Waves, complex numbers and Fourier transforms

2

The theory of X-ray and neutron scattering relies heavily on the mathematics of waves. This chapter provides a tutorial introduction to the basic physical concepts, and the associated analytical tools, needed for an understanding of wave phenomena.

2.1 Sinusoidal waves

An everyday description of a wave would be a 'wiggle', or something that goes up-and-down as you move forward. The progression of the fluctuations could refer to changes in 'height' with respect to position at a fixed time, or with respect to time at a fixed position. Several examples of geometrical waves are shown in Fig. 2.1; they are unusual in that they have points where there are abrupt changes in the value of the function or its gradient. What they have in common with the more familiar *sinusoidal* variation of Fig. 2.2 is a regularly repeating pattern.

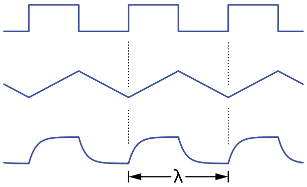


Fig. 2.1 Geometric examples of waves: square, triangular and exponential.

The *sine* and *cosine* curves of Fig. 2.2 are regarded as the archetypal waves, as they occur in many elementary physical situations; for example, the vibrations of an elastic string. Their smooth characteristics also make them amenable to analytical manipulation. An easy way of visualizing sinusoidal variations is to think about the projection of a circular motion onto the horizontal and vertical axes, as illustrated in Fig. 2.3.

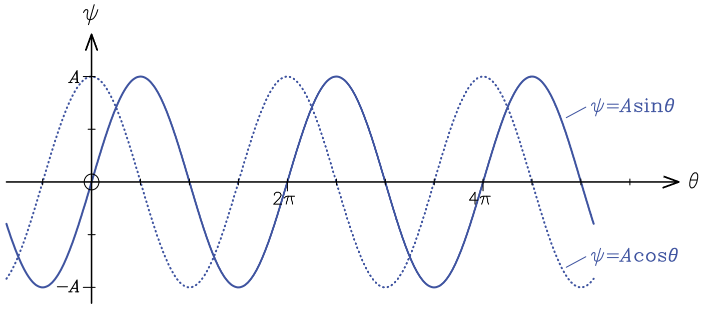


Fig. 2.2 The sinusoidal curves, or waves, $\psi = A\,\sin\theta$ and $\psi = A\cos\theta$.

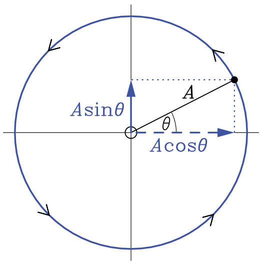


Fig. 2.3 The generation of sinusoidal variations through circular motion.

Elementary Scattering Theory, First Edition, D.S. Sivia, © D.S. Sivia 2011. Published in 2011 by Oxford University Press.

The two curves shown in Fig. 2.2 are identical apart from a lateral shift of $\pi/2$ radians, or $90°$: $\cos\theta = \sin(\theta + \pi/2)$. Hence, the general expression for a function of this type is

$$\psi = A\sin(\theta + \phi)\ , \tag{2.1}$$

where A is the *amplitude* of the wave and the angle ϕ, or the *phase*, controls its horizontal displacement with respect to $\sin\theta$. If the θ in eqn (2.1) varies linearly with position x, so that $\theta = kx$ where k is a constant, then we obtain a sinusoidal variation with respect to this physical coordinate:

$$\psi = A\sin(kx + \phi)\ . \tag{2.2}$$

Since the sine curve cycles around every 2π radians, the corresponding repeat distance, or *wavelength* λ, can be found from

$$\boxed{k = \frac{2\pi}{\lambda}}\ . \tag{2.3}$$

This is called the *wavenumber* and has SI units of $\mathrm{rad\,m^{-1}}$. Note that, as mentioned in Section 1.5, spectroscopists use the same term for $1/\lambda$ given in $\mathrm{cm^{-1}}$.

If the ϕ in eqn (2.2) itself varies linearly with time t, so that it can be written as $\phi = \phi_o - \omega t$ where ϕ_o and ω are constants, then we obtain the *travelling* wave

$$\psi = A\sin(kx - \omega t + \phi_o)\ . \tag{2.4}$$

That is to say, with $\omega > 0$, the sinusoidal variation in x moves steadily towards the right as time evolves; this is illustrated in Fig. 2.4. The crests and troughs of the translated wave will coincide with those of an earlier time after a duration T, called the *period*, given by

$$\boxed{\omega = \frac{2\pi}{T}}\ . \tag{2.5}$$

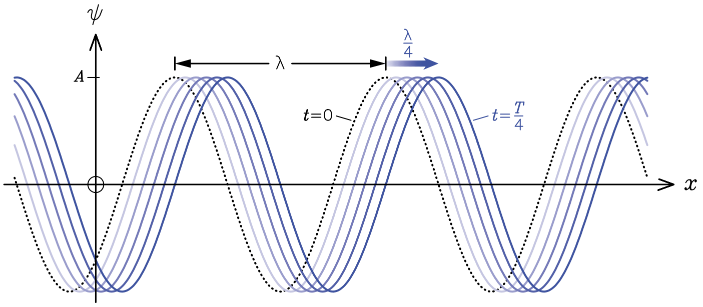


Fig. 2.4 The travelling wave of eqn (2.4) plotted as a function of x for several values of t, from zero to a quarter of the period.

The reciprocal of T, usually denoted by ν, is known as the *frequency* of the wave. It is related to its angular variant, ω, through

$$\omega \;=\; 2\pi\nu\,, \tag{2.6}$$

$$\nu \;=\; \frac{1}{T}$$

with ω specified in $\mathrm{rad\,s^{-1}}$ and ν in cycles per second or Hz (hertz). The speed of the wave, c, follows readily from the observation that it moves forward by a distance λ in a time T:

$$c \;=\; \frac{\lambda}{T} \;=\; \frac{\omega}{k} \;=\; \nu\lambda\,, \tag{2.7}$$

in agreement with the result quoted in eqn (1.16).

2.1.1 The direction of propagation

A negative prefactor was chosen for the ωt term in eqn (2.4) so that the wave would travel in the positive x direction; the opposite sign, $\psi = A\sin(kx+\omega t+\phi_{\mathrm{o}})$, gives a wave that moves backwards. In fact, a reversal also occurs with $\psi = A\sin(-kx-\omega t+\phi_{\mathrm{o}})$. Is there any reason for preferring one of these alternatives over the other to define the sense of the progression?

Conceptually, it would make more sense to associate the change of sign with the spatial term, rather than the temporal factor, because we are concerned with an orientation. This line of thought leads to the following generalization to accommodate fully the directional aspect of waves:

$$\psi \;=\; A\sin(\mathbf{k}\bullet\mathbf{r}-\omega t+\phi_{\mathrm{o}})\,, \tag{2.8}$$

$$\begin{aligned} \mathbf{r} &= (\,x\,,\;y\,,\;z\,) \\ \mathbf{k} &= (k_x, k_y, k_z) \\ \mathbf{k}\bullet\mathbf{r} &= k_x x + k_y y + k_z z \end{aligned}$$

where the bold script $\mathbf{k}$ and $\mathbf{r}$ are *vectors*, and the dot between them indicates their 'scalar multiplication'. The vector $\mathbf{r}$ denotes a general position in space, with coordinates x, y and z, but what do the three components, k_x, k_y and k_z, of the *wavevector* $\mathbf{k}$ represent? Its magnitude, or *modulus*, $|\mathbf{k}| = k$ is the familiar wavenumber of eqn (2.3), and its orientation indicates the direction of propagation. For a wave travelling along the x direction, with $k_y = k_z = 0$, the scalar product $\mathbf{k}\bullet\mathbf{r} = k_x x$ where $k_x = k$ for a forwards progression and $k_x = -k$ for the reverse.

$$\begin{aligned} |\mathbf{k}|^2 &= k^2 \\ &= k_x^2 + k_y^2 + k_z^2 \end{aligned}$$

Since $\mathbf{r}$ and $\mathbf{k}$ are generally three-dimensional vectors, the wave of eqn (2.8) tends to be a function of x, y, z and t. As such, it represents a travelling 'plane wave' rather than a moving oscillation on a string. That is to say ψ, which could be the air pressure in a sound wave, is uniform in planes perpendicular to $\mathbf{k}$, but its value varies sinusoidally with time in the direction of the wavevector in accordance with the wavelength of eqn (2.3), the period of eqn (2.5) and the speed in eqn (2.7). The situation is illustrated for the two-dimensional analogue in Fig. 2.5.

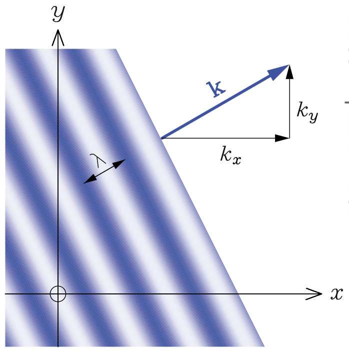
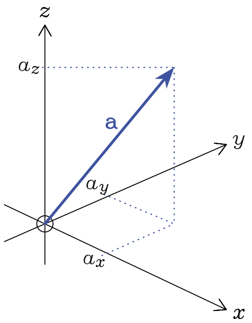


Fig. 2.5 The geometry of a plane wave.

Magnitudes, directions and vectors

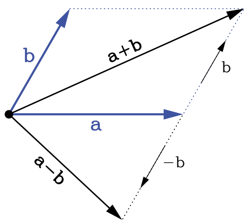


Quantities that have both a magnitude and a direction, such as a force, are called vectors. Unlike scalars, which only have a 'size', they cannot be quantified by a single number. They are defined by coordinates, or an array of numbers giving displacements with respect to a set of reference or *basis* axes. In the most common case of an x, y and z or *Cartesian* system, the vectors **a** and **b** can be written as

$$\mathbf{a} = (a_x\,, a_y\,, a_z) \quad \text{and} \quad \mathbf{b} = (b_x\,, b_y\,, b_z)\ .$$

Addition and subtraction are straightforward, in that the corresponding components are just combined separately:

$$\mathbf{a}+\mathbf{b} = (a_x+b_x\,,\ a_y+b_y\,,\ a_z+b_z)\ ,$$



with all the pluses replaced by minuses for a take away. The multiplication of a vector by a scalar, μ say, is also easy,

$$\mu\mathbf{a} = (\mu a_x\,, \mu a_y\,, \mu a_z)\ ,$$

and yields a vector with the original direction but an appropriately scaled length. The modulus, magnitude or length of a vector is given by *Pythagoras'* theorem; it's one for a *unit* or *normalized* vector.

A vector can be multiplied by another in two different ways. The first is a 'dot' or *scalar product*, which is a sum of the products of corresponding elements:

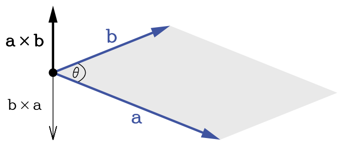
$$\mathbf{a}\bullet\mathbf{b} = a_x b_x + a_y b_y + a_z b_z \tag{2.9}$$

$$|\mathbf{a}|^2 = \mathbf{a}\bullet\mathbf{a} = a_x^2 + a_y^2 + a_z^2$$

$$\mathbf{a}\bullet\mathbf{b} = |\mathbf{a}|\,|\mathbf{b}|\cos\theta$$

and is geometrically the modulus of **a** times the modulus of **b** times the cosine of the angle between them. Vectors of non-zero length are perpendicular, or *orthogonal*, to each other if their dot product is zero; if they are also of unit length, they are said to be *orthonormal*.

Vectors can also be multiplied by a 'cross' or *vector product*. This is a bit more complicated since the result is a vector:



$$\mathbf{a}\times\mathbf{b} = (a_y b_z - b_y a_z\,,\ a_z b_x - b_z a_x\,,\ a_x b_y - b_x a_y)\ . \tag{2.10}$$

$$|\mathbf{a}\times\mathbf{b}| = |\mathbf{a}|\,|\mathbf{b}|\sin\theta$$

Geometrically, its magnitude is equal to the modulus of **a** times the modulus of **b** times the sine of the angle between them; this is also the area of the related parallelogram. The direction of the cross product is perpendicular to both **a** and **b**, and given by the 'right-hand screw rule': if the curl of the right-hand fingers indicates the sense of rotation needed to go from **a** to **b**, then the direction is given by the out-stretched thumb.

The physical meaning of a dot product holds irrespective of the dimensionality of the vectors and eqn (2.9) generalizes in an obvious way. The same is not true of a cross product, which is specific to a space of three dimensions (as considered here). The scalar product is also *symmetric* with respect to an interchange of **a** and **b** whereas the vector product is *antisymmetric*: the latter changes sign but the former does not. Division by a vector is not defined and must never be performed.

$$\mathbf{a}\bullet\mathbf{b} = \mathbf{b}\bullet\mathbf{a}$$

$$\mathbf{a}\times\mathbf{b} = -\,\mathbf{b}\times\mathbf{a}$$

2.1.2 The principle of superposition

A central feature of waves is that they pass through each other unaffected and, where overlapped, give a net result that is the sum of the individual contributions. This principle of *superposition* lies at the heart of scattering theory. Here we illustrate it with a couple of one-dimensional examples involving the combination of just two sinusoidal waves.

Consider first the case where the waves are identical but travelling in opposite directions. With the simplifying assignments that $A=1$ and $\phi_o=0$ in eqn (2.4), to reduce the algebraic clutter, the principle of superposition yields

$$\begin{aligned}\psi &= \sin(kx-\omega t)+\sin(-kx-\omega t)\\ &= -2\,\sin(\omega t)\cos(kx)\ ,\end{aligned}$$

$$\sin X+\sin Y = 2\,\sin\left(\tfrac{X+Y}{2}\right)\cos\left(\tfrac{X-Y}{2}\right)$$

where the second line follows from a trigonometric ‘factor formula’ and the antisymmetric properties of the sine function. This is called a *stationary* wave (Fig. 2.6), because there is no movement with time along the x direction. The separation of ψ into a product of spatial and temporal terms results in a purely ‘up-and-down’ oscillation, at a frequency of ω, with an amplitude that varies sinusoidally with wavelength $\lambda=2\pi/k$. The locations at which the amplitude is zero are called *nodes*.

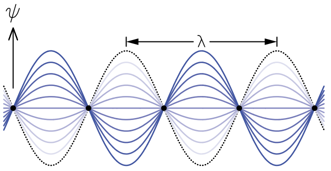


Fig. 2.6 A stationary wave, plotted as a function of x for several values of t.

As a second example, consider two waves travelling in the same direction with equal amplitudes but slightly different wavelengths and frequencies: $k\pm\Delta k$ and $\omega\pm\Delta\omega$, where the Δ-terms represent small departures from the average k and ω. With the simplification that $A{=}1$ and $\phi_o{=}0$, as before, ψ is now a product of two travelling waves:

$$\begin{aligned}\psi &= \sin\Big([k+\Delta k]x-[\omega+\Delta\omega]t\Big)+\sin\Big([k-\Delta k]x-[\omega-\Delta\omega]t\Big)\\ &= -2\,\sin(kx-\omega t)\cos\big(\Delta k\,x-\Delta\omega\,t\big)\ ,\end{aligned}$$

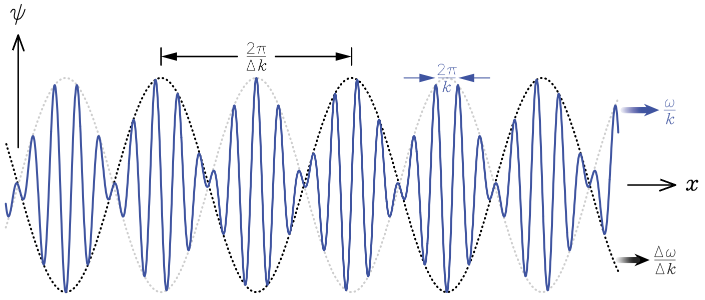


Fig. 2.7 The slowly varying ‘beating’ modulation, of wavelength $2\pi/\Delta k$, propagates with a speed of $\Delta\omega/\Delta k$, whereas the finer structure inside the envelope has the properties of the average wavelength and frequency, ω and k.

and is illustrated in Fig. 2.7. The amplitude of a sinusoid with the mean wavelength of $2\pi/k$, propagating with a speed ω/k, is modulated by a slowly varying envelope of wavelength $2\pi/\Delta k$, moving with a speed of $\Delta\omega/\Delta k$. This is the origin of the *beating* that is heard when neighbouring musical notes are played together: the sound becomes periodically louder and quieter.

Although we have only considered the combination of two similar waves, its generalization to the sum of many such components results in the formation of *wavepackets*; the beating modulation of Fig. 2.7 is just the most elementary example. The shape of the wavepacket will be preserved on propagation if all its constituents travel with the same speed c, when

$$\frac{\omega}{k} = c \qquad \text{and} \qquad \frac{\mathrm{d}\omega}{\mathrm{d}k} = c\,.$$

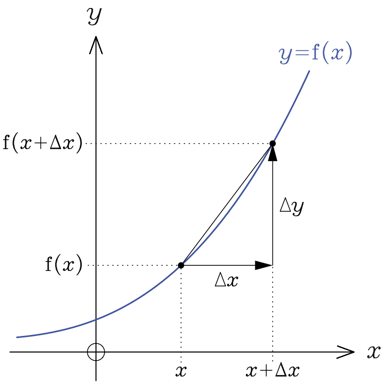
From Sivia and Rawlings (1999), *Foundations of Science Mathematics*, Oxford Chemistry Primers Series, **77**.

