Roger Penrose

THE ROAD TO REALITY

A Complete Guide to the Laws of the Universe

JONATHAN CAPE
LONDON

9
Fourier decomposition and hyperfunctions

9.1 Fourier series

Let us return to the question, raised in §6.1, of what Euler and his contemporaries might have regarded as an acceptable notion of 'honest function'. In §7.1, we settled on the holomorphic (complex-analytic) functions as best satisfying what Euler might well have had in mind. Yet, most mathematicians today would regard such a notion of a 'function' as being unreasonably restrictive. Who is right? We shall be coming to a very remarkable answer to this question at the end of this chapter. But first let us try to understand what the issues are.

In the application of mathematics to problems of the physical world, it is a frequent requirement that there be a flexibility that neither the holomorphic functions nor their real counterparts—the analytic (i.e. C^{ω}-) functions—appear to possess. Because of the uniqueness of analytic continuation, as described in §7.4, the global behaviour of a holomorphic function defined throughout some connected open region $\mathcal{D}$ of the complex plane, is completely fixed, once it is known in some small open subregion of $\mathcal{D}$. Similarly, an analytic function of a real variable, defined on some connected segment $\mathcal{R}$ of the real line $\mathbb{R}$ is also completely fixed once the function is known in some small open subregion of $\mathcal{R}$. Such rigidity seems inappropriate for the realistic modelling of physical systems.

It would be particularly awkward when the propagation of *waves* is under consideration. Wave propagation, which includes the sending of signals via the electromagnetic vibrations of radio waves or light, gains much of its utility from the fact that information can be transmitted by such means. The whole point of signalling, after all, is that there must be the potential for sending a message that might be unexpected by the receiver. If the form of the signal has to be given by an analytic function, then there is not the possibility of 'changing one's mind' in the middle of the message. Any small part of the signal would completely fix the signal in its entirety for all time. Indeed, wave propagation is frequently studied in terms of the question as to how discontinuities, or other deviations from analyticity, will actually propagate.

Let us consider waves and ask how such things are described mathematically. One of the most effective ways of studying wave forms is through the procedure known as *Fourier analysis*. Joseph Fourier was a French mathematician who lived from 1768 until 1830. He had been concerned with the question of decomposing periodic vibrations into their component 'sine-wave' parts. In music, this is basically what is involved in representing some musical sound in terms of its constituent 'pure tones'. The term 'periodic' means that the pattern (say of physical displacements of the object which is vibrating) exactly repeats itself after some period of time, or it could refer to periodicity in space, like the repeating patterns in a crystal or on wallpaper or in waves in the open sea. Mathematically, we say that a function f (say[1] of a real variable χ) is *periodic* if, for all χ, it satisfies

$$f(\chi + l) = f(\chi),$$

where l is some fixed number referred to as the *period*. Thus, if we 'slide' the graph of $y = f(\chi)$ along the χ-axis by an amount l, it looks just the same as it did before (Fig. 9.1a). (The way in which Fourier handled functions that need *not* be periodic—by use of the Fourier *transform*—will be described in §9.4.)

The 'pure tones' are things like $\sin\chi$ or $\cos\chi$ (Fig. 9.1b). These have period 2π, since

$$\sin(\chi + 2\pi) = \sin\chi, \qquad \cos(\chi + 2\pi) = \cos\chi,$$

these relations being manifestations of the periodicity of the single complex quantity $e^{i\chi} = \cos\chi + i\sin\chi$,

$$e^{i(\chi+2\pi)} = e^{i\chi},$$

which we encountered in §5.3. If we want periodicity l, rather than 2π, then we can 'rescale' the χ as it appears in the function, and take $e^{i2\pi\chi/l}$ instead of $e^{i\chi}$. The real and imaginary parts $\cos(2\pi\chi/l)$ and $\sin(2\pi\chi/l)$ will correspondingly also have period l. But this is not the only possibility. Rather than oscillating just once, in the period l, the function could oscillate twice, three times, or indeed n times, where n is any positive integer (see Fig. 9.1c), so we find that each of

$$e^{i\cdot 2\pi n\chi/l}, \quad \sin\left(\frac{2\pi n\chi}{l}\right), \quad \cos\left(\frac{2\pi n\chi}{l}\right)$$

has period l (in addition to having also a smaller period l/n). In music, these expressions, for $n = 2, 3, 4, \ldots$, are referred to as *higher harmonics*.

