, \phi, z) = \begin{cases} 1 & \text{for } \rho \leq 1, \\ 1/\rho & \text{for } \rho > 1. \end{cases}$$

Determine the variation of $\Phi(0, 0, z)$ along the z-axis.

21.24 Electrostatic charge is distributed in a sphere of radius R centred on the origin. Determine the form of the resultant potential $\phi(\mathbf{r})$ at distances much greater than R, as follows.

(a) Express in the form of an integral over all space the solution of

$$\nabla^2\phi = -\frac{\rho(\mathbf{r})}{\epsilon_0}.$$

(b) Show that, for $r \gg r'$,

$$|\mathbf{r} - \mathbf{r}'| = r - \frac{\mathbf{r} \cdot \mathbf{r}'}{r} + \mathrm{O}\left(\frac{1}{r}\right).$$

(c) Use results (a) and (b) to show that $\phi(\mathbf{r})$ has the form

$$\phi(\mathbf{r}) = \frac{M}{r} + \frac{\mathbf{d} \cdot \mathbf{r}}{r^3} + \mathrm{O}\left(\frac{1}{r^3}\right).$$

Find expressions for M and $\mathbf{d}$, and identify them physically.

21.25 Find, in the form of an infinite series, the Green's function of the ∇^2 operator for the Dirichlet problem in the region $-\infty < x < \infty$, $-\infty < y < \infty$, $-c \leq z \leq c$.

21.26 Find the Green's function for the three-dimensional Neumann problem

$$\nabla^2\phi = 0 \quad \text{for } z > 0 \qquad \text{and} \qquad \frac{\partial\phi}{\partial z} = f(x, y) \quad \text{on } z = 0.$$

Determine $\phi(x, y, z)$ if

$$f(x, y) = \begin{cases} \delta(y) & \text{for } |x| < a, \\ 0 & \text{for } |x| \geq a. \end{cases}$$

21.27 Determine the Green's function for the Klein–Gordon equation in a half-space as follows.

(a) By applying the divergence theorem to the volume integral

$$\int_V \left[\phi(\nabla^2 - m^2)\psi - \psi(\nabla^2 - m^2)\phi\right] dV,$$

obtain a Green's function expression, as the sum of a volume integral and a surface integral, for the function $\phi(\mathbf{r}')$ that satisfies

$$\nabla^2\phi - m^2\phi = \rho$$

in V and takes the specified form $\phi = f$ on S, the boundary of V. The Green's function, $G(\mathbf{r}, \mathbf{r}')$, to be used satisfies

$$\nabla^2 G - m^2 G = \delta(\mathbf{r} - \mathbf{r}')$$

and vanishes when $\mathbf{r}$ is on S.

(b) When V is all space, $G(\mathbf{r}, \mathbf{r}')$ can be written as $G(t) = g(t)/t$, where $t = |\mathbf{r} - \mathbf{r}'|$ and $g(t)$ is bounded as $t \to \infty$. Find the form of $G(t)$.

(c) Find $\phi(\mathbf{r})$ in the half-space $x > 0$ if $\rho(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}_1)$ and $\phi = 0$ both on $x = 0$ and as $r \to \infty$.

21.28 Consider the PDE $\mathcal{L}u(\mathbf{r}) = \rho(\mathbf{r})$, for which the differential operator $\mathcal{L}$ is given by

$$\mathcal{L} = \nabla \cdot [p(\mathbf{r})\nabla] + q(\mathbf{r}),$$

where $p(\mathbf{r})$ and $q(\mathbf{r})$ are functions of position. By proving the generalised form of Green's theorem,

$$\int_V (\phi\mathcal{L}\psi - \psi\mathcal{L}\phi)\, dV = \oint_S p(\phi\nabla\psi - \psi\nabla\phi) \cdot \hat{\mathbf{n}}\, dS,$$