Gradients, rates of change and differentiation

The relationship between two quantities, x and y say, can be visualized with the aid of a graph. While the intersections of the associated curve with the x and y axes may be of interest, it is often more important to know the slope at any given point; that is, how quickly y increases, or decreases, as x changes, and vice versa. This issue is at the heart of the topic of *differentiation*, and the related rules and formulae are simply ways of calculating the gradient algebraically.

Let us begin with a precise definition of what is meant by the slope of a curve. Suppose that y is related to x through some function called ‘f’, usually written as $y = \mathrm{f}(x)$, so that $\mathrm{f}(x) = m\,x + c$ for a general straight line, and $\mathrm{f}(x) = \sin(x)$ for a sinusoidal variation, and so on. Then, if the horizontal coordinate changes from x to $x+\Delta x$, where Δx represents a small increment, the value of y is altered from $\mathrm{f}(x)$ to $\mathrm{f}(x+\Delta x)$. The gradient, at a point x, is defined to be the ratio of the change in the vertical coordinate, Δy, to that of the horizontal increment, as Δx becomes vanishingly small. This can be stated formally as

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} = \lim_{\Delta x\to 0}\frac{\mathrm{f}(x+\Delta x)-\mathrm{f}(x)}{\Delta x}\,, \tag{2.11}$$



where $\mathrm{d}y/\mathrm{d}x$ is known as the *derivative*, or *differential coefficient*, and is pronounced ‘dy-by-dx’. The tendency of $\Delta x \to 0$ has to be approached gradually to ascertain the *limiting* value of the ratio $\Delta y/\Delta x$, as both increments are individually equal to zero when the condition is met. Strictly speaking, we should check that the same value of $\mathrm{d}y/\mathrm{d}x$ is obtained whether Δx is positive or negative, but this is assured as long as the curve $y = \mathrm{f}(x)$ is ‘smooth’; inconsistencies will arise if kinks and sudden breaks (or discontinuities) are present, and the function is said to be *non-differentiable* at those points.

$$y' = \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}(y) = \mathrm{f}'(x)$$

$$y'' = \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right) = \mathrm{f}''(x)$$

If the speed of the sinusoidal waves varies with their wavelength, because the frequency does not happen to be directly proportional to the wavenumber in the medium of interest, then the wavepacket will change with time. This phenomenon is called *dispersion*, and the relationship between ω and k which determines the nature of the 'spreading',

$$\omega = \omega(k) \,,$$

is called the 'dispersion relation' or the 'dispersion curve' (Fig. 2.8). For the non-dispersive case, $\omega = c\,k$, there is a unique speed, c, associated with the propagation of a wavepacket. The ratio ω/k and the derivative $\mathrm{d}\omega/\mathrm{d}k$ still yield useful characteristic speeds, however, when there is a dominant contribution from sinusoidal waves around a particular wavelength:

$$v_\phi = \frac{\omega}{k} \qquad \text{and} \qquad v_\mathrm{g} = \frac{\mathrm{d}\omega}{\mathrm{d}k} \,, \tag{2.12}$$

where v_ϕ is called the *phase velocity*, and gives the rate at which the crests and troughs of the local wavefront move, and v_g is the *group velocity*, which indicates how fast the envelope of the wavepacket travels.

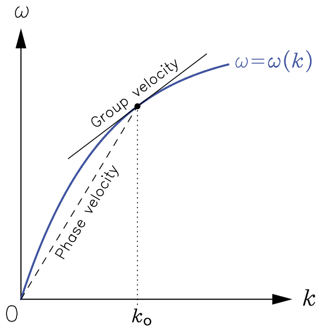


Fig. 2.8 A dispersion curve, and the related phase and group velocities for waves in the neighbourhood of k_o.

2.2 Complex numbers

The analysis of wave phenomena is aided greatly by the use of *complex numbers*. In particular, by a result which links an exponential to sines and cosines:

$$\boxed{\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta} \,, \tag{2.13}$$

where $\mathrm{i}^2 = -1$. Since complex numbers play a central role in theoretical work, we will devote a few pages to them; as with most of the mathematical background given in this book, the material is based on Sivia and Rawlings (1999).

2.2.1 Definition

If any number, integer or fraction, positive or negative, is multiplied by itself, then the result is always greater than, or equal, to zero. What, then, is the square root of -9? To address this question we need to invent an *imaginary* number, usually denoted by 'i', whose square is defined to be negative:

$$\mathrm{i}^2 = -1 \,. \tag{2.14}$$

A *real* number, say b (where $b^2 \geqslant 0$), times i is also imaginary; it's just b times bigger than i. If a is also an ordinary number, then the sum z of a and $\mathrm{i}b$,

$$\sqrt{-9} = \pm 3\mathrm{i}$$

$$z = a + \mathrm{i}b \,, \tag{2.15}$$

is known as a ‘complex’ number; this does not indicate an intrinsic difficulty with the concept, but highlights the hybrid nature of the entity. It consists of both a real part and an imaginary one:

$$\mathcal{R}e\,\{z\} = a \qquad \text{and} \qquad \mathcal{I}m\,\{z\} = b\,. \tag{2.16}$$

It may seem odd that $\mathcal{I}m\,\{z\}$ is b rather than $\mathrm{i}b$, but this is because it represents the size of the imaginary component.

2.2.2 Basic algebra

To add or subtract complex numbers, we simply add or subtract the real and imaginary parts separately:

$$a + \mathrm{i}b \pm (c + \mathrm{i}d) = a \pm c + \mathrm{i}(b \pm d)\,, \tag{2.17}$$

$$1 + 2\mathrm{i} - (5 - \mathrm{i}) = -4 + 3\mathrm{i}$$

where a, b, c and d are real. The usual rules of algebra apply for brackets and multiplication, except that every occurrence of i^2 is replaced by -1. Thus, it's easy to show that the product of $a + \mathrm{i}b$ and $c + \mathrm{i}d$ is given by

$$(a + \mathrm{i}b)\,(c + \mathrm{i}d) = ac - bd + \mathrm{i}(ad + bc)\,, \tag{2.18}$$

$$(1 + 2\mathrm{i})\,(3 - \mathrm{i}) = 5 + 5\mathrm{i}$$

since $\mathrm{i}^2 bd = -bd$. Division involves the use of a *complex conjugate*, so let us consider this first.

The conjugate of a complex number z, denoted by z^*, is defined to have the same real part but the opposite imaginary component; that is, $Re\{z^*\} = Re\{z\}$ and $\mathcal{I}m\,\{z^*\} = -\mathcal{I}m\,\{z\}$. In terms of eqn (2.15), therefore,

$$z^* = (a + \mathrm{i}b)^* = a - \mathrm{i}b\,. \tag{2.19}$$

Hence, complex numbers and their conjugates satisfy the following relationships:

$$\begin{aligned} z + z^* &= \;\; 2a &&= 2\,\mathcal{R}e\,\{z\} \\ z - z^* &= \;\; 2\mathrm{i}b &&= 2\,\mathrm{i}\,\mathcal{I}m\,\{z\} \\ z\,z^* &= a^2 + b^2 &&= |z|^2 \end{aligned} \tag{2.20}$$

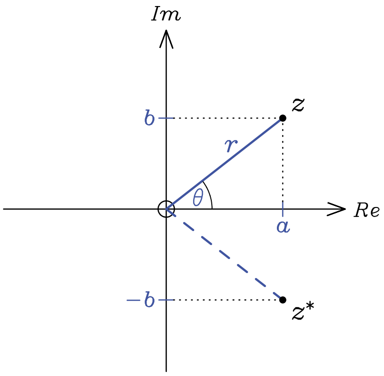
We will come to the meaning of $|z|$ shortly, but the important point about eqn (2.20) is that the product $z\,z^*$ is a real number. This feature enables us to calculate the real and imaginary part of the ratio of two complex numbers by multiplying both the top and bottom by the conjugate of the denominator

$$\frac{a + \mathrm{i}b}{c + \mathrm{i}d} = \frac{a + \mathrm{i}b}{c + \mathrm{i}d} \times \frac{c - \mathrm{i}d}{c - \mathrm{i}d} = \frac{ac + bd + \mathrm{i}(bc - ad)}{c^2 + d^2}\,. \tag{2.21}$$

To evaluate the ratio $(1{+}2\mathrm{i})/(3{-}\mathrm{i})$, for example, we multiply it by unity in the form $(3{+}\mathrm{i})/(3{+}\mathrm{i})$; this gives a real denominator of 10, and a complex numerator of $1{+}7\mathrm{i}$. Hence the result is $1/10 + \mathrm{i}7/10$.

2.2.3 The Argand diagram

So far we have considered complex numbers from an algebraic point of view; it is often helpful to think of them in geometrical terms. This is easily done with the aid of an *Argand diagram* where the horizontal, or x, axis of a graph is seen as representing the real part of a complex number, and the vertical, or y, axis gives the imaginary component. Thus the point with (x, y) coordinates (a, b) corresponds to the complex number $a + \mathrm{i}b$. Its conjugate z^* is a reflection in the real axis. An alternative way of specifying the location of a point on a graph is through its distance r from the origin, and the anticlockwise angle θ that this 'radius' makes with the (positive) real axis. In this system r is known as the *modulus*, magnitude or amplitude of z; θ is called its *argument* or phase.



The quantities r, θ, a and b in the Argand diagram are related through elementary trigonometry by

$$a = r\cos\theta \qquad \text{and} \qquad b = r\sin\theta\,, \tag{2.22}$$

or, in the reverse sense, by

$$r^2 = a^2 + b^2 \qquad \text{and} \qquad \theta = \tan^{-1}(b/a)\,. \tag{2.23}$$

A comparison between eqns (2.20) and (2.23) shows that $z\,z^* = r^2$, where $r = |z|$ is the modulus of the complex number. The second part of eqn (2.23) needs qualification because there is an ambiguity of $180°$ with $\tan^{-1}(b/a)$. The arctangent of unity, for example, could be either $45°$ or $-135°$. For complete consistency with eqn (2.22), $0<\theta<\pi$ (radians) for $b>0$ and $-\pi<\theta<0$ for $b<0$; θ is zero for $a>0$, and $\pm\pi$ for $a<0$, if $b=0$. It is also worth remembering that θ is only defined to within a factor of 2π, because we could add (or subtract) any integer number of $360°$ to it and obtain the same point in the Argand diagram.

2.2.4 The imaginary exponential

Perhaps the most important result in complex analysis concerns the exponential of an imaginary number:

$$\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta\,, \tag{2.24}$$

where θ is in radians. This equation can be verified by substituting $x = \mathrm{i}\theta$ in the *Taylor series* expansion for e^x, and collecting the odd and even powers of θ separately; remembering that $\mathrm{i}^2 = -1$, a comparison with the Taylor series for $\sin\theta$ and $\cos\theta$ yields eqn (2.24). The product of r and $\mathrm{e}^{\mathrm{i}\theta}$ allows a complex number to be expressed in a very compact form in terms of its modulus and argument:

$$\mathrm{e}^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots$$

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots$$

$$z = a + \mathrm{i}b = r(\cos\theta + \mathrm{i}\sin\theta) = r\,\mathrm{e}^{\mathrm{i}\theta}\,, \tag{2.25}$$

where a, b, r and θ are related through eqns (2.22) and (2.23). As can be seen from the Argand diagram, and verified by the symmetry properties of sines and cosines, its conjugate entails the replacement of θ with $-\theta$:

$$z^* \;=\; a - \mathrm{i}b \;=\; r\,(\cos\theta - \mathrm{i}\sin\theta) \;=\; r\,\mathrm{e}^{-\mathrm{i}\theta}\,, \tag{2.26}$$

from which the result $z\,z^* = r^2$ follows immediately.

Although the exponential form of a complex number is very useful when dealing with roots and logarithms, and provides a valuable insight into products and quotients, our interest here is in its relationship with waves. This hinges on eqn (2.24), which enables eqn (2.8) to be written as the imaginary part of

$$\psi \;=\; A\,\mathrm{e}^{\mathrm{i}(\mathbf{k}\bullet\mathbf{r}\,-\,\omega t)}\,, \tag{2.27}$$

where A is now a complex number whose modulus and argument give the amplitude and phase offset of the wave, respectively:

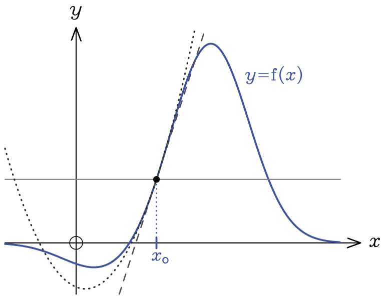
$$A \;=\; |A|\,\mathrm{e}^{\mathrm{i}\,\phi_\mathrm{o}}\,. \tag{2.28}$$

The real part of eqn (2.27) also represents the same wave, apart from a difference of $90°$ in the value of ϕ_o.

Taylor series

From Sivia and Rawlings (1999), *Foundations of Science Mathematics*, Oxford Chemistry Primers Series, **77**.

When dealing with a complicated function, it can be useful to approximate it with one of a simpler form. While the latter may not represent a complete and accurate description of the situation at hand, it frequently provides the only means of making analytical progress. There are many approximations that could be used, of course, but it is the one that captures the salient features that is most helpful. A Taylor series is appropriate when our principal interest lies in the behaviour of a function in the neighbourhood of a particular point.



Consider the curve $y = \mathrm{f}(x)$. The crudest approximation to this function is a horizontal line $y = a_0$, where a_0 is a constant; if $a_0 = \mathrm{f}(x_\mathrm{o})$, then it will even be correct at $x = x_\mathrm{o}$. A better approximation would be a sloping line $y = a_0 + a_1(x - x_\mathrm{o})$, where the coefficient a_1 allows for a non-zero gradient. Continuing along this path, we could add a quadratic (or curvature) term $a_2(x - x_\mathrm{o})^2$, a cubic contribution $a_3(x - x_\mathrm{o})^3$, and so on, to gain further improvements. Thus, a function $\mathrm{f}(x)$ can be approximated about the point x_o by using a *polynomial* expansion:

$$\mathrm{f}(x) \;\approx\; a_0 + a_1(x - x_\mathrm{o}) + a_2(x - x_\mathrm{o})^2 + a_3(x - x_\mathrm{o})^3 + \cdots. \tag{2.29}$$

$$a_n \;=\; \frac{1}{n!}\,\frac{\mathrm{d}^n f}{\mathrm{d}x^n}\bigg|_{x_\mathrm{o}}$$

This is the essence of a Taylor series. Its advantage is that the right-hand side of eqn (2.29) is usually easier to calculate, differentiate, integrate, and generally manipulate, than the expression on the left. The case of $x_\mathrm{o} = 0$, when the Taylor series simplifies, is called a *Maclaurin series*.

The benefit of using eqn (2.27) over (2.8) in wave analysis is that exponentials are easier to deal with mathematically than sinusoids; multiplication, differentiation and *integration*, for example, are more straightforward. As an illustration of this advantage, let's derive the 'compound angle' formulae for sines and cosines with complex numbers. Starting with the rule of eqn (1.2) for combining powers,

$$\mathrm{e}^{\mathrm{i}(\alpha+\beta)} = \mathrm{e}^{\mathrm{i}\alpha}\,\mathrm{e}^{\mathrm{i}\beta},$$

and expanding the exponentials with eqn (2.24),

$$\cos(\alpha+\beta) + \mathrm{i}\sin(\alpha+\beta) = (\cos\alpha + \mathrm{i}\sin\alpha)\,(\cos\beta + \mathrm{i}\sin\beta)\,,$$

the equating of the real and imaginary parts on the left- and right-hand sides yields the desired results:

$$\cos(\alpha+\beta) = \cos\alpha\,\cos\beta - \sin\alpha\,\sin\beta\,, \tag{2.30}$$

$$\sin(\alpha+\beta) = \sin\alpha\,\cos\beta + \cos\alpha\,\sin\beta\,. \tag{2.31}$$

As well as being the real and imaginary parts of $\exp(\mathrm{i}\theta)$, sines and cosines can also be expressed as

$$\cos\theta = \frac{\mathrm{e}^{\mathrm{i}\theta} + \mathrm{e}^{-\mathrm{i}\theta}}{2} \quad \text{and} \quad \sin\theta = \frac{\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{-\mathrm{i}\theta}}{2\mathrm{i}}\,, \tag{2.32}$$

which follow from the addition and subtraction, respectively, of eqn (2.24) with its complex conjugate.