One problem that Fourier addressed (and solved) was to find out how to express a general periodic function $f(\chi)$, of period l, as a sum of pure tones.

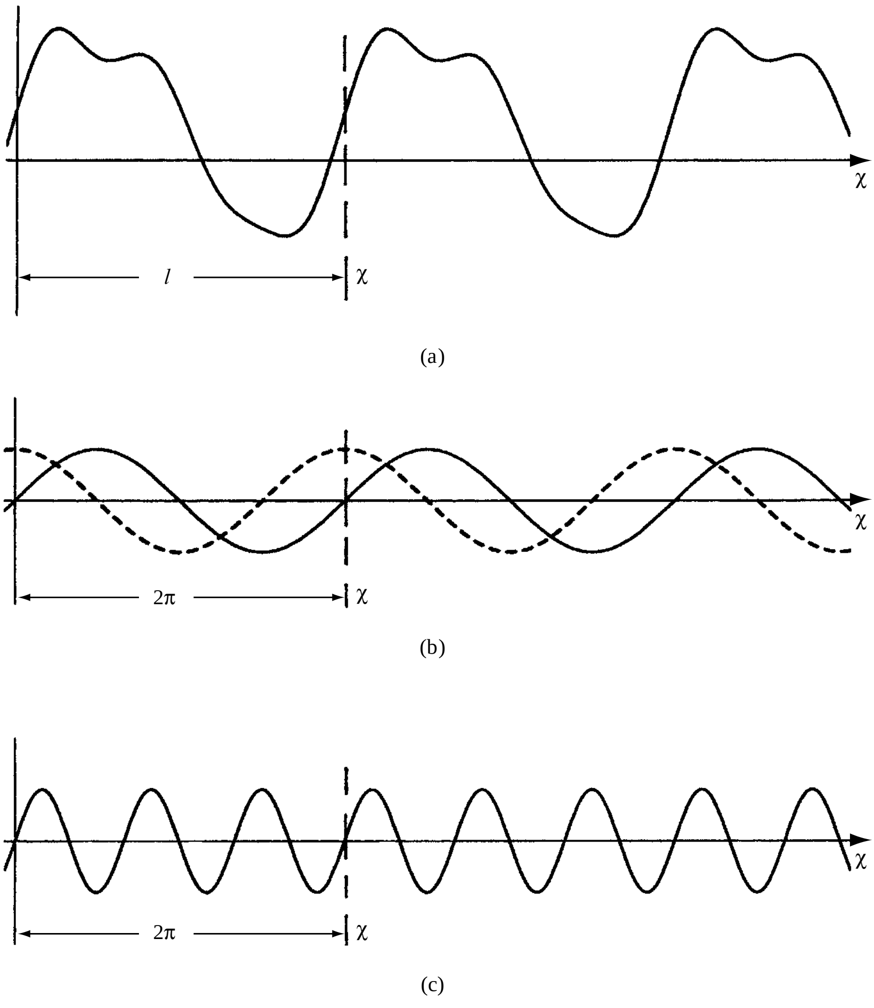

Fig. 9.1 Periodic functions. (a) $f(\chi)$ has period l if $f(\chi)=f(\chi+l)$ for all χ, meaning that if we slide the graph of $y=f(\chi)$ along the χ-axis by l, it looks just the same as before. (b) The basic 'pure tones' $\sin\chi$ or $\cos\chi$ (shown dotted) have period $l=2\pi$. (c) 'Higher harmonic' pure tones oscillate several times in the period l; they still have period l, while also having a shorter period ($\sin 3\chi$ is illustrated, having period $l=2\pi$ as well as the shorter period $2\pi/3$).