show that the solution of the PDE is given by

$$u(\mathbf{r}_0) = \int_V G(\mathbf{r}, \mathbf{r}_0)\rho(\mathbf{r})\, dV(\mathbf{r}) + \oint_S p(\mathbf{r}) \left[u(\mathbf{r}) \frac{\partial G(\mathbf{r}, \mathbf{r}_0)}{\partial n} - G(\mathbf{r}, \mathbf{r}_0) \frac{\partial u(\mathbf{r})}{\partial n} \right] dS(\mathbf{r}),$$

where $G(\mathbf{r}, \mathbf{r}_0)$ is the Green's function satisfying $\mathcal{L}G(\mathbf{r}, \mathbf{r}_0) = \delta(\mathbf{r} - \mathbf{r}_0)$.

21.7 Hints and answers

21.1 (a) $C \exp[\lambda(x^2 + 2y)]$; (b) $C(x^2 y)^\lambda$.

21.3 $u(x, y, t) = \sin(n\pi x/a)\sin(m\pi y/b)(A \sin \omega t + B \cos \omega t)$.

21.5 (a) $6u/r^2$, $-6u/r^2$, 0, $\ell = 2$ (or -3), $m = 0$;
(b) $2u/r^2$, $(\cot^2\theta - 1)u/r^2$; $-u/(r^2 \sin^2\theta)$, $\ell = 1$ (or -2), $m = \pm 1$.

21.7 Solutions of the form r^ℓ give ℓ as $-1, 1, 2, 4$. Because of the asymptotic form of ψ, an r^4 term cannot be present. The coefficients of the three remaining terms are determined by the two boundary conditions $\mathbf{u} = \mathbf{0}$ on the sphere and the form of ψ for large r.

21.9 Express $\cos^2\phi$ in terms of $\cos 2\phi$; $T(\rho, \phi) = A + B/2 + (B\rho^2/2a^2)\cos 2\phi$.

21.11 $(A \cos mx + B \sin mx + C \cosh mx + D \sinh mx)\cos(\omega t + \epsilon)$, with $m^4 a^4 = \omega^2$.

21.13 $E_n = 16\rho A^2 c^2/[(2n+1)^2\pi^2 L]$; $E = 2\rho c^2 A^2/L = \int_0^A [2Tv/(\frac{1}{2}L)]\, dv$.

21.15 Note that the boundary value function is a square wave that is *symmetric* in ϕ.

21.17 Since there is no heat flow at $x = \pm a$, use a series of period $4a$, $u(x, 0) = 100$ for $0 < x \le 2a$, $u(x, 0) = 0$ for $-2a \le x < 0$.

$$u(x, t) = 50 + \frac{200}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin\left[\frac{(2n+1)\pi x}{2a}\right]\exp\left[-\frac{k(2n+1)^2\pi^2 t}{4a^2 s}\right].$$

Taking only the $n = 0$ term gives $t \approx 2300$ s.

21.19 $u(x, t) = [a/(a^2 + 4\kappa t)^{1/2}]\exp[-x^2/(a^2 + 4\kappa t)]$.

21.21 Fourier-transform Poisson's equation to show that $\tilde{\rho}(\alpha) = \epsilon_0(\alpha^2 + q^2)\tilde{V}(\alpha)$.

21.23 Follow the worked example that includes result (21.95). For part of the explicit integration, substitute $\rho = z \tan\alpha$.

$$\Phi(0, 0, z) = \frac{z(1+z^2)^{1/2} - z^2 + (1+z^2)^{1/2} - 1}{z(1+z^2)^{1/2}}.$$

21.25 The terms in $G(\mathbf{r}, \mathbf{r}_0)$ that are additional to the fundamental solution are

$$\frac{1}{4\pi}\sum_{n=2}^{\infty}(-1)^n\left\{\left[(x-x_0)^2+(y-y_0)^2+(z+(-1)^n z_0-nc)^2\right]^{-1/2}\right.$$
$$\left.+\left[(x-x_0)^2+(y-y_0)^2+(z+(-1)^n z_0+nc)^2\right]^{-1/2}\right\}.$$

21.27 (a) As given in equation (21.86), but with $\mathbf{r}_0$ replaced by $\mathbf{r}'$.