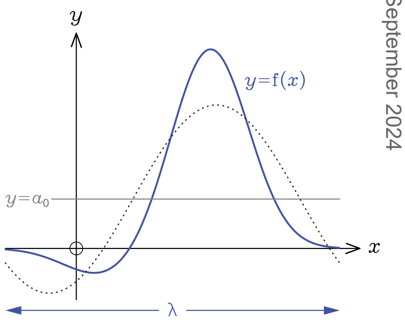
2.3 Fourier series

Let's begin wave analysis by considering how periodic signals, such as those in Fig. 2.1, can be decomposed into the sum of sinusoids. Suppose that the function $\mathrm{f}(x)$ repeats itself after a 'distance' of λ, so that

$$\mathrm{f}(x) = \mathrm{f}(x+\lambda)\,. \tag{2.33}$$

This has the same periodicity as sines and cosines of wavenumber $k = 2\pi/\lambda$. A simple approximation to $\mathrm{f}(x)$, which matches its wavelength, is therefore

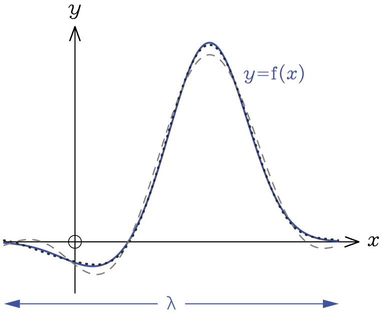
$$\mathrm{f}(x) \approx a_0 + a_1\cos(kx) + b_1\sin(kx)\,, \tag{2.34}$$



where a_0, a_1 and b_1 are constants whose values need to be selected in some way. The crudest assignment would be to set both a_1 and b_1 equal to zero, giving an invariant $\mathrm{f}(x) \approx a_0$, but the *linear* combination of $\sin(kx)$ and $\cos(kx)$ allows for a sinusoidal variation with the correct period and an appropriate amplitude and phase:

$$a\cos(kx) + b\sin(kx) = A\sin(kx+\phi)\,,$$

where $a = A\sin\phi$ and $b = A\cos\phi$, in accordance with eqn (2.31).



The sines and cosines of $2kx$, $3kx$, $4kx$, and so on, also satisfy the periodicity of eqn (2.33); they just go through several, or many, complete cycles in the interval λ. We can obtain a better approximation to $\mathrm{f}(x)$, therefore, by including contributions from these higher-order terms:

$$\begin{aligned}\mathrm{f}(x) \;\approx\; & a_0 + a_1\cos(kx) + a_2\cos(2kx) + a_3\cos(3kx) + \cdots \\ & + b_1\sin(kx) + b_2\sin(2kx) + b_3\sin(3kx) + \cdots\end{aligned} \tag{2.35}$$

This expansion is called a *Fourier series*, and eqn (2.34) is simply the first-order version of it which contains only the lowest, or fundamental, harmonic.

We will come to the evaluation of the coefficients a_n and b_n, for integer n, shortly but note that one of the sets goes to zero if $\mathrm{f}(x)$ possesses a symmetry about the y-axis:

$$\sin(\theta) = -\sin(-\theta)$$
$$\cos(\theta) = \cos(-\theta)$$

$$\mathrm{f}(x) = \begin{cases} \mathrm{f}(-x) & \Longrightarrow\; b_n = 0 \\ -\mathrm{f}(-x) & \Longrightarrow\; a_n = 0 \end{cases} \tag{2.36}$$

because sines and cosines are *odd* and *even* functions, respectively. The generalization of eqn (2.35) explains why the invariant term is designated as a_0, and why there is no corresponding b_0 (apart from its general redundancy): they are the coefficients of $\cos(0)=1$ and $\sin(0)=0$, with the b_0 being unnecessary since it adds nothing to the Fourier series.

2.3.1 Orthogonality and the Fourier coefficients

A prescription for the a_n and b_n in eqn (2.35) presents itself once we realize that the related sine and cosine functions are *orthogonal*. By this we mean that the *integral* of the product of any two over the interval of the period λ will be zero, unless they happen to be exactly the same functions:

$$\int_0^{\lambda} \sin(mkx)\sin(nkx)\,\mathrm{d}x \;=\; \begin{cases} 0 & \text{if } m \neq n\,, \\ \frac{\lambda}{2} & \text{if } m = n\,, \end{cases} \tag{2.37}$$

with an identical expression for $\cos(mkx)\cos(nkx)$, but $n \neq 0$, and

$$\int_0^{\lambda} \sin(mkx)\cos(nkx)\,\mathrm{d}x \;=\; 0\,. \tag{2.38}$$

Although these sines and cosines aren't perpendicular in a geometrical sense, this type of integral is the functional analogue of a dot product which is zero for orthogonal vectors.

If we multiply eqn (2.35) through by one of the sine or cosine functions, $\sin(mkx)$ or $\cos(mkx)$, and integrate the resultant products over the period λ, then all but one of the terms on the right-hand side will be zero due to eqns (2.37) and (2.38). The surviving $m = n$ contributions yield the formulae for the Fourier coefficients:

$$a_n = \frac{2}{\lambda}\int_0^\lambda \mathrm{f}(x)\cos(nkx)\,\mathrm{d}x \quad \text{and} \quad b_n = \frac{2}{\lambda}\int_0^\lambda \mathrm{f}(x)\sin(nkx)\,\mathrm{d}x \tag{2.39}$$

for $n = 1, 2, 3, \ldots$, from which eqn (2.36) can be verified. If eqn (2.35) is integrated over the period λ as it stands, then the constant a_0 is seen to be the average value of $\mathrm{f}(x)$:

$$a_0 = \frac{1}{\lambda}\int_0^\lambda \mathrm{f}(x)\,\mathrm{d}x\ . \tag{2.40}$$

$$\int_0^\lambda \sin(nkx)\,\mathrm{d}x = 0$$

$$\int_0^\lambda \mathrm{d}x = \lambda$$

Cumulative properties and integrals

While differentiation is concerned with the slope of $y = \mathrm{f}(x)$, integration deals with the 'area under the curve'. This relates to the average and cumulative behaviour of y, over some range in x.

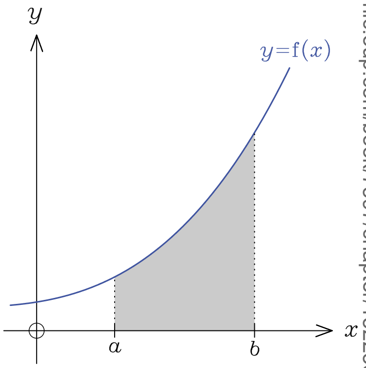
To set up a definition of an integral, consider the region bounded by the straight lines $x = a$, $x = b$ and $y = 0$, and the curve $y = \mathrm{f}(x)$. The size of the enclosure can be estimated by approximating it as a whole series of narrow vertical strips, and adding together the areas of these contiguous rectangular blocks. If the x-axis between a and b is divided into N equal intervals, then the width of each strip is given by $\Delta x = (b-a)/N$; the corresponding heights of the thin blocks are equal to the values of the function $\mathrm{f}(x)$ at their central positions. In other words, the area of the j^{th} strip, which is at $x = x_j$ and of height $y = \mathrm{f}(x_j)$, is $\mathrm{f}(x_j)\,\Delta x$; the index j ranges from 1 to N, of course, with $x_1 = a + \Delta x/2$ and $x_N = b - \Delta x/2$. As N tends to infinity, $\Delta x \to 0$ and the approximation to the area under the curve becomes ever more accurate. This limiting form of the summation procedure defines an integral

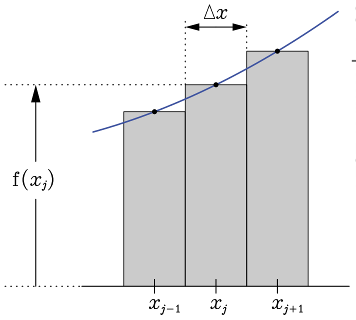
$$\int_a^b y\ \mathrm{d}x \ = \ \int_a^b \mathrm{f}(x)\,\mathrm{d}x \ = \ \lim_{N\to\infty}\sum_{j=1}^{N} \mathrm{f}(x_j)\,\Delta x\ , \tag{2.41}$$

where the symbol $\int \mathrm{d}x$ is read as the 'integral, from a to b, with respect to x'. The use of the term 'area' in the above discussion needs some qualification, in that it can be negative; this is because the 'height' of a strip $\mathrm{f}(x_j) < 0$ whenever the curve $y = \mathrm{f}(x)$ lies below the x-axis (and even the 'width' $\Delta x < 0$ if $b < a$).

Although an integral is defined as the limiting form of a summation, it is usually calculated analytically by noting that 'integration is the reverse of differentiation'. While this may not be obvious, it is easily illustrated with an example from everyday kinematics: the distance travelled by a car (say) is the integral of the speed with respect to time, and speed is the rate of change of distance with time (a derivative).

From Sivia and Rawlings (1999), *Foundations of Science Mathematics*, Oxford Chemistry Primers Series, **77**.





2.3.2 The complex Fourier series

The Fourier series of eqn (2.35) can be written in a very compact form by using complex numbers:

$$\mathrm{f}(x) \;=\; \sum_{n=-\infty}^{\infty} c_n\, \mathrm{e}^{\mathrm{i}nkx}\,, \tag{2.42}$$

where the Σ stands for a summation over integer values of n, from $-\infty$ to ∞, and the approximation has been replaced by an equality to denote a definition. The right-hand side of eqn (2.42) will yield a real function, $\mathrm{f}(x)$, as long as the complex coefficients c_n satisfy the conjugacy condition

$$c_{-n} \;=\; c_n^*\,. \tag{2.43}$$

This follows from eqn (2.20) because the contribution to the sum from pairs of positive and negative values of n of equal magnitudes will then be

$$\left(c_n\, \mathrm{e}^{\mathrm{i}nkx}\right)^* \;=\; c_n^*\, \mathrm{e}^{-\mathrm{i}nkx}$$

$$\begin{aligned} c_n\, \mathrm{e}^{\mathrm{i}nkx} + c_{-n}\, \mathrm{e}^{-\mathrm{i}nkx} \;&=\; 2\,\mathcal{R}e\left\{c_n\, \mathrm{e}^{\mathrm{i}nkx}\right\} \\ &=\; a_n \cos(nkx) + b_n \sin(nkx)\;, \end{aligned}$$

for $n \neq 0$, where we have substituted $2\,c_n = a_n - \mathrm{i}b_n$ in the second line to obtain consistency with eqn (2.35). In fact, the formula for the complex coefficients is simply

$$c_n \;=\; \tfrac{1}{\lambda}\int_0^{\lambda} \mathrm{f}(x)\, \mathrm{e}^{-\mathrm{i}nkx}\, \mathrm{d}x\;, \tag{2.44}$$

with $c_0 = a_0$.

2.4 Fourier transforms

We began our discussion of Fourier series by considering how a periodic function could be decomposed into, or approximated by, a sum of sinusoidal waves. The analysis can be extended to the non-periodic case by letting $\lambda \to \infty$, so that no repetitions are required of $\mathrm{f}(x)$ within a finite interval. To carry out this limiting procedure, it is helpful to define

$$\Delta k \;=\; \frac{2\pi}{\lambda} \qquad \text{and} \qquad k_n \;=\; n\,\Delta k$$

because the wavenumber of the fundamental harmonic, Δk, shrinks gradually to zero as λ gets ever larger and k_n approaches a continuum even with integer n. The imaginary exponentials of eqns (2.42) and (2.44) can then be written as

$$\mathrm{e}^{\mathrm{i}k_n x} \qquad \text{and} \qquad \mathrm{e}^{-\mathrm{i}k_n x}\,,$$

and the coefficients expressed as

$$c_n = \alpha\, \mathrm{F}(k_n)\, \Delta k\,,$$

where α is a constant and $\mathrm{F}(k)$ is a continuous function of k. With these substitutions, eqns (2.42) and (2.44) become

$$\mathrm{f}(x) = \alpha \sum_{n=-\infty}^{\infty} \mathrm{F}(k_n)\, \mathrm{e}^{\mathrm{i}k_n x}\, \Delta k \quad \text{and} \quad \mathrm{F}(k_n) = \tfrac{1}{2\pi\alpha} \int\limits_{-\lambda/2}^{\lambda/2} \mathrm{f}(x)\, \mathrm{e}^{-\mathrm{i}k_n x}\, \mathrm{d}x\,,$$

where we have expressed the integral over a period as being from $-\lambda/2$ to $\lambda/2$, instead of 0 to λ, for a more symmetrical appearance. The limit of $\lambda\to\infty$, when $\Delta k\to 0$, can now be taken safely, and yields the integrals

$$\boxed{\mathrm{f}(x) = \tfrac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \mathrm{F}(k)\, \mathrm{e}^{\mathrm{i}kx}\, \mathrm{d}k} \tag{2.45}$$

and

$$\boxed{\mathrm{F}(k) = \tfrac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \mathrm{f}(x)\, \mathrm{e}^{-\mathrm{i}kx}\, \mathrm{d}x} \tag{2.46}$$

as being the continuum versions of eqns (2.42) and (2.44), where we have set $\alpha = 1/\sqrt{2\pi}$ for aesthetic reasons of symmetry. These come as a linked pair, and define a *Fourier transform* and its *inverse*; which one is called which is quite arbitrary.

While the exponents of a Fourier transform and its inverse must have opposite signs, their precise definitions are a matter of convention; the choice of α is up to us, for example. If the wavenumber is taken to be $1/\lambda$ instead of $2\pi/\lambda$, as done by spectroscopists, then the exponents will be $\pm\,\mathrm{i}2\pi kx$ and neither integral will have a scaling term:

$$\mathrm{f}(x) = \int\limits_{-\infty}^{\infty} \mathrm{F}(k)\, \mathrm{e}^{\mathrm{i}2\pi kx}\, \mathrm{d}k \qquad \text{and} \qquad \mathrm{F}(k) = \int\limits_{-\infty}^{\infty} \mathrm{f}(x)\, \mathrm{e}^{-\mathrm{i}2\pi kx}\, \mathrm{d}x\,,$$

where k is measured in cycles per unit length, typically cm^{-1}, rather than the SI radians per metre.

Although our goal is to gain a physical insight into Fourier transforms, we first need to discuss some of their formal properties. Basic symmetries are a good place to start, as the most common one is the continuum analogue of eqn (2.43):

$$\mathrm{f}(x) = \mathrm{f}(x)^* \quad \Longleftrightarrow \quad \mathrm{F}(-k) = \mathrm{F}(k)^*\,, \tag{2.47}$$

which states that the Fourier transform of a real function is 'conjugate symmetric'. If one of them possesses a symmetry about the origin, then so too will the other:

$$\mathrm{f}(x) = \begin{cases} \mathrm{f}(-x) \\ -\mathrm{f}(-x) \end{cases} \iff \mathrm{F}(k) = \begin{cases} \mathrm{F}(-k)\,, \\ -\mathrm{F}(-k)\,. \end{cases} \tag{2.48}$$

Equations (2.47) and (2.48) can be combined to show that the Fourier transform of a real and symmetric function is also real and even, whereas that of a real and antisymmetric function is imaginary and odd; this is equivalent to eqn (2.36).

The substitution of $k=0$ in eqns (2.46) and (2.47) reveals $\mathrm{F}(0)$ to be proportional to the area under the curve $y=\mathrm{f}(x)$,

$$\mathrm{F}(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{f}(x)\,\mathrm{d}x\,, \tag{2.49}$$

and necessarily real if $\mathrm{f}(x)=\mathrm{f}(x)^*$. It will equal zero if $\mathrm{f}(x)=-\mathrm{f}(-x)$. Technically, the integral of the modulus, $|\mathrm{f}(x)|$, must be bounded (or finite) if its Fourier transform is to exist everywhere; this is known as the *Dirichlet* condition.