For each n, there will generally be a different magnitude of that pure tone's contribution to the total, and this will depend upon the wave form (i.e. upon the shape of the graph $y=f(\chi)$). Some simple examples are illustrated in Fig. 9.2. Usually, the number of different pure tones that contribute to $f(\chi)$ will be infinite, however. More specifically, what Fourier required was the

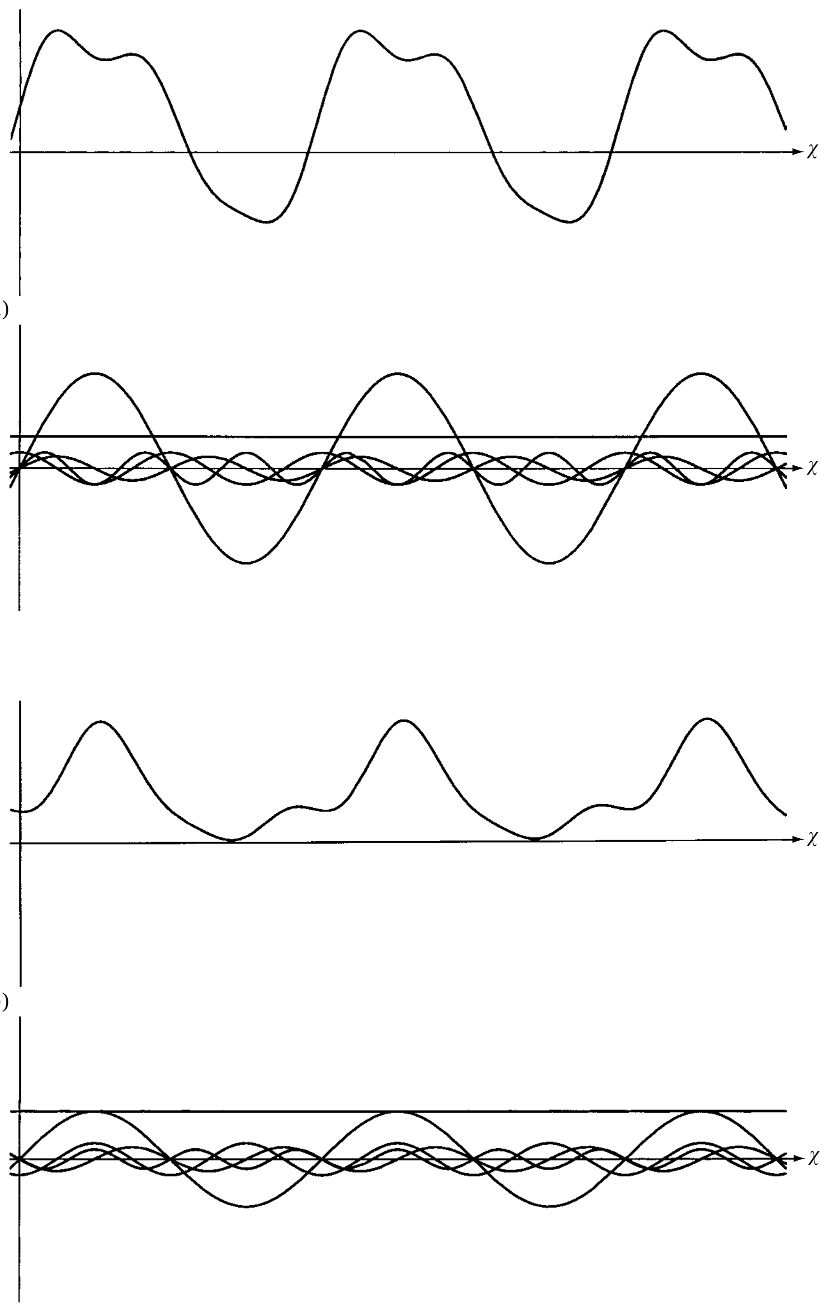

Fig. 9.2 Examples of Fourier decomposition of periodic functions. The wave form (shape of the graph) is determined by the Fourier coefficients. The functions and their individual Fourier components beneath. (a) $f(\chi) = \frac{2}{3} + 2\sin\chi + \frac{1}{3}\cos 2\chi + \frac{1}{4}\sin 2\chi + \frac{1}{3}\sin 3\chi$. (*b*) $f(\chi) = \frac{1}{2} + \sin\chi - \frac{1}{3}\cos 2\chi - \frac{1}{4}\sin 2\chi - \frac{1}{5}\sin 3\chi$.

collection of coefficients c, a_1, b_1, a_2, b_2, a_3, b_3, a_4, in the decomposition of $f(\chi)$ into its constituent pure tones, as given by the expression

$$f(\chi) = c + a_1 \cos \omega\chi + b_1 \sin \omega\chi + a_2 \cos 2\omega\chi + b_2 \sin 2\omega\chi + \\ a_3 \cos 3\omega\chi + b_3 \sin 3\omega\chi + \cdots,$$

where, in order to make the expressions look simpler, I have written them in terms of the *angular frequency* ω (nothing to do with the 'ω' of §§5.4,5, §8.1) given by $\omega = 2\pi/l$.