(b) Move the origin to $\mathbf{r}'$ and integrate the defining Green's equation to obtain

$$4\pi t^2\frac{dG}{dt} - m^2\int_0^t G(t')\, 4\pi t'^2\, dt' = 1,$$

leading to $G(t) = [-1/(4\pi t)]e^{-mt}$.

(c) $\phi(\mathbf{r}) = [-1/(4\pi)](p^{-1}e^{-mp} - q^{-1}e^{-mq})$, where $p = |\mathbf{r}-\mathbf{r}_1|$ and $q = |\mathbf{r}-\mathbf{r}_2|$ with $\mathbf{r}_1 = (x_1, y_1, z_1)$ and $\mathbf{r}_2 = (-x_1, y_1, z_1)$.

22

Calculus of variations

In chapters 2 and 5 we discussed how to find stationary values of functions of a single variable $f(x)$, of several variables $f(x, y, \dots)$ and of constrained variables, where $x, y, \dots$ are subject to the n constraints $g_i(x, y, \dots) = 0$, $i = 1, 2, \dots, n$. In all these cases the forms of the functions f and g_i were known, and the problem was one of finding the appropriate values of the variables x, y etc.

We now turn to a different kind of problem in which we are interested in bringing about a particular condition for a given expression (usually maximising or minimising it) by varying the *functions* on which the expression depends. For instance, we might want to know in what shape a fixed length of rope should be arranged so as to enclose the largest possible area, or in what shape it will hang when suspended under gravity from two fixed points. In each case we are concerned with a general maximisation or minimisation criterion by which the function $y(x)$ that satisfies the given problem may be found.

The calculus of variations provides a method for finding the function $y(x)$. The problem must first be expressed in a mathematical form, and the form most commonly applicable to such problems is an *integral*. In each of the above questions, the quantity that has to be maximised or minimised by an appropriate choice of the function $y(x)$ may be expressed as an integral involving $y(x)$ and the variables describing the geometry of the situation.

In our example of the rope hanging from two fixed points, we need to find the shape function $y(x)$ that minimises the gravitational potential energy of the rope. Each elementary piece of the rope has a gravitational potential energy proportional both to its vertical height above an arbitrary zero level and to the length of the piece. Therefore the total potential energy is given by an integral for the whole rope of such elementary contributions. The particular function $y(x)$ for which the value of this integral is a minimum will give the shape assumed by the hanging rope.

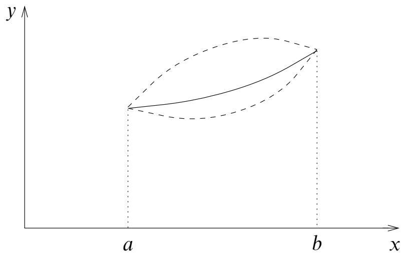


Figure 22.1 Possible paths for the integral (22.1). The solid line is the curve along which the integral is assumed stationary. The broken curves represent small variations from this path.

So in general we are led by this type of question to study the value of an integral whose integrand has a specified form in terms of a certain function and its derivatives, and to study how that value changes when the form of the function is varied. Specifically, we aim to find the function that makes the integral *stationary*, i.e. the function that makes the value of the integral a local maximum or minimum. Note that, unless stated otherwise, y' is used to denote dy/dx throughout this chapter. We also assume that all the functions we need to deal with are sufficiently smooth and differentiable.

22.1 The Euler–Lagrange equation

Let us consider the integral

$$I = \int_a^b F(y, y', x)\, dx, \tag{22.1}$$

where a, b and the form of the function F are fixed by given considerations, e.g. the physics of the problem, but the curve $y(x)$ is to be chosen so as to make stationary the value of I, which is clearly a function, or more accurately a *functional*, of this curve, i.e. $I = I[y(x)]$. Referring to figure 22.1, we wish to find the function $y(x)$ (given, say, by the solid line) such that first-order small changes in it (for example the two broken lines) will make only second-order changes in the value of I.