2.4.1 Convolution theorem

One of the most useful results in Fourier theory concerns the *convolution* of two functions. Mathematically, the convolution of $\mathrm{g}(x)$ and $\mathrm{h}(x)$ is defined by

$$\mathrm{g}(x) \otimes \mathrm{h}(x) = \int_{-\infty}^{\infty} \mathrm{g}(t)\,\mathrm{h}(x-t)\,\mathrm{d}t\,, \tag{2.50}$$

where $\mathrm{g}\otimes\mathrm{h}$ is read as 'g convolved with h', and physically represents a 'blurring' of $\mathrm{g}(x)$ by $\mathrm{h}(x)$. This can be understood from the example of Fig. 2.9, where $\mathrm{g}(x)$ consists of four spikes, or δ-*functions*, and $\mathrm{h}(x)$ is a broad asymmetric function. The convolution is carried out

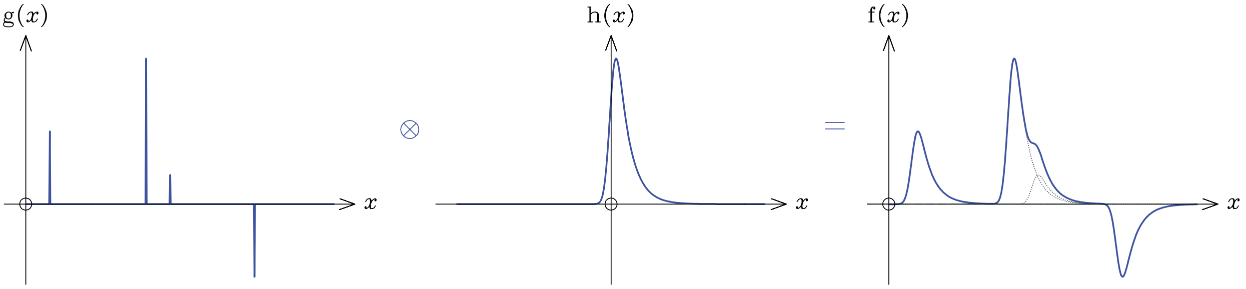


Fig. 2.9 The convolution of the spiky function $\mathrm{g}(x)$ with the broad asymmetric function $\mathrm{h}(x)$: $\mathrm{f}(x) = \mathrm{g}(x) \otimes \mathrm{h}(x)$.

by replacing each of the the sharp peaks in $\mathrm{g}(x)$ with scaled copies of $\mathrm{h}(x)$ and adding together the four contributions; those from the two closely spaced components in the middle, shown by dotted grey lines, combine to give a resultant function where the constituent doublet is no longer resolved clearly. Although it's not as easy to visualize it the other way around, eqn (2.50) can equally be thought of as the blurring of $\mathrm{h}(x)$ by $\mathrm{g}(x)$.

$$\mathrm{g}(x) \otimes \mathrm{h}(x) \;=\; \mathrm{h}(x) \otimes \mathrm{g}(x)$$

The convolution theorem states that the Fourier transform of the convolution of two functions is proportional to the product of their Fourier transforms:

$$\boxed{\mathrm{f}(x) = \mathrm{g}(x) \otimes \mathrm{h}(x) \quad \Longleftrightarrow \quad \mathrm{F}(k) = \sqrt{2\pi}\, \mathrm{G}(k) \times \mathrm{H}(k)} \;, \qquad (2.51)$$

where $\mathrm{F}(k)$, $\mathrm{G}(k)$ and $\mathrm{H}(k)$ are the Fourier transforms of $\mathrm{f}(x)$, $\mathrm{g}(x)$ and $\mathrm{h}(x)$, respectively, according to eqn (2.46). Given the reciprocity between a Fourier transform and its inverse,

$$\mathrm{f}(x) = \mathrm{g}(x) \times \mathrm{h}(x) \quad \Longleftrightarrow \quad \mathrm{F}(k) = \tfrac{1}{\sqrt{2\pi}}\, \mathrm{G}(k) \otimes \mathrm{H}(k)\,. \qquad (2.52)$$

The power of eqn (2.51) will be illustrated in a physical sense in the next section, and throughout this book, but its computational benefit stems from the fact that it's much easier to multiply functions than to convolve them. To work out $\mathrm{g}(x)\otimes\mathrm{h}(x)$ numerically, for example, it's quicker to use a *fast Fourier transform* (FFT) computer subroutine to calculate $\mathrm{G}(k)$ and $\mathrm{H}(k)$, and inverse Fourier transform their product, than to compute the integral of eqn (2.50) directly.

$$\mathrm{g}(x) \otimes \mathrm{h}(x) \;= \; \tfrac{1}{2\pi} \int\limits_{-\infty}^{\infty} \mathrm{G}(k)\; \mathrm{H}(k)\; \mathrm{e}^{\mathrm{i}kx}\; \mathrm{d}k$$

Putting $k=0$ in eqn (2.51), and interpreting $\mathrm{F}(0)$, $\mathrm{G}(0)$ and $\mathrm{H}(0)$ with eqn (2.49), shows that the area under the convolution is equal to the product of the corresponding individual integrals:

$$\int\limits_{-\infty}^{\infty} \big[\,\mathrm{g}(x) \otimes \mathrm{h}(x)\big]\, \mathrm{d}x \;=\; \int\limits_{-\infty}^{\infty} \mathrm{g}(x)\, \mathrm{d}x \;\times\; \int\limits_{-\infty}^{\infty} \mathrm{h}(t)\, \mathrm{d}t\,. \qquad (2.53)$$

This can be seen from the example of Fig. 2.10, where an array of different shaped peaks is convolved with a Gaussian. Although the

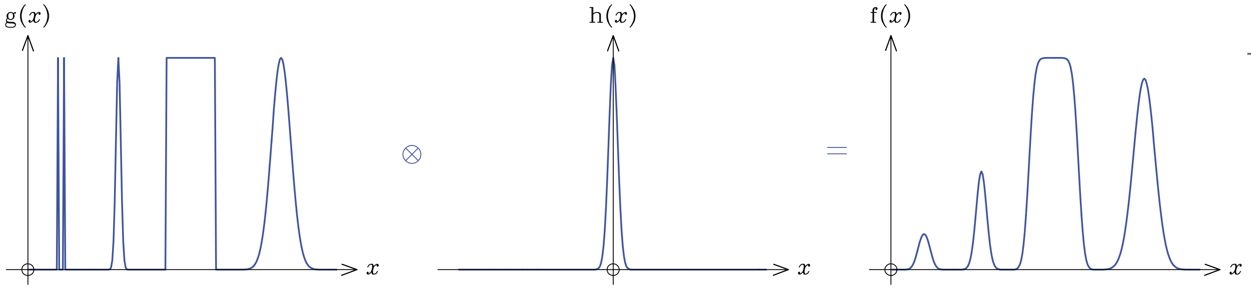
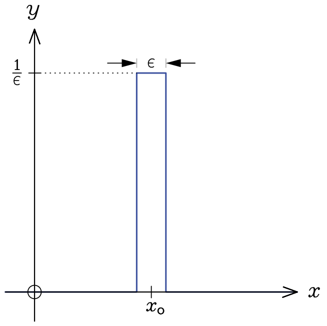


Fig. 2.10 The convolution of a function with an array of different shaped peaks, $\mathrm{g}(x)$, with a Gaussian, $\mathrm{h}(x)$.



The Dirac δ-function

A Dirac δ-function, $\delta(x-x_o)$, is a sharp spike of unit area at a given location, x_o; its simplicity as a 'point impulse' makes it a useful test object for studying equations that model physical situations. Mathematically, it is defined by

$$\delta(x-x_o) = 0 \quad \text{if } x \neq x_o \qquad \text{and} \qquad \int_{-\infty}^{\infty} \delta(x-x_o)\,\mathrm{d}x = 1\,,$$

and can be thought of as the limiting form of a variety of functions as they become ever narrower. Of these the most straightforward is a rectangular column of width ϵ, centred on $x = x_o$, and height $1/\epsilon$; this acquires the properties of $\delta(x-x_o)$ in the limit of $\epsilon \to 0$. An important corollary of the above definition is

$$\int_{a}^{b} \mathrm{f}(x)\,\delta(x-x_o)\,\mathrm{d}x = \begin{cases} \mathrm{f}(x_o) & \text{if } a < x_o < b\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{2.54}$$

so that integrals involving a δ-function are easy to evaluate.

$$\delta(x-x_o) \otimes \mathrm{h}(x) = \mathrm{h}(x-x_o)$$

two spikes on the left of $\mathrm{g}(x)$ merge into one in $\mathrm{f}(x)$, because they are very closely spaced compared with the width of $\mathrm{h}(x)$, the areas of the various components in the blurred output are proportional to those of the input signal. The amplitudes of the narrowest peaks are affected the most, since their relative spreading is the greatest as a result of the convolution; the slowly varying parts of the structure change the least.

2.4.2 Auto-correlation function

The last Fourier concept that we need to consider concerns the *auto-correlation function*, or ACF, which provides information on the distance distribution of the various structures in $\mathrm{f}(x)$. Mathematically, the ACF of $\mathrm{f}(x)$ is defined by

$$\mathrm{ACF}(x) = \int_{-\infty}^{\infty} \mathrm{f}(t)^{*}\,\mathrm{f}(x+t)\,\mathrm{d}t\,, \tag{2.55}$$

and is real if $\mathrm{f}(x)^{*} = \mathrm{f}(x)$. Although this looks like a self-convolution, or $\mathrm{f}(x)^{*} \otimes \mathrm{f}(-x)$, it's not the best way to think about eqn (2.55). The ACF is largest at the origin,

$$\mathrm{f}(x)^{*}\,\mathrm{f}(x) = \left|\mathrm{f}(x)\right|^{2} \geqslant 0$$

$$\mathrm{ACF}(0) = \int_{-\infty}^{\infty} \mathrm{f}(t)^{*}\,\mathrm{f}(t)\,\mathrm{d}t\,,$$

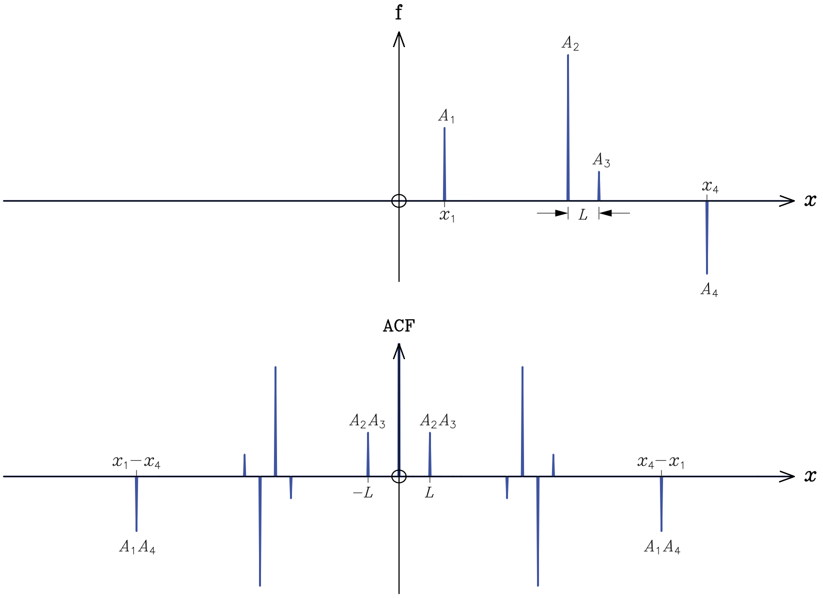


Fig. 2.11 An $\mathrm{f}(x)$ consisting of four sharp peaks and its auto-correlation function. The spike at the origin of the ACF should be three times higher than drawn, and has been suppressed for clarity. The relationship of the closest and farthest peaks in $\mathrm{f}(x)$ to their corresponding mutual contributions in the ACF is indicated.

because everything correlates with itself. The value of the ACF at a distance L away from the origin is calculated by multiplying $\mathrm{f}(x)$ with a copy that's displaced by L relative to it, $\mathrm{f}(L+x)$, and integrating the product; its magnitude is a measure of how much structure there is in $\mathrm{f}(x)$ separated by a distance of L. This can be understood most easily by considering the ACF of a function that consists of a few sharp peaks, such as that shown in Fig. 2.11. Basically, two spikes at x_1 and x_2 in $\mathrm{f}(x)$, with amplitudes A_1 and A_2, will contribute a symmetric pair of very sharp components at $\pm(x_1-x_2)$, and magnitude A_1A_2, towards the ACF of $\mathrm{f}(x)$; they will also add an amount $A_1^2+A_2^2$ to the ACF at the origin.

$$\mathrm{ACF}(-x) \;=\; \mathrm{ACF}(x)^*$$

The reason for discussing the ACF is its linear relationship to the modulus of a Fourier transform:

$$\mathrm{ACF}(x) \;=\; \int_{-\infty}^{\infty} \big|\mathrm{F}(k)\big|^2\,\mathrm{e}^{\mathrm{i}\,k\,x}\,\mathrm{d}k\;, \qquad (2.56)$$

where $\mathrm{F}(k)$ is given by eqn (2.46). While a Fourier transform and its inverse contain the same information, albeit in different ways, and it's possible to switch between one and the other through eqns (2.45) and (2.46), the situation becomes less straightforward if only $\big|\mathrm{F}(k)\big|$ is available. We can begin to appreciate the problems caused by such a loss of the Fourier phase by comparing the relative complexity of the ACF with $\mathrm{f}(x)$ in Fig. 2.11. The ACF, which is directly available

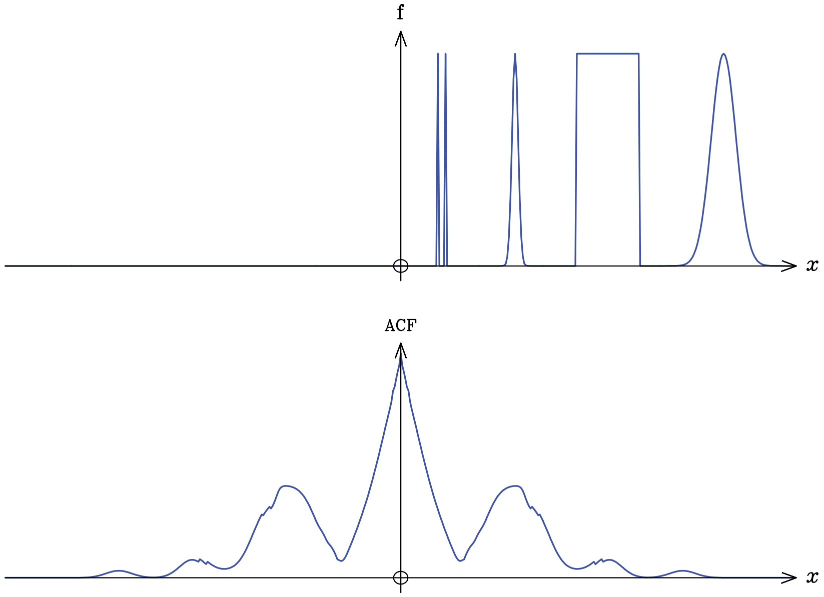


Fig. 2.12 An $\mathrm{f}(x)$ containing a variety of peaks and its auto-correlation function.

from $\left|\mathrm{F}(k)\right|$ through eqn (2.56), is much harder to interpret in terms of the underlying structure; for a diffuse case, such as that in Fig. 2.12, it's almost impossible.

2.5 Fourier optics and physical insight

So far, we have discussed Fourier transforms in a largely abstract context. Now let's try to gain some physical insight into their properties with the aid of diffraction experiments familiar from high school physics. First, though, we need to establish the link between optics and Fourier transforms.

The geometry of the diffraction experiment is shown in Fig. 2.13, where a travelling plane wave passes through a set of slits and produces a pattern of dark and light bands on a very distant screen. We have made the problem one-dimensional for simplicity, but will indicate its generalization later. The nature of the aperture is defined by the function $\mathrm{A}(x)$, which describes how much light passes through it at position x; this is called the *aperture function*. It usually only takes values of zero or one, corresponding to complete opaqueness and transparency respectively, but it could in principle be complex with $0 \leqslant \left|\mathrm{A}(x)\right| \leqslant 1$.

To calculate the diffraction pattern, the principle of superposition tells us that we need to add up all the waves that emerge through the aperture. The amplitude of the contribution from the narrow region between x and $x + \Delta x$ is proportional to $\mathrm{A}(x)\,\Delta x$, but what

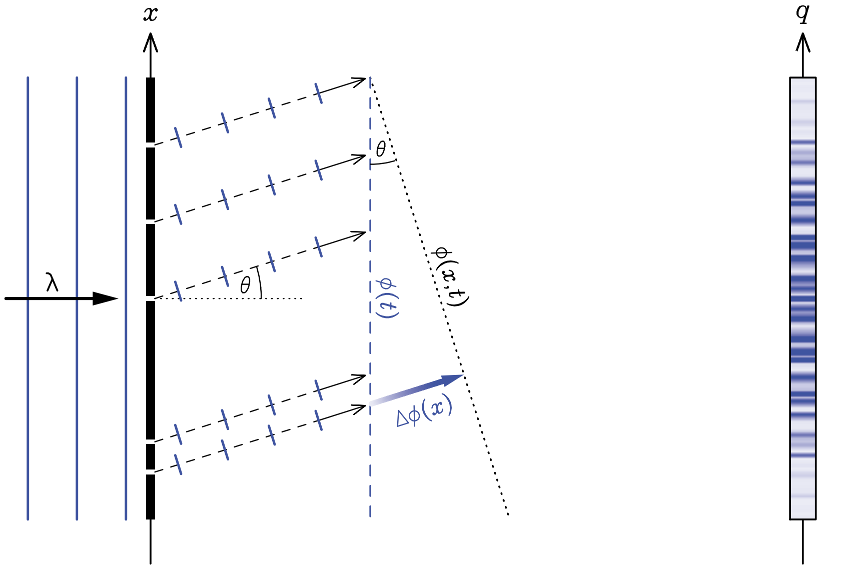


Fig. 2.13 The geometry for Fraunhofer diffraction by a one-dimensional aperture, $\mathrm{A}(x)$; the interference pattern of interest, $\mathrm{I}(q)$, is projected onto a distant screen.

about its phase ϕ? That depends on both x and the angle of propagation relative to the incident wave, θ, as well as the time t. The phase will be invariant with position parallel to the incoming wavefront, but will gain a relative factor of

$$\Delta\phi \;=\; \left(\tfrac{2\pi}{\lambda}\right)\, x \sin\theta$$

in the direction of θ due to the associated path difference of $x\sin\theta$. Hence, the complex contribution to the resultant wave is

$$\Delta\psi \;=\; \psi_{\mathrm{o}}\, \mathrm{A}(x)\, \mathrm{e}^{\mathrm{i}\,qx}\, \Delta x\ ,$$

where $q = 2\pi\sin\theta/\lambda$ and the temporal variation has been absorbed into the 'constant' of proportionality, ψ_{o}. The diffracted wave, ψ, is the sum of all such terms; in the limit $\Delta x \to 0$, it becomes the Fourier transform of the aperture function:

$$\psi(q) \;=\; \psi_{\mathrm{o}} \int\limits_{-\infty}^{\infty} \mathrm{A}(x)\, \mathrm{e}^{\mathrm{i}\,qx}\, \mathrm{d}x\ . \tag{2.57}$$

Thus we met Fourier transforms a (very) long time ago but did not realize it! Before reminding ourselves of the results from elementary diffraction experiments, and trying to understand them in terms what we've now learnt about Fourier transforms, we need to make a few qualifying remarks.