Some readers may well feel that this expression for $f(\chi)$ still looks unduly complicated—and such a reader is indeed correct. The formula actually looks a lot tidier if we incorporate the cos and sin terms together as complex exponentials $\left(e^{iA\chi} = \cos A\chi + i \sin A\chi\right)$, so that

$$f(\chi) = \cdots + \alpha_{-2}e^{-2i\omega\chi} + \alpha_{-1}e^{-i\omega\chi} + \alpha_0 + \alpha_1 e^{i\omega\chi} + \alpha_2 e^{2i\omega\chi} + \alpha_3 e^{3i\omega\chi} + \cdots,$$

where[2],[9.1]

$$a_n = \alpha_n + \alpha_{-n}, \qquad b_n = i\alpha_n - i\alpha_{-n}, \qquad c = \alpha_0$$

for $n = 1, 2, 3, 4, \ldots$. The expression looks even tidier if we put $z = e^{i\omega\chi}$, and define the function $F(z)$ to be just the same quantity as $f(\chi)$ but now expressed in terms of the new complex variable z. For then we get

$$F(z) = \cdots + \alpha_{-2}z^{-2} + \alpha_{-1}z^{-1} + \alpha_0 z^0 + \alpha_1 z^1 + \alpha_2 z^2 + \alpha_3 z^3 + \cdots,$$

where

$$F(z) = F(e^{i\omega\chi}) = f(\chi).$$

And we can make it look tidier still by using the summation sign $\sum$, which here means 'add together all the terms, for all integer values of r':

$$F(z) = \sum \alpha_r z^r.$$

This looks like a power series (see §4.3), except that there are *negative* as well as positive powers. It is called a *Laurent* series. We shall be seeing the importance of this expression in the next section.[9.2]

9.2 Functions on a circle

The Laurent series certainly gives us a very economical way of representing Fourier series. But this expression also suggests an interesting

[9.1] Show this.

[9.2] Show that when F is analytic on the unit circle the coefficients α_n, and hence the a_n, b_n, and c, can be obtained by use of the formula $\alpha_n = (2\pi i)^{-1} \oint z^{-n-1}F(z)\, dz$.

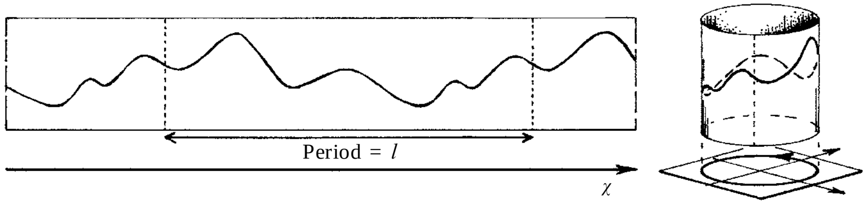

Fig. 9.3 A periodic function of a real variable χ may be thought of as defined on a circle of circumference l where we 'wrap up' the real axis of χ into the circle. With $l=2\pi$, we may take this circle as the unit circle in the complex plane.

alternative perspective on Fourier decomposition. Since a periodic function simply repeats itself endlessly, we may think of such a function (of a real variable χ) as being defined on a *circle* (Fig. 9.3), where the function's period l is the length of the circle's circumference, χ measuring distance around the circle. Rather than simply going off in a straight line, these distances now wrap around the circle, so that the periodicity is automatically taken into account.