Writing this in a more mathematical form, let us suppose that $y(x)$ is the function required to make I stationary and consider making the replacement

$$y(x) \to y(x) + \alpha\eta(x), \tag{22.2}$$

where the parameter α is small and $\eta(x)$ is an arbitrary function with sufficiently amenable mathematical properties. For the value of I to be stationary with respect to these variations, we require

$$\left.\frac{dI}{d\alpha}\right|_{\alpha=0} = 0 \qquad \text{for all } \eta(x). \tag{22.3}$$

Substituting (22.2) into (22.1) and expanding as a Taylor series in α we obtain

$$\begin{aligned} I(y,\alpha) &= \int_a^b F(y+\alpha\eta, y'+\alpha\eta', x)\,dx \\ &= \int_a^b F(y,y',x)\,dx + \int_a^b \left(\frac{\partial F}{\partial y}\alpha\eta + \frac{\partial F}{\partial y'}\alpha\eta'\right)\,dx + \mathrm{O}(\alpha^2). \end{aligned}$$

With this form for $I(y,\alpha)$ the condition (22.3) implies that for all $\eta(x)$ we require

$$\delta I = \int_a^b \left(\frac{\partial F}{\partial y}\eta + \frac{\partial F}{\partial y'}\eta'\right)\,dx = 0,$$

where δI denotes the first-order variation in the value of I due to the variation (22.2) in the function $y(x)$. Integrating the second term by parts this becomes

$$\left[\eta\frac{\partial F}{\partial y'}\right]_a^b + \int_a^b \left[\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right]\eta(x)\,dx = 0. \tag{22.4}$$

In order to simplify the result we will assume, for the moment, that the end-points are fixed, i.e. not only a and b are given but also $y(a)$ and $y(b)$. This restriction means that we require $\eta(a) = \eta(b) = 0$, in which case the first term on the LHS of (22.4) equals zero at both end-points. Since (22.4) must be satisfied for arbitrary $\eta(x)$, it is easy to see that we require

$$\frac{\partial F}{\partial y} = \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right). \tag{22.5}$$

This is known as the *Euler–Lagrange* (EL) equation, and is a differential equation for $y(x)$, since the function F is known.

22.2 Special cases

In certain special cases a first integral of the EL equation can be obtained for a general form of F.

22.2.1 F does not contain y explicitly

In this case $\partial F/\partial y = 0$, and (22.5) can be integrated immediately giving

$$\frac{\partial F}{\partial y'} = \text{ constant.} \tag{22.6}$$

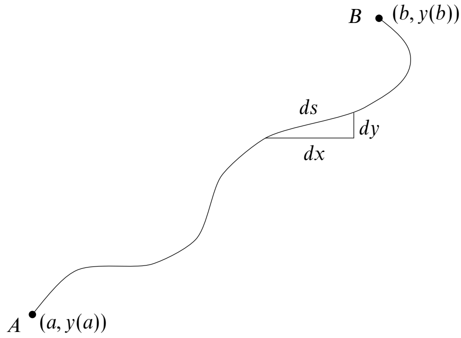


Figure 22.2 An arbitrary path between two fixed points.

►*Show that the shortest curve joining two points is a straight line.*

Let the two points be labelled A and B and have coordinates $(a, y(a))$ and $(b, y(b))$ respectively (see figure 22.2). Whatever the shape of the curve joining A to B, the length of an element of path ds is given by

$$ds = \left[(dx)^2 + (dy)^2\right]^{1/2} = (1 + y'^2)^{1/2} dx,$$

and hence the total path length along the curve is given by

$$L = \int_a^b (1 + y'^2)^{1/2}\, dx. \tag{22.7}$$

We must now apply the results of the previous section to determine that path which makes L stationary (clearly a minimum in this case). Since the integral does not contain y (or indeed x) explicitly, we may use (22.6) to obtain