The first point is essentially a technicality, but the above analysis assumes that we are considering *Fraunhofer* diffraction. This is the limit where the projection screen is so far away that all the waves reaching a particular point can be considered to be travelling in parallel directions. The equations becomes more cumbersome when this approximation does not hold, and leads to the theory of *Fresnel* diffraction.

The more serious point of note is that the observed, or measured, diffraction pattern is not the complex function $\psi(q)$ but its intensity, or modulus-squared, $\mathrm{I}(q)$:

$$\mathrm{I}(q) \;=\; \left|\psi(q)\right|^2 \;=\; \psi(q)\,\psi(q)^* \,. \tag{2.58}$$

The difficulties caused by such a loss of phase information, in terms of ascertaining the aperture function from its diffraction pattern, have been alluded to in Section 2.4.2, but we will encounter them again throughout this book.

2.5.1 Young's double slit

A first introduction to interference experiments usually involves a *Young's double slit*. This consists of a pair of very narrow slits that are separated by a distance d, and give rise to a diffraction pattern of uniformly spaced dark and light bands which become closer together as d increases. Let's try to understand this theoretically by using eqns (2.57) and (2.58).

The aperture function for a Young's double slit can be modeled by two δ-functions located at a distance of $d/2$ on either side of an arbitrarily defined origin,

$$\mathrm{A}(x) \;=\; \delta\big(x-\tfrac{d}{2}\big) + \delta\big(x+\tfrac{d}{2}\big)\,,$$

$$\int_{-\infty}^{\infty} \delta(x-x_{\mathrm{o}})\,\mathrm{e}^{\mathrm{i}qx}\,\mathrm{d}x \;=\; \mathrm{e}^{\mathrm{i}qx_{\mathrm{o}}}$$

and is plotted in Fig. 2.14. Since δ-functions are easy to integrate, from eqn (2.54), the Fourier transform of $\mathrm{A}(x)$ is readily shown to yield

$$\begin{aligned}\psi(q) \;&=\; \psi_{\mathrm{o}}\left(\mathrm{e}^{\mathrm{i}qd/2} + \mathrm{e}^{-\mathrm{i}qd/2}\right) \\ &=\; \psi_{\mathrm{o}}\;2\cos\!\left(\tfrac{qd}{2}\right),\end{aligned}$$

where we have used eqn (2.32) in writing the second line. The product of this diffracted wave with its complex conjugate, $\psi(q)^*$, leads to the prediction

$$\mathrm{I}(q) \;\propto\; \left[\cos\!\left(\tfrac{qd}{2}\right)\right]^2 \;\propto\; 1+\cos(qd)\;, \tag{2.59}$$

$$\cos 2\theta \;=\; 2\cos^2\theta - 1$$

where all the multiplicative prefactors not involving q, such as $\left|\psi_{\mathrm{o}}\right|^2$, have been omitted and a trigonometric double angle formula used on the far right-hand side. This pattern of 'uniform cosine fringes' is plotted in Fig. 2.14.

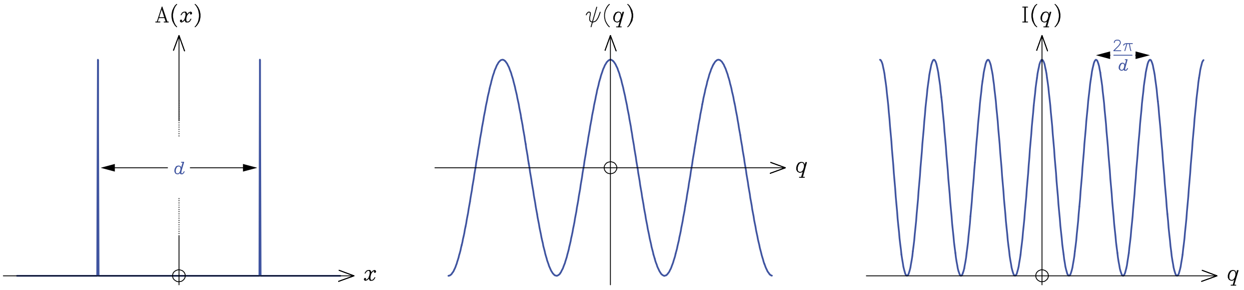


Fig. 2.14 The aperture function for a Young's double slit, A(x), its Fourier transform, $\psi(q)$, and the diffraction pattern, I(q).

The theoretical result in eqn (2.59) is consistent with the experimental observations: the dark and light bands are equally spaced, of uniform intensity and become closer together in inverse proportion to the distance d between the slits. The last feature is a universal property of Fourier transforms: the length scales which characterize a function and its Fourier transform are inversely related to each other. This leads to the use of the terminology *reciprocal space* when referring to the Fourier domain.

2.5.2 A single wide slit

Another common interference experiment involves a single wide slit that gives rise to a diffraction pattern where the intensity of the light bands diminishes rapidly away from a central bright region, which is itself twice as broad as the rest. Let's also try to understand this theoretically.

If we take the x-origin to be in the middle of the slit of width w, then the aperture function becomes

$$\mathrm{A}(x) \;=\; \begin{cases} 1 & \text{if } |x| \leqslant \frac{w}{2}\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

and is plotted in Fig. 2.15. According to eqn (2.57), therefore,

$$\psi(q) \;=\; \psi_{\mathrm{o}} \int\limits_{-w/2}^{w/2} \mathrm{e}^{\mathrm{i}qx}\,\mathrm{d}x\,.$$

This Fourier transform is easy to evaluate, because the integration of an exponential is straightforward, and yields

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{\mu x}\right) \;=\; \mu\,\mathrm{e}^{\mu x}$$

$$\psi(q) \;=\; \psi_{\mathrm{o}}\left[\frac{\mathrm{e}^{\mathrm{i}qx}}{\mathrm{i}q}\right]_{-w/2}^{w/2} \;=\; \frac{\psi_{\mathrm{o}}}{\mathrm{i}q}\left(\mathrm{e}^{\mathrm{i}qw/2}-\mathrm{e}^{-\mathrm{i}qw/2}\right).$$

The difference of the imaginary exponentials on the far right-hand side can be recognized as being equal to $2\mathrm{i}$ times $\sin(qw/2)$ from

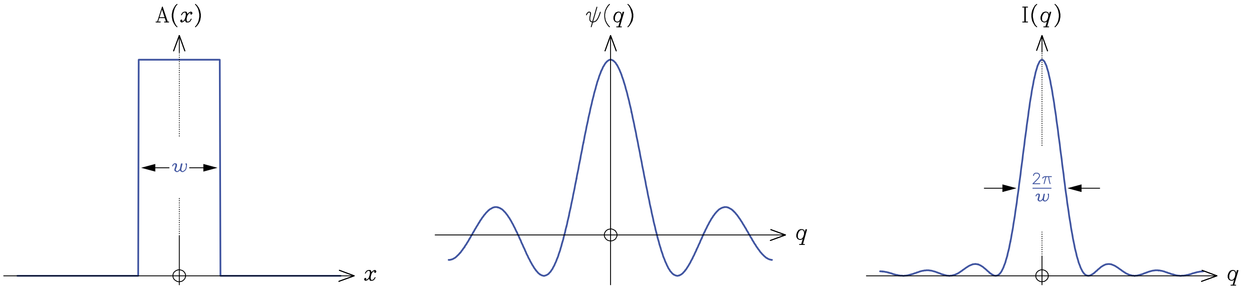


Fig. 2.15 The aperture function for a single wide slit, $\mathrm{A}(x)$, its Fourier transform, $\psi(q)$, and the diffraction pattern, $\mathrm{I}(q)$.

eqn (2.32). With this substitution, the modulus-squared of $\psi(q)$ leads to the prediction

$$\mathrm{I}(q) \;\propto\; \left[\tfrac{1}{q}\sin\!\left(\tfrac{qw}{2}\right)\right]^2 \;\propto\; \frac{1-\cos(qw)}{q^2}\;, \qquad (2.60)$$

$$\operatorname{sinc}\theta \;=\; \frac{\sin\theta}{\theta} \;\rightarrow 1 \text{ as } \theta\rightarrow 0$$

which is shown in Fig. 2.15 and consistent with the *sinc*-squared behaviour of the observed diffraction pattern. We again see the inverse relationship between the width of the aperture function and the spread of the diffraction pattern: as one of them becomes broader the other gets narrower.

2.5.3 A diffraction grating

A diffraction grating is an aperture consisting of a large number of thin, parallel and equally spaced lines. In one dimension, it can be modelled as a periodic array of δ-functions:

$$\mathrm{A}(x) \;=\; \sum_{m=-\infty}^{\infty} \delta(x-md)\;,$$

$$\int_{-\infty}^{\infty} \delta(x-md)\,\mathrm{e}^{\mathrm{i}qx}\,\mathrm{d}x \;=\; \mathrm{e}^{\mathrm{i}qmd}$$

where d is the distance between the grating lines. Swapping the order of integration and summation, and using eqn (2.54), the Fourier transform of eqn (2.57) reduces to

$$\psi(q) \;=\; \psi_{\mathrm{o}} \sum_{m=-\infty}^{\infty} \mathrm{e}^{\mathrm{i}qdm}\,.$$

The nature of $\psi(q)$ becomes apparent once we realize that it's proportional to the sum of complex numbers that are of unit magnitude but varying phase. They will add up coherently if the product qd is an integer number of 2π, yielding a huge resultant sum, but cancel out otherwise. Hence,

$$\psi(q) \;\propto\; \sum_{n=-\infty}^{\infty} \delta(q-n\,q_{\mathrm{o}})\;, \qquad (2.61)$$

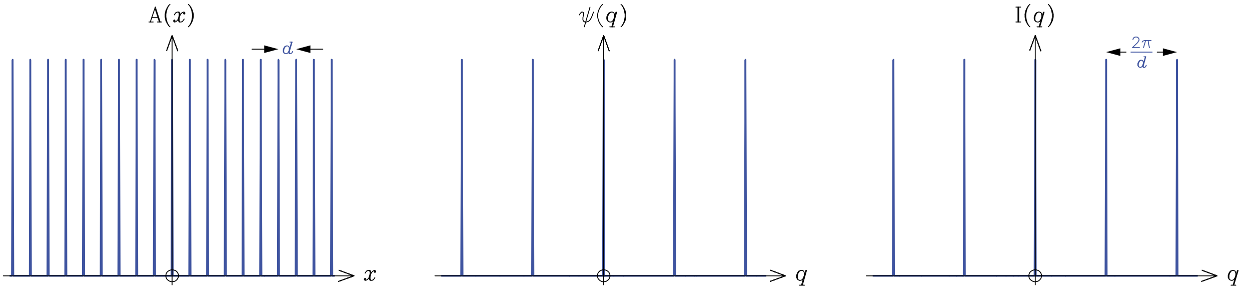


Fig. 2.16 The aperture function for a diffraction grating, $\mathrm{A}(x)$, its Fourier transform, $\psi(q)$, and the diffraction pattern, $\mathrm{I}(q)$.

where $q_\mathrm{o} = 2\pi/d$. The diffraction pattern has the same structure as the grating, therefore, but the spacing of the lines is inversely related to d (Fig. 2.16).

In terms of the physical set up of Fig. 2.13, where $q = 2\pi\sin\theta/\lambda$, sharp bright lines are seen when

$$\frac{2\pi\sin\theta}{\lambda} = \frac{2\pi n}{d}$$

$$n\,\lambda \;=\; d\,\sin\theta\,, \tag{2.62}$$

for $n = 0, \pm 1, \pm 2, \ldots, \pm n_\mathrm{max}$, where the trigonometric constraint that $|\sin\theta| \leqslant 1$ imposes a cutoff on the highest observable order n_max. If the spacing of the diffraction grating is known, such as 500 lines per millimetre (so that $d = 2\,\mu$m), then eqn (2.62) provides the basis for an accurate measurement of the wavelength of the illumination. If white light is used for the experiment, then the intense central line is accompanied by increasingly dispersed rainbows for the higher orders; this is because each of the wavelengths that makes up white light satisfies eqn (2.62) for a slightly different angle θ for a given value of $n \neq 0$.

2.5.4 The convolution theorem in action

Although a real diffraction grating isn't infinite as assumed above, we expect the analysis to be a very good approximation for one that is sufficiently large. The case of a grating of limited extent w can be addressed by combining the results of eqns (2.60) and (2.61) through the convolution theorem: as the aperture function can be expressed as a product of an infinite grating with line spacing d and a single slit of width w, as illustrated in Fig. 2.17, the Fourier transform of the finite grating is equal to the convolution of the Fourier transforms of the infinite grating and the single wide slit. The resultant diffraction pattern is simply that of the infinite grating but with each of the δ-functions replaced by a narrow sinc-squared function, as shown in Fig. 2.17. A qualification is in order here, in that $\mathrm{I}(q) \propto |\mathrm{G}(q)|^2 \otimes |\mathrm{H}(q)|^2$ is only an approximation (albeit a good one); strictly speaking, $\mathrm{I}(q) \propto |\,\mathrm{G}(q) \otimes \mathrm{H}(q)\,|^2$. Given the inverse relationship between the length scales of a function and its Fourier trans-

$$\mathrm{A}(x) \;=\; \mathrm{g}(x) \times \mathrm{h}(x)$$

$$\therefore\;\; \psi(q) \;=\; \psi_\mathrm{o}\,\mathrm{G}(q) \otimes \mathrm{H}(q)$$

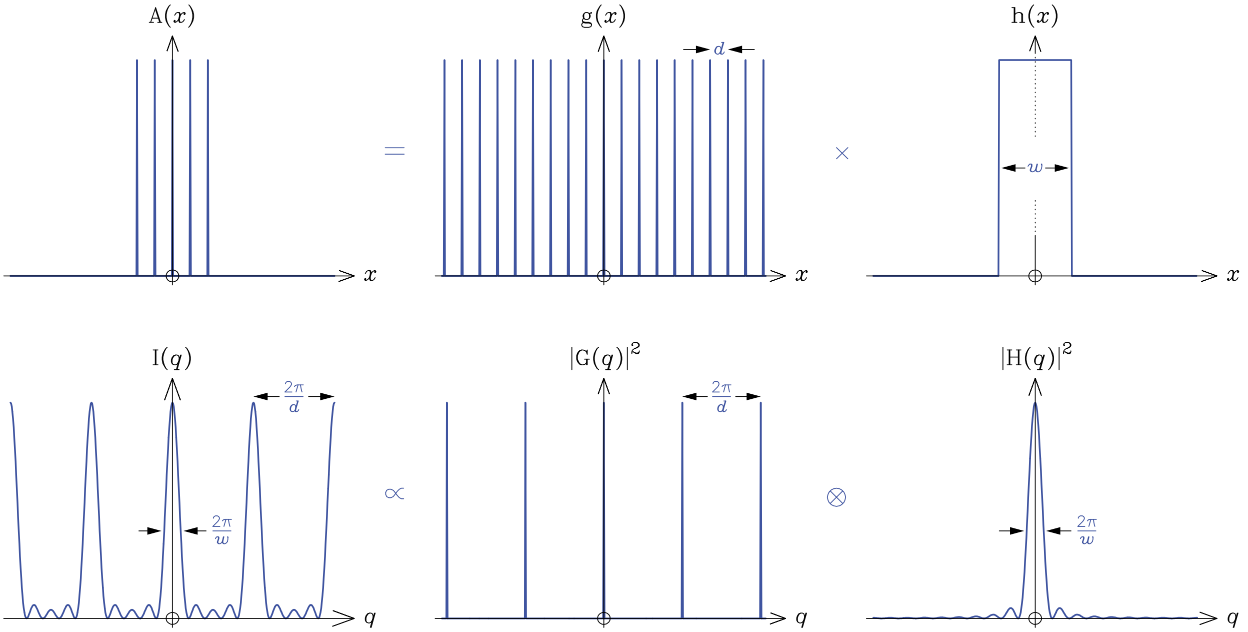


Fig. 2.17 The diffraction pattern from a grating of limited extent, w, can be evaluated from a knowledge of the Fourier transforms of an infinite grating, with line spacing d, and a single slit, of width w, through the use of the convolution theorem.

form, the width of the large diffraction peaks tells us about the size w of the grating whereas the distance between them indicates the d-spacing of its lines. As the number of grating lines goes up, so that the ratio w/d increases, the principal peaks become narrower and more low-level wiggles appear between them.

$$\mathrm{A}(x) = \mathrm{g}(x) \otimes \mathrm{h}(x)$$

$$\therefore \quad \psi(q) = \psi_{\mathrm{o}}\, \mathrm{G}(q) \times \mathrm{H}(q)$$

The convolution theorem also enables us to ascertain the diffraction pattern for a pair of broad slits from the results of eqns (2.59) and (2.60). Taking each to be of width w, and separated by d, the aperture function can be seen as a convolution of an ideal Young's double slit with a narrow but finite single slit, as in Fig. 2.18. Since the Fourier transform of the former is then equal to the product of those of the latter, the intensity of the uniform cosine fringes that we'd expect from a perfect Young's double slit is modulated by a slowly varying sinc-squared function.