For convenience (at least for the time being), I take this circle to be the unit circle in the complex plane, whose circumference is 2π, and I take the period l to be 2π. Accordingly,

$$\omega = 1, \quad \text{so } z = e^{i\chi}.$$

(For any other value of the period, all we need to do is to reinstate ω by rescaling the χ-variable appropriately.) The different cos and sin terms that represent the various 'pure tones' of the Fourier decomposition are now simply represented as positive or negative powers of z, namely $z^{\pm n}$ for the nth harmonics. On the unit circle, these powers just give us the oscillatory cos and sin terms that we require; see Fig. 9.4.

We now have this very tidy way of representing the Fourier decomposition of some periodic function $f(\chi)$. We think of $f(\chi) = F(z)$ as defined on the unit circle in the z-plane, with $z = e^{i\chi}$, and then the Fourier decomposition is just the Laurent series description of this function, in terms of a complex variable z. But the advantage is not just a matter of tidiness. This representation also provides us with deeper insights into the nature of Fourier series and of the kind of function that they can represent. More significantly for the eventual purpose of this book, it has important connections with quantum mechanics and, therefore, for our deeper understanding of Nature. This comes about through the magic of complex numbers, for we can also use our Laurent series expression when z lies away from the unit circle. It turns out that

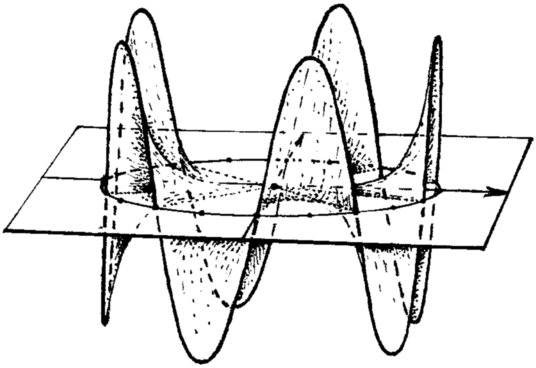

Fig. 9.4 On the unit circle, the real and imaginary parts of the function z^n appear as nth harmonic cos and sin waves (the real and imaginary parts of $e^{in\chi}$, respectively, where $z = e^{i\chi}$). Here, for $n = 5$, the real part of z^5 is plotted.

this series tells us something important about $F(z)$, for z lying *on* the unit circle, in terms of what the series does when z lies *off* the unit circle.

Now, let us recall (from §4.4) the notion of a circle of convergence, within which a power series converges and outside of which it diverges. There is a close analogue of this for a Laurent series: the *annulus of convergence*. This is the region lying strictly between two circles in the complex plane, both centred at the origin (see Fig. 9.5a). This is simple to understand once we have the notion of circle of convergence for an ordinary power series. The part of the series with positive powers,[3]

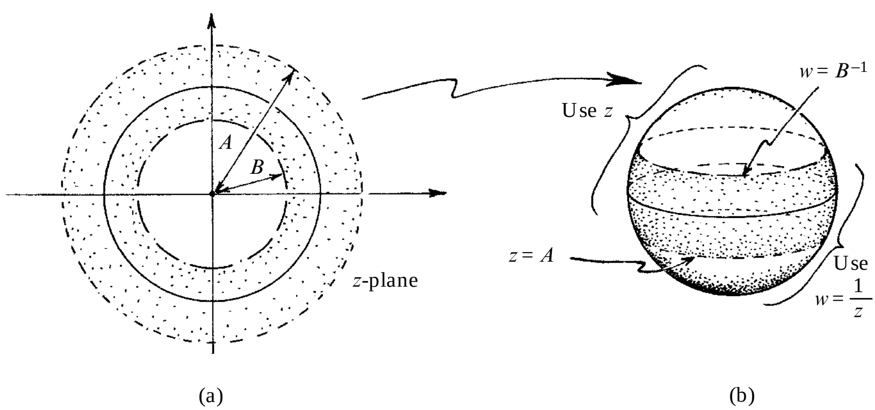

Fig. 9.5 (a) The annulus of convergence for a Laurent series $F(z) = F^+ + \alpha_0 + F^-$, where $F^+ = \cdots + \alpha_{-2}z^{-2} + \alpha_{-1}z^{-1}$, $F^- = \alpha_1 z^1 + \alpha_2 z^2 + \cdots$. The radius of convergence for F^+ is A and, in terms of $w = z^{-1}$, for F^- is B^{-1}. (b) The same, on the Riemann sphere (see Fig. 8.7), where z refers to the extended northern hemisphere and w $(= z^{-1})$ to the extended southern hemisphere.