$$k = \frac{\partial F}{\partial y'} = \frac{y'}{(1 + y'^2)^{1/2}}.$$

where k is a constant. This is easily rearranged and integrated to give

$$y = \frac{k}{(1 - k^2)^{1/2}} x + c,$$

which, as expected, is the equation of a straight line in the form $y = mx + c$, with $m = k/(1 - k^2)^{1/2}$. The value of m (or k) can be found by demanding that the straight line passes through the points A and B and is given by $m = [y(b) - y(a)]/(b - a)$. Substituting the equation of the straight line into (22.7) we find that, again as expected, the total path length is given by

$$L^2 = [y(b) - y(a)]^2 + (b - a)^2. \blacktriangleleft$$

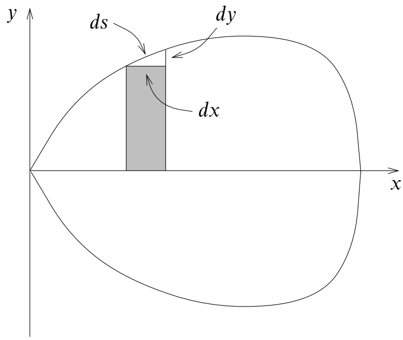


Figure 22.3 A convex closed curve that is symmetrical about the x-axis.

22.2.2 *F* does not contain *x* explicitly

In this case, multiplying the EL equation (22.5) by y' and using

$$\frac{d}{dx}\left(y'\frac{\partial F}{\partial y'}\right) = y'\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) + y''\frac{\partial F}{\partial y'}$$

we obtain

$$y'\frac{\partial F}{\partial y} + y''\frac{\partial F}{\partial y'} = \frac{d}{dx}\left(y'\frac{\partial F}{\partial y'}\right).$$

But since F is a function of y and y' only, and not explicitly of x, the LHS of this equation is just the total derivative of F, namely dF/dx. Hence, integrating we obtain

$$F - y'\frac{\partial F}{\partial y'} = \text{ constant.} \tag{22.8}$$

►*Find the closed convex curve of length l that encloses the greatest possible area.*

Without any loss of generality we can assume that the curve passes through the origin and can further suppose that it is symmetric with respect to the x-axis; this assumption is not essential. Using the distance s along the curve, measured from the origin, as the independent variable and y as the dependent one, we have the boundary conditions $y(0) = y(l/2) = 0$. The element of area shown in figure 22.3 is then given by

$$dA = y\,dx = y\left[(ds)^2 - (dy)^2\right]^{1/2},$$

and the total area by

$$A = 2\int_0^{l/2} y(1 - y'^2)^{1/2}\,ds; \tag{22.9}$$

here y' stands for dy/ds rather than dy/dx. Since the integrand does not contain s explicitly, we can use (22.8) to obtain a first integral of the EL equation for y, namely

$$y(1-y'^2)^{1/2} + yy'^2(1-y'^2)^{-1/2} = k,$$

where k is a constant. On rearranging this gives

$$ky' = \pm(k^2 - y^2)^{1/2},$$

which, using $y(0) = 0$, integrates to

$$y/k = \sin(s/k). \tag{22.10}$$

The other end-point, $y(l/2) = 0$, fixes the value of k as $l/(2\pi)$ to yield

$$y = \frac{l}{2\pi}\sin\frac{2\pi s}{l}.$$

From this we obtain $dy = \cos(2\pi s/l)\,ds$ and since $(ds)^2 = (dx)^2 + (dy)^2$ we find also that $dx = \pm\sin(2\pi s/l)\,ds$. This in turn can be integrated and, using $x(0) = 0$, gives x in terms of s as

$$x - \frac{l}{2\pi} = -\frac{l}{2\pi}\cos\frac{2\pi s}{l}.$$

We thus obtain the expected result that x and y lie on the circle of radius $l/(2\pi)$ given by

$$\left(x - \frac{l}{2\pi}\right)^2 + y^2 = \frac{l^2}{4\pi^2}.$$

Substituting the solution (22.10) into the expression for the total area (22.9), it is easily verified that $A = l^2/(4\pi)$. A much quicker derivation of this result is possible using plane polar coordinates. ◄