2.5.5 Multi-dimensional generalization

$$\mathbf{k} = (k_x, k_y, k_z)$$

Having illustrated Fourier transforms and the use of the convolution theorem with one-dimensional versions of familiar high school experiments, let's indicate the multi-dimensional generalization of eqn (2.57). A closer examination of Fig. 2.13 reveals q to be the x-component of the wavevector $\mathbf{k}$ of Section 2.1.1:

$$k_x = \left(\tfrac{2\pi}{\lambda}\right) \sin\theta = q \qquad \text{and} \qquad k_z = \left(\tfrac{2\pi}{\lambda}\right) \cos\theta \,,$$

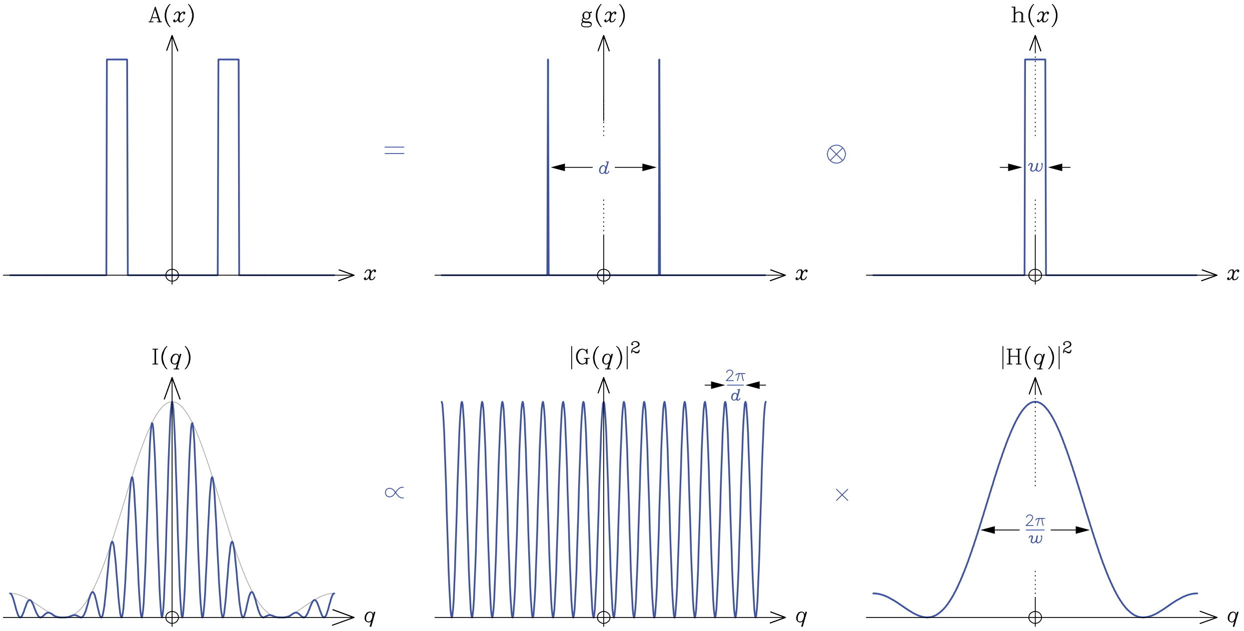


Fig. 2.18 The diffraction pattern from a pair of slits of width w and separation d can be evaluated from a knowledge of the Fourier transforms of a Young's double slit, of spacing d, and a single slit, of width w, with the convolution theorem.

where we have taken z to be the original direction of propagation, from the aperture to the projection screen, and $|\mathbf{k}| = 2\pi/\lambda$. With this observation, it seems plausible that the two-dimensional diffraction pattern, $\mathrm{I}(k_x, k_y)$, from an aperture in the x–y plane, $\mathrm{A}(x, y)$, with y coming out of the page in Fig. 2.13, might be given by the modulus-squared of

$$\psi(k_x, k_y) \;=\; \psi_\mathrm{o} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{A}(x, y)\, \mathrm{e}^{\mathrm{i}(k_x x + k_y y)}\, \mathrm{d}x\, \mathrm{d}y\,. \qquad (2.63)$$

This *double integral*, over the surface area of the aperture, simplifies to the product of two one-dimensional integrals if the aperture function is *separable*:

$$\psi(k_x, k_y) \;=\; \psi_\mathrm{o} \int_{-\infty}^{\infty} \mathrm{A}_1(x)\, \mathrm{e}^{\mathrm{i}k_x x}\, \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{A}_2(y)\, \mathrm{e}^{\mathrm{i}k_y y}\, \mathrm{d}y$$

if $\mathrm{A}(x, y) = \mathrm{A}_1(x)\, \mathrm{A}_2(y)$. The substitution of either $\delta(y)$ or a constant for $\mathrm{A}_2(y)$, and the properties of δ-functions, allows us to confirm that eqn (2.63) reduces to the one-dimensional form of eqn (2.57) if the aperture is either infinitesimally thin or invariant with respect to y. Strictly speaking,

$$\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}(q - q_\mathrm{o})t}\, \mathrm{d}t \;=\; 2\pi\, \delta(q - q_\mathrm{o})$$

$$\psi \to \psi_1(k_x) \quad \text{as} \quad \mathrm{A} \to \mathrm{A}_1(x)\, \delta(y) \qquad \text{and} \qquad \psi \to \psi_1(k_x)\, \delta(k_y) \quad \text{as} \quad \mathrm{A} \to \mathrm{A}_1(x)$$

From Sivia and Rawlings (1999), *Foundations of Science Mathematics*, Oxford Chemistry Primers Series, **77**.

Multiple integrals

In ordinary integration, we are concerned with the area under the curve $y = \mathrm{f}(x)$. Many functions of interest in real life entail several variables, and *multiple integrals* are a natural extension of the one-dimensional ideas to deal with multivariate problems.

To get a feel for how multiple integrals arise, let's consider a couple of physical examples. Suppose that we wish to calculate the force exerted on a wall by a gale. If the pressure P was constant across the whole face with area A, then the total force is simply $\mathrm{P}\times\mathrm{A}$. With a varying pressure $\mathrm{P}(x,y)$, the answer is not so obvious. This situation can be handled by thinking about the wall as consisting of many small square segments, each with area $\delta x\,\delta y$, so that the total force is the sum of all the contributions $\mathrm{P}(x,y)\,\delta x\,\delta y$; in the limiting case when $\delta x\to 0$ and $\delta y\to 0$, we have

$$\text{Force} \;=\; \iint\limits_{\text{wall}} \mathrm{P}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

where the double integral indicates that the infinitesimal summation is being carried out over a two-dimensional surface (in the x and y directions). Incidentally, if the wall does not have a conventional (rectangular) shape then its area can be calculated similarly according to

$$\text{Area} \;=\; \iint\limits_{\text{wall}} \mathrm{d}x\,\mathrm{d}y\,.$$

The double integral is also called a *surface integral*.

Another illustration is provided by quantum mechanics where the modulus-squared of the wave function, $|\psi(x,y,z)|^2$, of an electron (say) gives the *probability density* of finding it at some point in space. The chances that the electron is in a small (cuboid) region of volume $\delta x\,\delta y\,\delta z$ is then $|\psi(x,y,z)|^2\,\delta x\,\delta y\,\delta z$. Hence, the probability of finding it within a finite domain V is given by

$$\text{Probability} \;=\; \iiint\limits_{\mathrm{V}} |\psi(x,y,z)|^2\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,,$$

which is known as a *triple*, or *volume*, *integral*.

and demonstrates the reciprocal Fourier relationship between the widths of $\mathrm{A}_2(y)$ and $\psi_2(k_y)$ in the limit of complete invariance versus a δ-function. A careful consideration of the situation, in a manner analogous to that used to derive eqn (2.57), shows eqn (2.63) to be the correct two-dimensional extension.

The most common case of two-dimensional diffraction is from a circular hole, but a rectangle is easier to deal with analytically. This is because the aperture function of the latter, which is equal to one inside the rectangle and zero outside it, is separable and yields a

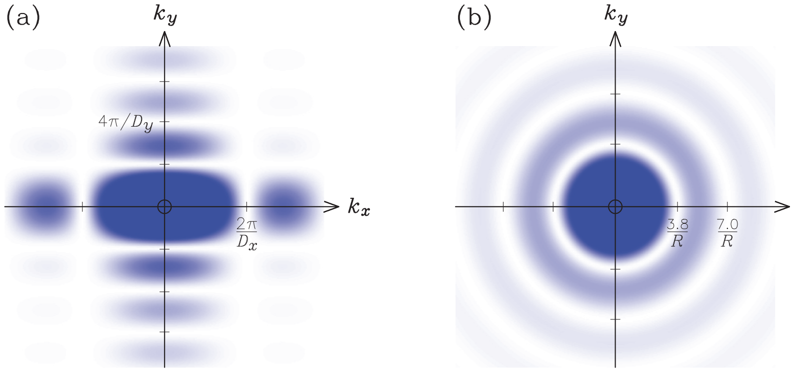


Fig. 2.19 Diffraction patterns from two-dimensional apertures: (a) a rectangular opening of size D_x by D_y, and (b) a circular hole of radius R.

Fourier transform that is just a product of the now familiar sinc functions in k_x and k_y; its modulus-squared is shown in Fig. 2.19(a). The evaluation of the integral of eqn (2.63) is less straightforward for a circular aperture, but the resultant diffraction pattern is plotted in Fig. 2.19(b). It is circularly symmetric, depending only on $k_x^2 + k_y^2$, and is similar to a sinc function in the radial direction; the behaviour is formally governed by a J_1 *Bessel function*. The central bright region is called an *Airy disc*, and its spread is the basis of the resolution formula of eqn (1.8).

Having seen the formulae for the Fourier transforms of one- and two-dimensional functions, in eqns (2.45), (2.46) and (2.63), we can state the M-dimensional generalization succinctly by using vector notation:

$$\boxed{\mathrm{f}(\mathbf{r}) \;=\; (2\pi)^{-\frac{M}{2}} \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty} \mathrm{F}(\mathbf{k})\;\mathrm{e}^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\;\mathrm{d}^M\mathbf{k}} \qquad (2.64)$$

$\mathrm{d}^3\mathbf{k} \;=\; \mathrm{d}k_x\,\mathrm{d}k_y\,\mathrm{d}k_z$

and

$\mathbf{k}\bullet\mathbf{r} \;=\; k_x x + k_y y + k_z z + \cdots$

$$\boxed{\mathrm{F}(\mathbf{k}) \;=\; (2\pi)^{-\frac{M}{2}} \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty} \mathrm{f}(\mathbf{r})\;\mathrm{e}^{\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\;\mathrm{d}^M\mathbf{r}} \qquad (2.65)$$

$\mathrm{d}^3\mathbf{r} \;=\; \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$

where $\mathbf{r} = (x, y, z, \dots)$ and $\mathbf{k} = (k_x, k_y, k_z, \dots)$ have M components, with corresponding hyper-volume elements $\mathrm{d}^M\mathbf{r}$ and $\mathrm{d}^M\mathbf{k}$. We reiterate that a Fourier transform and its inverse come as a linked pair, but which one is called which is arbitrary. Their precise definitions are also a matter of convention. No multiplicative prefactors are required if the wavevector is specified in cycles rather than radians per unit length, for example, when the $\mathbf{k}$ in the exponents is replaced with $2\pi\mathbf{k}$.

2.6 Fourier data analysis

The analysis of data from X-ray and neutron scattering experiments is similar to the task of making inferences about the aperture function from its diffraction pattern. If we knew that $\mathrm{A}(x)$ consisted of a small number of slits, n say, of equal spacing d, as in Fig. 2.17 with $w=(n-1)\,d$, then an examination of the width and separation of the principal peaks in $\mathrm{I}(q)$ readily provides the desired parameters n and d. In less well informed circumstances, however, all we have to go on is the relationship between $\mathrm{A}(x)$ and $\mathrm{I}(q)$ enshrined in eqns (2.57) and (2.58). How can the data then be analysed and what difficulties are likely to arise?

2.6.1 The phase problem

Ignoring matters of practicality for the moment, the most relevant mathematical operation that can be performed on a diffraction pattern is a Fourier transform:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \mathrm{I}(\mathbf{k})\ \mathrm{e}^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\ \mathrm{d}^M\mathbf{k}\ \propto\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \mathrm{A}(\mathbf{t})^*\,\mathrm{A}(\mathbf{r}+\mathbf{t})\ \mathrm{d}^M\mathbf{t}\ ,$$

giving the M-dimensional generalization of the ACF of eqns (2.55) and (2.56), where

$$\mathrm{I}(\mathbf{k})\ =\ \big|\psi(\mathbf{k})\big|^2 \quad \text{and} \quad \psi(\mathbf{k})\ =\ \psi_{\mathrm{o}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \mathrm{A}(\mathbf{r})\ \mathrm{e}^{\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\ \mathrm{d}^M\mathbf{r}\ .$$

Whereas the correspondence between $\mathrm{A}(\mathbf{r})$ and the complex function $\psi(\mathbf{k})$ is one-to-one, implying that there is no loss of information in the transformation, the same is not true of $\mathrm{A}(\mathbf{r})$ and the real and positive diffraction pattern $\mathrm{I}(\mathbf{k})$. Only the auto-correlation function of $\mathrm{A}(\mathbf{r})$ can be ascertained unambiguously from $\mathrm{I}(\mathbf{k})$, and we have already seen, in Figs. 2.11 and 2.12, how much more difficult it is to interpret the ACF than $\mathrm{A}(\mathbf{r})$.

The simplest way of appreciating how the lack of phase, $\arg\{\psi(\mathbf{k})\}$, in a diffraction pattern results in a loss of uniqueness about $\mathrm{A}(\mathbf{r})$ is to consider the Fourier transform of an aperture function that has been shifted by $\mathbf{r}_{\mathrm{o}}$,

$$\psi_{\mathrm{o}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \mathrm{A}(\mathbf{r}+\mathbf{r}_{\mathrm{o}})\ \mathrm{e}^{\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\ \mathrm{d}^M\mathbf{r}\ =\ \psi(\mathbf{k})\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}_{\mathrm{o}}}\ ,$$

which differs from that of $\mathrm{A}(\mathbf{r})$ only through an additional factor of $-\mathbf{k}{\bullet}\mathbf{r}_{\mathrm{o}}$ in its argument; the intensity, or modulus-squared,

$\mathrm{e}^{\theta}\,\mathrm{e}^{-\theta}\ =\ \mathrm{e}^{0}=1$

$$\Big[\psi(\mathbf{k})\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}_{\mathrm{o}}}\Big]\Big[\psi(\mathbf{k})\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}_{\mathrm{o}}}\Big]^*\ =\ \psi(\mathbf{k})\,\psi(\mathbf{k})^*\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}_{\mathrm{o}}}\,\mathrm{e}^{\mathrm{i}\mathbf{k}\bullet\mathbf{r}_{\mathrm{o}}}\ ,$$

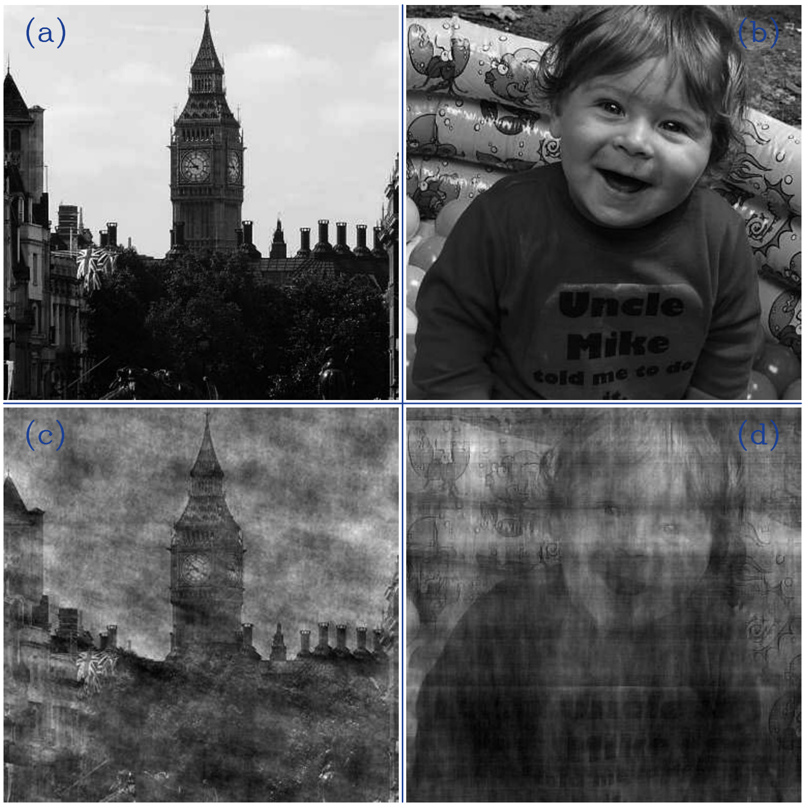


Fig. 2.20 The phase problem: (c) has the Fourier phases of (a) and the Fourier amplitudes of (b), while (d) has the phases of (b) and the amplitudes of (a).

is unchanged. The amplitude of a Fourier transform is, therefore, insensitive to translation. The same is true of the inversion of a real function, so that $\mathrm{A}(\mathbf{r})$ and $\mathrm{A}(-\mathbf{r})$ give identical diffraction patterns if $\mathrm{A}(\mathbf{r})=\mathrm{A}(\mathbf{r})^*$. This provides another elementary demonstration of the loss of uniqueness without the Fourier phase.