$$F^- = \alpha_1 z^1 + \alpha_2 z^2 + \alpha_3 z^3 + \cdots,$$

will have an ordinary circle of convergence, of radius A, say, and that part of the series converges for all values of z whose modulus is less than A. With regard to the part of the series with negative powers, that is,

$$F^+ = \cdots + \alpha_{-3} z^{-3} + \alpha_{-2} z^{-2} + \alpha_{-1} z^{-1},$$

we can understand it as just an ordinary power series in the reciprocal variable $w = 1/z$. There will be a circle of convergence in the w-plane, of radius $1/B$, say, and that part of the series will converge for values of w whose modulus is smaller than $1/B$. (We are really talking about the Riemann sphere here, as described in Chapter 8—see Fig. 8.7, with the z-coordinate referring to one hemisphere and the w-coordinate referring to the other. See Fig. 9.5b. We shall explore the Riemann sphere aspect of this in the next section.) For values of z whose moduli are greater than B, therefore, the negative-power part of the series will converge. Provided that $B < A$, these two convergence regions will overlap, and we get the annulus of convergence for the entire Laurent series. Note that the whole Fourier or Laurent series for the function $f(\chi) = F(e^{i\chi}) = F(z)$ is given by

$$F(z) = F^+ + \alpha_0 + F^-,$$

where the additional constant term α_0 must be included.

In the present situation, we ask for convergence *on* the unit circle, since this is where we can have $z = e^{i\chi}$ for real values of χ, and the question of the convergence of our Fourier series for $f(\chi)$ is precisely the question of the convergence of the Laurent series for $F(z)$ when z lies on the unit circle. Thus, we seem to need $B < 1 < A$, ensuring that the unit circle indeed lies within the annulus of convergence. Does this mean that, for convergence of the Fourier series, we necessarily require the unit circle to lie within the annulus of convergence?

This would indeed be the case if $f(\chi)$ is analytic (i.e. C^ω); for then the function $f(\chi)$ can be extended to a function $F(z)$ that is holomorphic throughout some open region that includes the unit circle.[4] But, if $f(\chi)$ is not analytic, an interesting question arises. In this case, either the annulus of convergence shrinks down to become the unit circle itself—which, strictly speaking, is not allowed for a genuine annulus of convergence, because the annulus of convergence ought to be an open region, which the unit circle is not—or else the unit circle becomes the outer or inner boundary of the annulus of convergence. These questions will be important for us in §§9.6,7.

For the moment, let us not worry about what happens when $f(\chi)$ in not analytic, and consider the simpler situation that arises when $f(\chi)$ is analytic. Then we have the unit circle in the z-plane strictly contained within a genuine annulus of convergence for $F(z)$, this being bounded by circles (centred at the origin) of radii A and B, with $B < 1 < A$. The part of the Laurent series with positive powers, F^-, converges for points in the z-plane whose moduli are smaller than A and the part with negative powers, F^+, converges for points in the z-plane whose moduli are greater than B, so both converge within the annulus itself (and, in a very trivial sense, the constant term α_0 obviously 'converges' for all z). This provides us with a 'splitting' of the function $F(z)$ into two parts, one holomorphic inside the outer circle and the other holomorphic outside the inner circle, these being defined, respectively, by the series expressions for F^- and F^+.

There is a (mild) ambiguity about whether the constant term α_0 is to be included with F^- or with F^+ in this splitting. In fact, it is better just to live with this ambiguity. For there is a symmetry between F^- and F^+, which is made clearer if we adopt the Riemann sphere picture that was alluded to above (see Fig. 9.5b). This gives us a more complete picture of the situation, so let us explore this next.