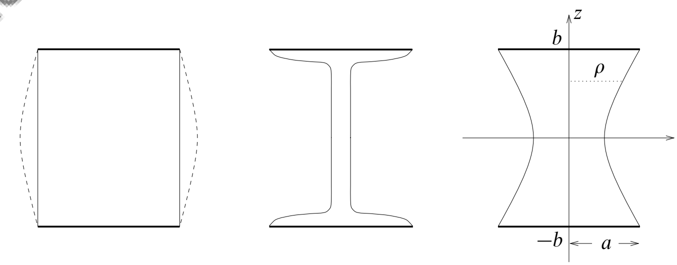
The previous two examples have been carried out in some detail, even though the answers are more easily obtained in other ways, expressly so that the method is transparent and the way in which it works can be filled in mentally at almost every step. The next example, however, does not have such an intuitively obvious solution.

►*Two rings, each of radius a, are placed parallel with their centres 2b apart and on a common normal. An open-ended axially symmetric soap film is formed between them (see figure 22.4). Find the shape assumed by the film.*

Creating the soap film requires an energy γ per unit area (numerically equal to the surface tension of the soap solution). So the stable shape of the soap film, i.e. the one that minimises the energy, will also be the one that minimises the surface area (neglecting gravitational effects).

It is obvious that any convex surface, shaped such as that shown as the broken line in figure 22.4(a), cannot be a minimum but it is not clear whether some shape intermediate between the cylinder shown by solid lines in (a), with area $4\pi ab$ (or twice this for the double surface of the film), and the form shown in (b), with area approximately $2\pi a^2$, will produce a lower total area than both of these extremes. If there is such a shape (e.g. that in figure 22.4(c)), then it will be that which is the best compromise between two requirements, the need to minimise the ring-to-ring distance measured on the film surface (a) and the need to minimise the average waist measurement of the surface (b).

We take cylindrical polar coordinates as in figure 22.4(c) and let the radius of the soap film at height z be $\rho(z)$ with $\rho(\pm b) = a$. Counting only one side of the film, the element of



(a) (b) (c)

Figure 22.4 Possible soap films between two parallel circular rings.

surface area between z and $z + dz$ is

$$dS = 2\pi\rho\left[(dz)^2 + (d\rho)^2\right]^{1/2},$$

so the total surface area is given by

$$S = 2\pi\int_{-b}^{b}\rho(1+\rho'^2)^{1/2}\,dz. \tag{22.11}$$

Since the integrand does not contain z explicitly, we can use (22.8) to obtain an equation for ρ that minimises S, i.e.

$$\rho(1+\rho'^2)^{1/2} - \rho\rho'^2(1+\rho'^2)^{-1/2} = k,$$

where k is a constant. Multiplying through by $(1+\rho'^2)^{1/2}$, rearranging to find an explicit expression for ρ' and integrating we find

$$\cosh^{-1}\frac{\rho}{k} = \frac{z}{k} + c.$$

where c is the constant of integration. Using the boundary conditions $\rho(\pm b) = a$, we require $c = 0$ and k such that $a/k = \cosh b/k$ (if b/a is too large, no such k can be found). Thus the curve that minimises the surface area is

$$\rho/k = \cosh(z/k),$$

and in profile the soap film is a catenary (see section 22.4) with the minimum distance from the axis equal to k. ◄

22.3 Some extensions

It is quite possible to relax many of the restrictions we have imposed so far. For example, we can allow end-points that are constrained to lie on given curves rather than being fixed, or we can consider problems with several dependent and/or independent variables or higher-order derivatives of the dependent variable. Each of these extensions is now discussed.