The importance of the phase of a Fourier transform can be illustrated dramatically with graphical examples of the type shown in Fig. 2.20. Here two photographs, pertaining to any subject or scene, are Fourier transformed numerically and the phases of one assigned to the amplitudes of the other. Each of these hybrid sets of complex coefficients is then inverse Fourier transformed, and the resultant pictures examined visually. Instinctively we would guess that the outcome of this numerical experiment will be a complete mess, for why should the Fourier phases of one distribution of light intensity have anything to do with the amplitudes from another; if not, we might expect to see some sort of mixture of the two sources. What we find in practice is certainly degraded compared to the originals, but each output only resembles the scene which contributed the Fourier phase with no hint of that from which the amplitudes were taken.

It seems that most of the structural information in a Fourier transform resides in its phase; and since this is missing in diffraction data, it makes their analysis difficult in general without additional prior knowledge.

2.6.2 Truncation effects and windowing

Even when the main interest is in the ACF of the aperture function, and the absence of Fourier phase is not a problem, the limited sampling of a diffraction pattern causes difficulties in practice. In the simplest one-dimensional case, when $\mathrm{I}(q)$ is available only within the range $|q| \leqslant q_{\max}$, the truncated Fourier integral

$$\int_{-q_{\max}}^{q_{\max}} \mathrm{I}(q)\ \mathrm{e}^{-\mathrm{i}qx}\,\mathrm{d}q \;=\; 2\int_{0}^{q_{\max}} \mathrm{I}(q)\ \cos(qx)\ \mathrm{d}q\ , \tag{2.66}$$

$\mathrm{A}(x) = \mathrm{A}(x)^* \implies \mathrm{I}(q) = \mathrm{I}(-q)$

where the cosine equivalent on the right assumes that the aperture function is real, yields an estimate of the ACF that is corrupted by ripples with a characteristic wavelength of $2\pi/q_{\max}$. These artefacts can be understood with the aid of the convolution theorem, and Fig. 2.21, by considering eqn (2.66) to be the Fourier transform of the product of the full but unmeasured diffraction pattern, $\mathrm{J}(q)$, and a ‘top-hat’ function of width $2\,q_{\max}$, $\mathrm{H}(q)$. The result is, therefore, the true but unknown auto-correlation function, acf, convolved with a

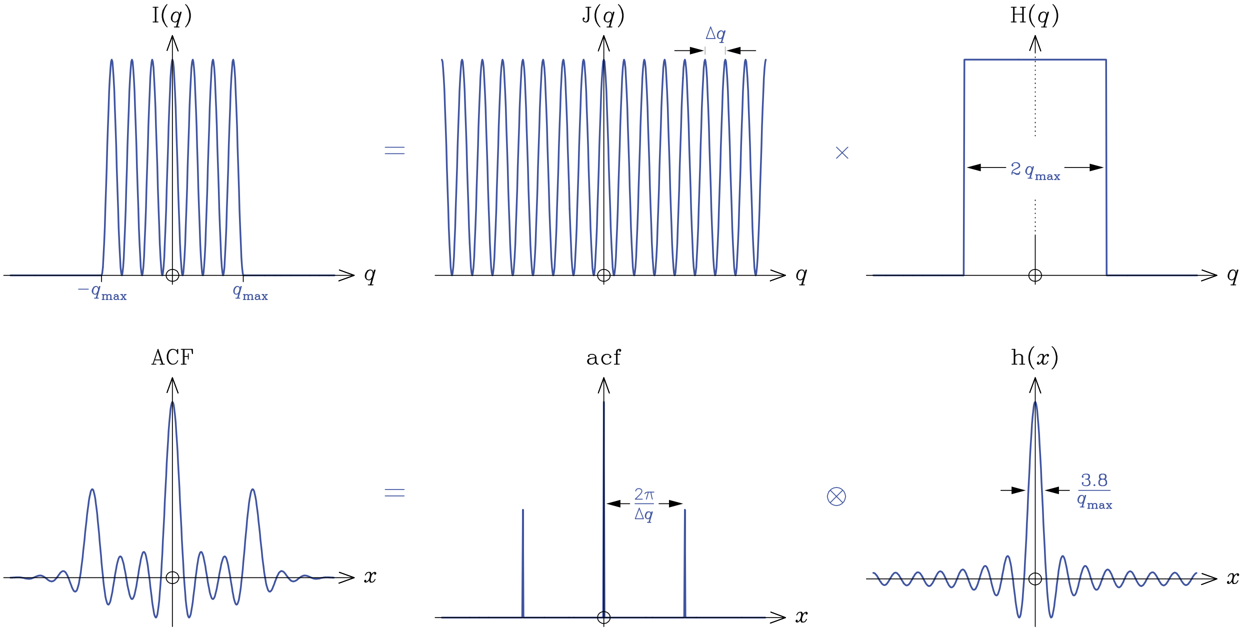


Fig. 2.21 The Fourier transform of a diffraction pattern of limited q-extent, $\mathrm{I}(q)$, yields an ACF of the aperture function which is corrupted by truncation ripples associated with $q_{\max}$; their origin is easily understood from the convolution theorem.

sinc function, $\mathrm{h}(x)$, whose central peak has a full width at half maximum (FWHM) of about $3.8/q_{\max}$.

The messy picture due to the truncation ripples can be cleaned up greatly by multiplying the incomplete diffraction pattern, $\mathrm{I}(q)$, with a *window* function, $\mathrm{W}(q)$, which decays smoothly from one at the origin to zero around $\pm q_{\max}$, before the (inverse) Fourier transform is calculated. This is illustrated in Fig. 2.22 with the ubiquitous *Gaussian*,

$$\mathrm{W}(q) \;=\; \exp\!\left(-\frac{q^2}{2\,\sigma^2}\right), \tag{2.67}$$

$$\mathrm{FWHM} \;=\; \sqrt{8\ln 2}\;\sigma \;\approx\; 2.35\,\sigma$$

whose *standard deviation* σ was chosen somewhat arbitrarily as $q_{\max}/2$. The resultant auto-correlation function, denoted by acf, is said to be a *filtered* version of the ACF given by $\mathrm{I}(q)$. The suppression of the truncation ripples can also be understood from the convolution theorem, which tells us that $\mathrm{acf}(x) = \mathrm{ACF}(x) \otimes \mathrm{w}(x)$, because the subsidiary oscillations are averaged out through a blurring with the filter $\mathrm{w}(x)$. The latter is just the Fourier transform of the windowing function, $\mathrm{W}(q)$, with

$$\mathrm{w}(x) \;\propto\; \exp\!\left(-\frac{\sigma^2 x^2}{2}\right) \tag{2.68}$$

for the case of eqn (2.67). Although the spurious peaks and troughs are increasingly reduced as $\mathrm{w}(x)$ becomes broader, requiring $\mathrm{W}(q)$

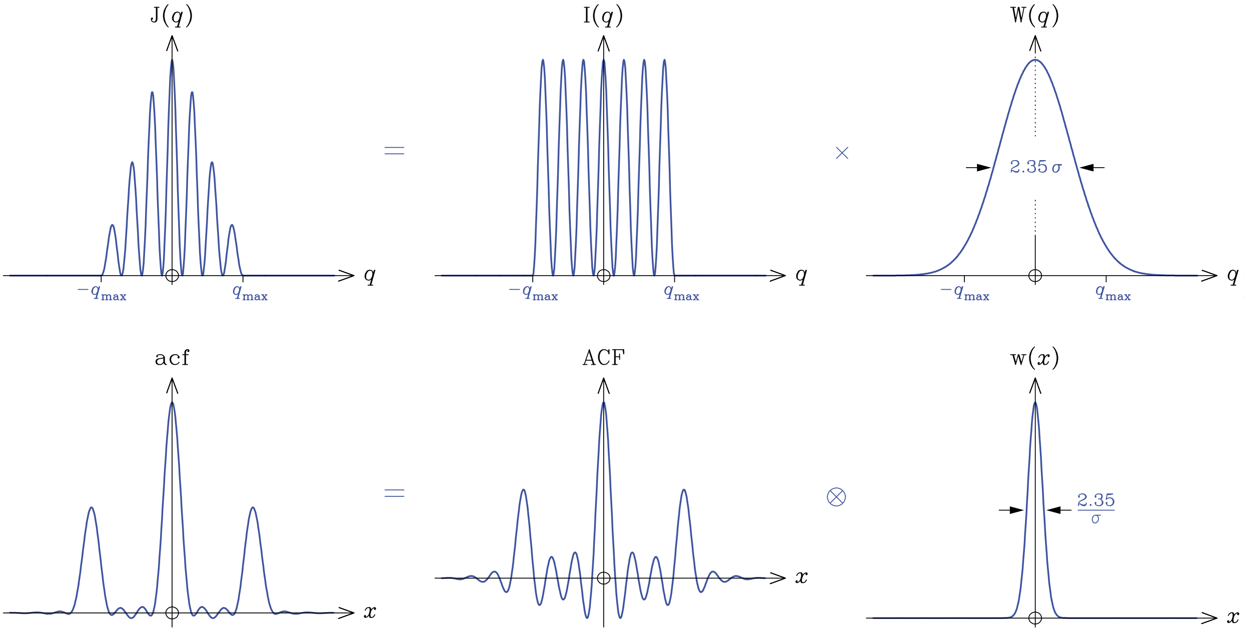


Fig. 2.22 Truncation ripples can be suppressed by multiplying the diffraction pattern, $\mathrm{I}(q)$, by a 'window' function, $\mathrm{W}(q)$, which decays smoothly to zero over a q-range comparable to that of the measurements, before calculating the (inverse) Fourier transform; the resultant 'filtered' auto-correlation function can also be understood from the convolution theorem.

to be narrower, the drawback is that intrinsically sharp features of the auto-correlation function are smeared out even more. Thus filtering is a matter of striking a balance between the suppression of the truncation ripples and a further loss of resolution. A variety of windowing functions have been developed to try to best achieve this end.

Diffraction measurements are often unattainable at low q-values as well as high ones, so that $\mathrm{I}(q)$ is available only within the range $q_{\min} \leqslant |q| \leqslant q_{\max}$. The truncated Fourier integral,

$$\int_{q_{\min}}^{q_{\max}} \mathrm{I}(q)\, \cos(qx)\, \mathrm{d}q\ ,$$

then yields an estimate of the ACF of the aperture function that is plagued by both low and high frequency artefacts. Structure in $\mathrm{A}(x)$ which is longer than around $2\pi/q_{\min}$ or narrower than about $2\pi/q_{\max}$ cannot be inferred reliably. The difficulty caused by a lack of $\mathrm{I}(0)$ is easiest to appreciate since, from eqns (2.57) and (2.58), it relates to the area under $\mathrm{A}(x)$:

$$\mathrm{I}(0)\ \propto\ \left|\ \int_{-\infty}^{\infty} \mathrm{A}(x)\, \mathrm{d}x \right|^2\ \propto\ \int_{-\infty}^{\infty} \mathrm{ACF}(x)\, \mathrm{d}x\ , \qquad \textbf{(2.69)}$$

where the equivalent expression on the far right follows from the x-integral of eqn (2.55), or the inverse of eqn (2.56) with k (or q) set to zero. As the truncated Fourier integral implicitly assumes that $\mathrm{I}(0)=0$ if $q_{\min} \neq 0$, the resultant ACF will contain equal amounts of positive and negative structure to ensure a net null area. Apart from at the origin, $q=0$, the diffraction pattern is insensitive to the addition of a constant to $\mathrm{A}(x)$ or its ACF.

2.6.3 Noise and probability theory

In practice, the analysis of a diffraction pattern is also limited by the *noise* in the measurement process and the extent to which the details of the experimental setup are understood and modelled. The task is not really one of calculating an inverse Fourier transform, which isn't possible in a strict mathematical sense, but a matter of making *inferences* about the aperture function given incomplete and noisy data. The tool for dealing with and quantifying uncertainty is *probability theory*, as developed by Laplace (1812), and the reader is referred to Sivia (1996) for an extended tutorial. A brief overview is given below.

Data analysis: a Bayesian tutorial, Sivia (1996), Oxford University Press; 2nd edition (2006) with Skilling.

The generic data analysis problem can be stated as follows: Given a set of N measurements $\{D_k\}$, for $k=1,2,3,\ldots,N$, and some pertinent information H, what can we infer about the object of interest $\mathrm{A}(x)$? The Fourier nature of the experiment enters the analysis through the equation that predicts the k^{th} data point, F_k, for a given aperture $\mathrm{A}(x)$:

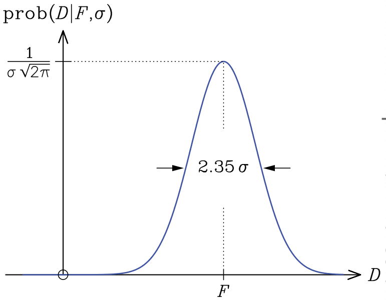
$$F_k \;=\; \mathrm{f}\Big(\mathrm{I}(q)\,,k\Big)\,, \tag{2.70}$$

where

$$\mathrm{I}(q) \;=\; \left|\,\psi_{\mathrm{o}}\int\limits_{-\infty}^{\infty}\mathrm{A}(x)\;\mathrm{e}^{\mathrm{i}qx}\;\mathrm{d}x\,\right|^2 \tag{2.71}$$

and ‘f’ is the function that models the measurement process. In the simplest case $F_k = \mathrm{I}(q_k)$, but a more common situation involves the convolution of $\mathrm{I}(q)$ with an instrumental *response*, or *resolution*, function, $\mathrm{R}(q)$, and the addition of a slowly varying background signal, $\mathrm{B}(q)$:

$$F_k \;=\; \int\limits_{-\infty}^{\infty}\mathrm{I}(q)\;\mathrm{R}(q_k - q)\;\mathrm{d}q \;\;+\;\mathrm{B}(q_k)\;. \tag{2.72}$$



The noise, or the expected mismatch between F_k and D_k, is usually quantified through an *error-bar*, σ_k, which is a shorthand way of assigning a Gaussian probability for the likelihood of the k^{th} datum:

$$\mathrm{prob}\Big(\boldsymbol{D}_k\,\Big|\,\mathrm{A}(x)\,,H\Big) \;=\; \frac{1}{\sigma_k\sqrt{2\pi}}\;\exp\!\left[-\,\frac{\left(\boldsymbol{D}_k - \boldsymbol{F}_k\right)^2}{2\,\sigma_k^{\,2}}\right], \tag{2.73}$$

where the vertical bar ‘$|$’ means ‘given’ (so that all items to the right of this conditioning symbol are taken as being true) and the comma is read as the conjunction ‘and’. A knowledge of eqns (2.70)–(2.72) and, hopefully, the related resolution and background functions, as well as the error-bars, is implicitly assumed in H. If the N measurements, $\{D_k\}$, are *independent*, in that the noise associated with one is unrelated to that of another (as far as H is concerned), then their joint likelihood is just the product of the individual contributions:

$$\mathrm{prob}\Big(\{\boldsymbol{D}_k\}\,\Big|\,\mathrm{A}(x)\,,H\Big) \;=\; \prod_{k=1}^{N}\;\mathrm{prob}\Big(\boldsymbol{D}_k\,\Big|\,\mathrm{A}(x)\,,H\Big)\;.$$

In conjunction with eqn (2.73), therefore, the *likelihood function* for the data can be written as

$$\mathrm{prob}\Big(\{\boldsymbol{D}_k\}\,\Big|\,\mathrm{A}(x)\,,H\Big) \;\propto\; \exp\!\left(-\,\frac{\chi^2}{2}\right), \tag{2.74}$$

where

$$\chi^2 \;=\; \sum_{k=1}^{N}\left(\frac{F_k - D_k}{\sigma_k}\right)^2 \tag{2.75}$$

is the sum of the squares of the *normalized residuals*.

Our inference, or ‘state of knowledge’, about the aperture function in the light of the data and H is not encapsulated by the likelihood function but by the *posterior probability*,

$$\mathrm{prob}\Big(\mathrm{A}(x)\,\Big|\,\{\boldsymbol{D}_k\},H\Big)\ ,$$

where the positions of $\{\boldsymbol{D}_k\}$ and $\mathrm{A}(x)$ are reversed with respect to the conditioning symbol. The $\mathrm{A}(x)$ which gives the largest value for the posterior probability can be regarded as the ‘best’ estimate of the aperture function, while the range of the alternatives that yield a reasonable fraction of the maximum probability gives an indication of the uncertainty. The likelihood function is related to the posterior probability through *Bayes’ theorem*,

$$\mathrm{prob}\Big(\mathrm{A}(x)\,\Big|\,\{\boldsymbol{D}_k\},H\Big) \;=\; \frac{\mathrm{prob}\Big(\{\boldsymbol{D}_k\}\,\Big|\,\mathrm{A}(x),H\Big)\times\mathrm{prob}\Big(\mathrm{A}(x)\,\Big|\,H\Big)}{\mathrm{prob}\Big(\{\boldsymbol{D}_k\}\,\Big|\,H\Big)}\ ,$$

where the second term in the numerator is called the *prior probability*, and represents our state of knowledge (or ignorance) about the aperture function before the analysis of the data, and the denominator usually constitutes an uninteresting proportionality constant (required for normalization) since it doesn’t explicitly mention $\mathrm{A}(x)$. The latter plays a crucial role when comparing different assumptions or models, however, such as H_1 versus H_2, and is referred to as the ‘global likelihood’, ‘prior predictive’ or simply the *evidence* in that context.