9.3 Frequency splitting on the Riemann sphere

The coordinates z and w $(= 1/z)$ give us two patches covering the Riemann sphere. The unit circle becomes the equator of the sphere and the annulus is now just a 'collar' of the equator. We think of our splitting of $F(z)$ as expressing it as a sum of two parts, one of which extends holomorphically into the southern hemisphere—called the *positive-frequency* part of $F(z)$—as defined by $F^+(z)$, together with whatever portion of the constant term we choose to include, and the other, extending holomorphically into the northern hemisphere—called the *negative-frequency* part of $F(z)$—as defined by $F^-(z)$ and the remaining portion of the constant term. If we ignore the constant term, this splitting is uniquely determined by this holomorphicity requirement for the extension into one or other of the two hemispheres.[9.3]

It will be handy, from time to time, to refer to the 'inside' and the 'outside' of a circle (or other closed loop) drawn on the Riemann sphere by appealing to an *orientation* that is to be assigned to the circle. The standard orientation of the unit circle in the z-plane is given in terms of the direction of increase of the standard θ-coordinate, i.e. anticlockwise. If we reverse this orientation (e.g. replacing θ by $-\theta$), then we interchange positive with negative frequency. Our convention for a general closed loop is to be consistent with this. The orientation is anticlockwise if the 'clock face' is on the inside of the loop, so to speak, whereas it would be clockwise if the 'clock face' were to be placed on the outside of the loop. This serves to define the 'inside' and 'outside' of an oriented closed loop. Figure 9.6 should clarify the issue.

[9.3] Can you see why?

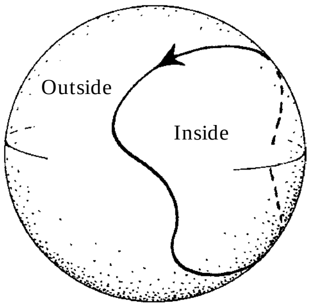

Fig. 9.6 An orientation assigned to a closed loop on the Riemann sphere defines its 'inside' and 'outside' as indicated: this orientation is anticlockwise for a 'clock face' inside the loop (and clockwise if outside).

This splitting of a function into its positive- and negative-frequency parts is a crucial ingredient of quantum theory, and most particularly of quantum field theory, as we shall be seeing in §24.3 and §§26.2–4. The particular formulation that I have given here is not quite the most usual way that this splitting is expressed, but it has some considerable advantages in a number of different contexts (particularly in twistor theory, for example; see §33.10). The usual formulation is not so concerned with holomorphic extensions as with the Fourier expansion directly. The positive-frequency components are those given by multiples of $e^{-in\chi}$, where n is positive, as opposed to those given by multiples of $e^{in\chi}$, which are negative-frequency components. A positive-frequency function is one composed entirely of positive-frequency components.

However, this description does not reveal the full generality of what is involved in this splitting. There are many holomorphic mappings of the Riemann sphere to itself which send each hemisphere to itself, but which do not preserve the north or south poles (i.e. the points $z = 0$ or $z = \infty$).[9.4] These preserve the positive/negative-frequency splitting but do not preserve the individual Fourier components $e^{-in\chi}$ or $e^{in\chi}$. Thus, the issue of the splitting into positive and negative frequencies (crucial to quantum theory) is a more general notion than the picking out of individual Fourier components.

In normal discussions of quantum mechanics, the positive/negative-frequency splitting refers to functions of *time* t, and we do not usually think of time as going round in a circle. But we can use a simple transformation to obtain the full range of t, from the 'past limit' $t = -\infty$ to the 'future limit' $t = \infty$, from a χ that goes once around the circle—here I take χ to range between the limits $\chi = -\pi$ and $\chi = \pi$ (so $z = e^{i\chi}$ ranges round the unit circle in the complex plane, in an anticlockwise direction, from the point $z = -1$ and back to $z = -1$ again; see Fig. 9.7). Such a transformation is given by

[9.4] Which are these mappings, explicitly?