A quantitative discussion of the aperture function is contingent on a parametric description of $\mathrm{A}(x)$, of course, and its choice is a reflection of the information H at hand. If it were known that we were dealing with a pair of slits of equal finite width, as in Fig. 2.18 for example, then $\mathrm{A}(x)$ would be defined by the two parameters d and w as follows:

$$\mathrm{A}(x) \;=\; \begin{cases} 1 & \text{if } \left|x\pm\frac{d}{2}\right|\leqslant\frac{w}{2}\,, \\ 0 & \quad\text{otherwise}\,, \end{cases}$$

where $d>w>0$. If very little information was available, then we might use the formulation

$$\mathrm{A}(x) \;=\; \sum_{j=1}^{M} c_j\,\mathrm{G}_j(x)\ ,$$

where the M coefficients, $\{c_j\}$, define the aperture function through a linear combination of suitable *basis functions*, $\mathrm{G}_j(x)$. Although a larger value of M provides greater flexibility in the range of $\mathrm{A}(x)$ that can be modelled, a more careful choice of the $\mathrm{G}_j(x)$ can reduce the number required and, thereby, aid many aspects of the data analysis task.

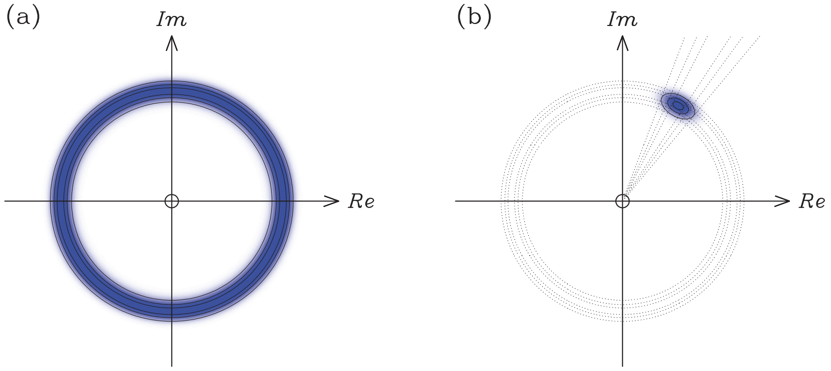


Fig. 2.23 The likelihood constraint on the value of a Fourier coefficient given (a) a noisy measurement of its modulus-squared and (b) additional phase information.

When the aperture function is defined by only a few parameters, the data tend to impose a strong constraint on their allowed values. The likelihood function dominates the posterior probability in this case and the prior, which is relatively broad to represent ignorance, is largely irrelevant. For the likelihood function of eqns (2.74) and (2.75), therefore, the optimal parameters of $\mathrm{A}(x)$ are those that yield the smallest value of χ^2; this is called the *least-squares* estimate. When little is known about the aperture function beforehand, and its description entails a large number of parameters to reflect this initial ignorance, it becomes important to give due consideration to the prior to encode whatever weak information is available about $\mathrm{A}(x)$. For example, positivity, bounds, local smoothness and so on. This leads to the use of *regularization* procedures, or constrained optimization, such as *maximum entropy*.

The computational task of finding the maximum of the posterior probability distribution and determining its spread, in the space of the parameters used to describe $\mathrm{A}(x)$, can be a very challenging one. If we have a good initial estimate of the optimal solution, then an efficient gradient algorithm, such as *Newton–Raphson*, can often be employed. Otherwise we may need to use the slower, but more robust, *Monte Carlo* methods. These sorts of practical considerations can make it tempting to ignore the noise and limited coverage of the data, and try to emulate an inverse Fourier transform in some way. For the ACF, and with appropriate filtering, this can provide a useful quick method for a qualitative analysis.

As mentioned earlier, the loss of the Fourier phase in diffraction experiments causes a serious difficulty for ascertaining the aperture function. We can begin to appreciate this from a probabilistic point of view by considering the constraint that the likelihood function imposes on the value of a Fourier coefficient, $\psi(q_k)$, when only its modulus-squared can be measured; this is shown pictorially in Fig. 2.23(a). Unlike the case of Fig. 2.23(b), where additional phase information is available, the permissible values of $\psi(q_k)$ do not shrink towards a unique point in the Argand plane even in the limit of noiseless data; they reduce instead to a thin circular region, with a phase ambiguity of 2π radians.

2.7 A list of useful formulae

To finish off this principally mathematical chapter, covering the important prerequisites for scattering theory, we give a list of some useful formulae.

Powers and logarithms *(see Section 1.1)*

$$a^M a^N = a^{M+N}\,, \qquad (a^M)^N = a^{MN}\,,$$

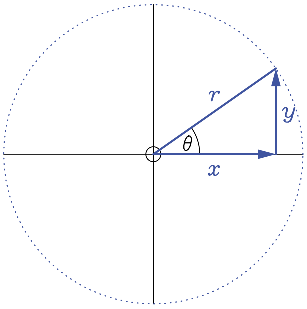
$$a^0 = 1\,, \qquad a^{-N} = 1/a^N \quad \text{and} \quad a^{1/p} = \sqrt[p]{a} \;\; \text{(integer p)}\,.$$

$$y = a^x \iff x = \log_a(y)$$

$$\log(AB) = \log(A) + \log(B) \quad \text{and} \quad \log(A/B) = \log(A) - \log(B)\,,$$

$$\log(A^\beta) = \beta\log(A) \quad \text{and} \quad \log_b(A) = \log_a(A) \times \log_b(a)\,.$$

Trigonometry



$$\sin\theta = \frac{y}{r} = \frac{1}{\operatorname{cosec}\theta}\,, \quad \cos\theta = \frac{x}{r} = \frac{1}{\sec\theta}\,, \quad \tan\theta = \frac{y}{x} = \frac{\sin\theta}{\cos\theta} = \frac{1}{\cot\theta}\,.$$

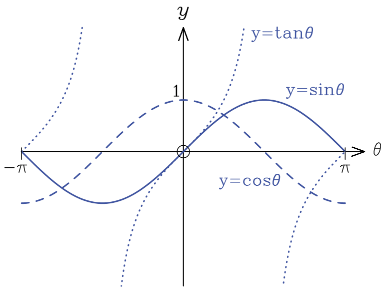
$$x^2 + y^2 = r^2 \iff \sin^2\theta + \cos^2\theta \equiv 1$$

$$\tan^2\theta + 1 \equiv \sec^2\theta$$

$$\cot^2\theta + 1 \equiv \operatorname{cosec}^2\theta$$

$$\sin(A \pm B) = \sin A\cos B \pm \cos A\sin B \implies \sin 2\theta = 2\sin\theta\cos\theta$$

$$\cos(A \pm B) = \cos A\cos B \mp \sin A\sin B \implies \cos 2\theta = \cos^2\theta - \sin^2\theta$$



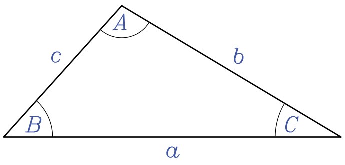
$$2\sin A\cos B = \sin(A{+}B) + \sin(A{-}B)$$

$$2\cos A\cos B = \cos(A{+}B) + \cos(A{-}B)$$

$$-2\sin A\sin B = \cos(A{+}B) - \cos(A{-}B)$$

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

$$c^2 = a^2 + b^2 - 2ab\cos C$$



Power series, sums and expansions *(see Section 2.2.4)*

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

$$n! = n\times(n-1)\times(n-2)\times\cdots\times3\times2\times1$$

$$\mathrm{e}^x = \exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

$$\mathrm{e} = \mathrm{e}^1 = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \cdots = 2.718\ldots$$

$$\log_{\mathrm{e}}(1+x) = \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \cdots \qquad \big(|x|<1\big)$$

$$(1+x)^p = 1 + p\,x + \frac{p\,(p-1)}{2!}\,x^2 + \frac{p\,(p-1)(p-2)}{3!}\,x^3 + \cdots \qquad \big(|x|<1\big)$$

$$\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + \cdots$$

$$(a+b)^n = \sum_{k=0}^{n} {}^nC_k\; a^k\; b^{n-k}$$

$${}^nC_k = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}\,, \qquad 0! = 1$$

$$\sum_{k=1}^{n} a + (k-1)\,d = \frac{n}{2}\Big[2\,a + (n-1)\,d\Big]$$

$$1 + 2 + 3 + \cdots + N = \frac{N\,(N+1)}{2}$$

$$\sum_{k=1}^{n} a\,r^{k-1} = \frac{a\,(1-r^n)}{1-r} \quad\longrightarrow\quad \frac{a}{1-r} \quad \text{as } n\to\infty \text{ and } |r|<1$$

Vectors *(see Section 2.1.1)*

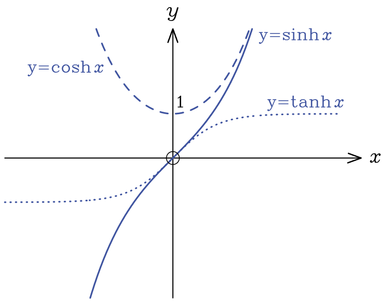
$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\bullet\mathbf{c})\,\mathbf{b} - (\mathbf{a}\bullet\mathbf{b})\,\mathbf{c}$$

Complex numbers *(see Section 2.2)*

$$\sinh x = \frac{\mathrm{e}^x - \mathrm{e}^{-x}}{2} = -\,\mathbf{i}\,\sin(\mathbf{i}\,x)$$

$$\cosh x = \frac{\mathrm{e}^x + \mathrm{e}^{-x}}{2} = \cos(\mathbf{i}\,x)$$

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{\mathrm{e}^{2x} - 1}{\mathrm{e}^{2x} + 1}$$



Differentiation and integration *(see Sections 2.1.2 and 2.3.1)*

$$y'' = \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)$$

$$\frac{\mathrm{d}^n y}{\mathrm{d}x^n} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}^{n-1} y}{\mathrm{d}x^{n-1}}\right) , \qquad y' = \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{\mathrm{d}x/\mathrm{d}y} ,$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(uv) = u\frac{\mathrm{d}v}{\mathrm{d}x} + v\frac{\mathrm{d}u}{\mathrm{d}x} , \qquad \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{u}{v}\right) = \frac{v\,u' - u\,v'}{v^2} ,$$

$$(uv)'' = u\,v'' + 2\,u'v' + u''v$$

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}(uv) = \sum_{k=0}^{n} {}^{n}C_k \, \frac{\mathrm{d}^k u}{\mathrm{d}x^k}\,\frac{\mathrm{d}^{n-k} v}{\mathrm{d}x^{n-k}} , \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u} \times \frac{\mathrm{d}u}{\mathrm{d}x} = \frac{\mathrm{d}y/\mathrm{d}t}{\mathrm{d}x/\mathrm{d}t} .$$

$$\boxed{\frac{\mathrm{df}}{\mathrm{d}x} = \mathrm{g}(x) \iff \int_a^b \mathrm{g}(x)\,\mathrm{d}x = \Big[\mathrm{f}(x) + c\Big]_a^b = \mathrm{f}(b) - \mathrm{f}(a)}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{a(t)}^{b(t)} \mathrm{g}(x)\,\mathrm{d}x = \mathrm{g}(b)\,\frac{\mathrm{d}b}{\mathrm{d}t} - \mathrm{g}(a)\,\frac{\mathrm{d}a}{\mathrm{d}t}$$

$$\int u\,\frac{\mathrm{d}v}{\mathrm{d}x}\,\mathrm{d}x = u\,v - \int v\,\frac{\mathrm{d}u}{\mathrm{d}x}\,\mathrm{d}x \quad \text{and} \quad \int \mathrm{g}(u)\,\frac{\mathrm{d}u}{\mathrm{d}x}\,\mathrm{d}x = \int \mathrm{g}(u)\,\mathrm{d}u$$

$$y = \sin\theta \iff \theta = \sin^{-1} y$$

$\mathrm{f}(x)$	$\frac{\mathrm{df}}{\mathrm{d}x}$	$\mathrm{f}(x)$	$\frac{\mathrm{df}}{\mathrm{d}x}$	$\mathrm{f}(x)$	$\frac{\mathrm{df}}{\mathrm{d}x}$
x^n	$n\,x^{n-1}$	$\ln x$	$1/x$	$\sinh x$	$\cosh x$
e^x	e^x	$\log_a(x)$	$(x \ln a)^{-1}$	$\cosh x$	$\sinh x$
a^x	$a^x \ln a$	e^{-x^2}	$-2\,x\,\mathrm{e}^{-x^2}$	$\tanh x$	$\mathrm{sech}^2\, x$
$\sin x$	$\cos x$	$\sin^{-1} x$	$\left(1-x^2\right)^{-1/2}$	$\sinh^{-1} x$	$\left(1+x^2\right)^{-1/2}$
$\cos x$	$-\sin x$	$\cos^{-1} x$	$-\left(1-x^2\right)^{-1/2}$	$\cosh^{-1} x$	$\left(x^2-1\right)^{-1/2}$
$\tan x$	$\sec^2 x$	$\tan^{-1} x$	$\left(1+x^2\right)^{-1}$	$\tanh^{-1} x$	$\left(1-x^2\right)^{-1}$

$$\mathrm{erf}(-x) = -\,\mathrm{erf}(x)$$

$$\int_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t = \frac{\sqrt{\pi}}{2}\,\mathrm{erf}(x) \quad \text{and} \quad \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}(x-x_\mathrm{o})t}\,\mathrm{d}t = 2\pi\,\delta(x-x_\mathrm{o})$$

$$\mathrm{erf}(\infty) = 1$$

(see Section 2.4.1)

Fourier transforms *(see Sections 2.4 and 2.5.5)*

$$\mathrm{F}(k) \;=\; \int_{-\infty}^{\infty} \mathrm{f}(x)\, \mathrm{e}^{-\mathrm{i}kx}\, \mathrm{d}x \quad\Longleftrightarrow\quad \mathrm{f}(x) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{F}(k)\, \mathrm{e}^{\mathrm{i}kx}\, \mathrm{d}k$$

$$\int_{-\infty}^{\infty} \big|\mathrm{f}(x)\big|\, \mathrm{d}x \;<\; \infty$$

$\mathrm{f}(x)$	$\mathrm{F}(k)$	$\mathrm{f}(x)$	$\mathrm{F}(k)$	
$\mathrm{f}(x+x_\mathrm{o})$	$\mathrm{e}^{\mathrm{i}kx_\mathrm{o}}\,\mathrm{F}(k)$	$\dfrac{\mathrm{d}\mathrm{f}}{\mathrm{d}x}$	$\mathrm{i}k\,\mathrm{F}(k)$	$\mathrm{f}(x)=0$ for $x=\pm\infty$
$\mathrm{f}(x)\otimes\mathrm{g}(x)$	$\mathrm{G}(k)\,\mathrm{F}(k)$	$\mathrm{f}(x)\,\mathrm{g}(x)$	$\frac{1}{2\pi}\,\mathrm{G}(k)\otimes\mathrm{F}(k)$	
$\delta(x-x_\mathrm{o})$	$\mathrm{e}^{-\mathrm{i}kx_\mathrm{o}}$	1	$2\pi\,\delta(k)$	
$\delta(x+d)+\delta(x-d)$	$2\cos(kd)$	$\delta(x+d)-\delta(x-d)$	$2\mathrm{i}\sin(kd)$	
$\begin{cases} 1 & \text{if } \lvert x\rvert\leqslant w \\ 0 & \text{otherwise} \end{cases}$	$\frac{2}{k}\sin(kw)$	$\displaystyle\sum_{n=-\infty}^{\infty}\delta(x-nd)$	$\displaystyle\sum_{m=-\infty}^{\infty}\delta\left(k-\frac{2\pi m}{d}\right)$	
$\begin{cases} \mathrm{e}^{-ax} & \text{for } x\geqslant 0 \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{1}{a+\mathrm{i}k}$	$\frac{a}{\pi(a^2+x^2)}$	$\mathrm{e}^{-a\lvert k\rvert}$	$a>0$
$\frac{1}{\sigma\sqrt{2\pi}}\exp\!\left(-\frac{x^2}{2\sigma^2}\right)$	$\mathrm{e}^{-\sigma^2k^2/2}$			

2.7.1 Dimensional analysis

Theoretical analysis involves the use of equations for understanding physical phenomena in a quantitative manner. Since the derivation of the relationships can be mathematically complicated, it's always worth carrying out 'sanity checks' on the formulae before applying them in detailed calculations; this is a good way of detecting algebraic mistakes and typographical errors. A requirement to simplify to familiar or intuitive results in elementary cases is one part of this approach, but the need for 'dimensional consistency' provides an even more basic test.

Physical parameters related to mechanics can be analysed in terms of their associated dimensions of length, L, time, T, and mass, M. Thus velocity, being a displacement per unit time, has dimensions of $L\,T^{-1}$; acceleration, being the rate of change of velocity, $L\,T^{-2}$; force, from Newton's second law of motion, $M\,L\,T^{-2}$; energy, from $work = force \times distance$, $M\,L^2\,T^{-2}$; and so on. While it may be necessary to add charge, Q, and temperature, Θ, to the basic list for dealing with electromagnetism and thermodynamics, the balance implied by an $=$ symbol means that the dimensions on both sides of an equation must match up (or else something has gone wrong). Indeed, the dimensions of every component separated by a $+$ or $-$ must be the same, and the arguments of functions, such as $\exp$, $\log$, $\sin$ and $\cos$, should be dimensionless.