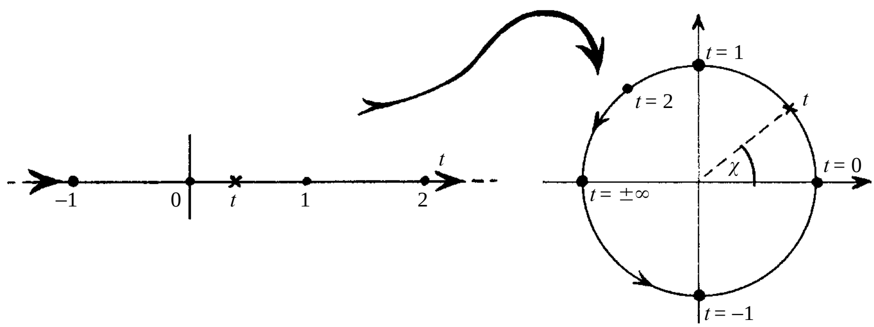

Fig. 9.7 In quantum mechanics, positive/negative-frequency splitting refers to functions of time t, not assumed periodic. The splitting of Fig. 9.5 can still be applied, for the full range of t (from $-\infty$ to $= +\infty$) if we use the transformation of relating t to $z(= e^{i\chi})$, where we go around unit circle, anticlockwise, from $z = -1$ and back to $z = -1$ again, so χ goes from $-\pi$ to π.

$$t = \tan\frac{1}{2}\chi.$$

The graph of this relationship is given in Fig. 9.8 and a simple geometrical description is provided in Fig. 9.9.

An advantage of this particular transformation is that it extends holomorphically to the entire Riemann sphere, this being a transformation that we already considered in §8.3 (see Fig. 8.8), which takes the unit circle (z-plane) into the real line (t-plane):[9.5]

$$t = \frac{z-1}{iz+i}, \quad z = \frac{-t+i}{t+i}.$$

The interior of the unit circle in the z-plane corresponds to the upper half-t-plane and the exterior of the z-unit circle corresponds to the lower half-t-plane. Hence, positive-frequency functions of t are those that extend holomorphically into the lower half-plane of t and negative-frequency ones, into the upper half-plane. (There is, however, a significant additional

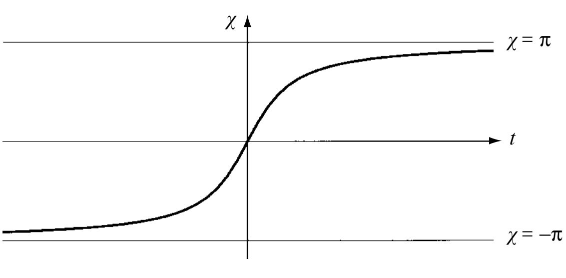

Fig. 9.8 Graph of $t = \tan\chi/2$.

[9.5] Show that this gives the same t as above.

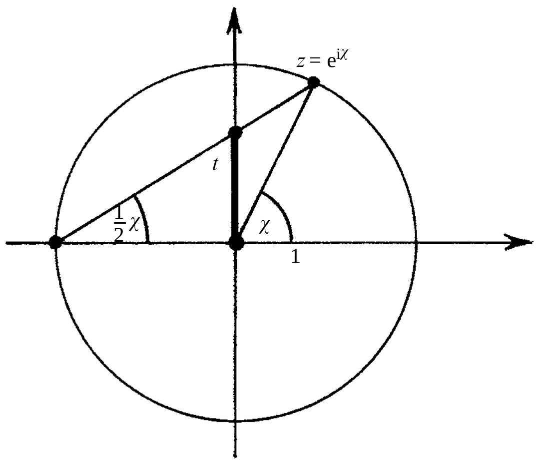

Fig. 9.9 Geometry of $t = \tan\frac{\chi}{2}$.

technicality that we have to be careful about how we deal with the point '∞' of the t-plane; but this is handled appropriately if we always think in terms of the Riemann sphere, rather than simply the complex t-plane.)

In standard presentations, however, the notion of 'positive frequency' in terms of a time-coordinate t, is not usually stated in the particular way that I have just presented it here, but rather in terms of what is called the *Fourier transform* of $f(\chi)$. The answer is actually the same[5] as the one that I have given, but since Fourier transforms are of crucial significance for quantum mechanics in any case (and also in many other areas), it will be important to explain here what this transform actually is.

9.4 The Fourier transform

Basically, a Fourier transform is the limiting case of a Fourier series when the period l of our periodic function $f(\chi)$ is taken to get larger and larger until it becomes infinite. In this infinite limit, there is no restriction of periodicity on $f(\chi)$ at all: it is just an ordinary function.[6] This has considerable advantages when we are studying wave propagation and the potential for sending of 'unexpected' signals. For then we do not want to insist that the form of the signal be periodic. The Fourier transform allows us to consider such 'one-off' signals, while still analysing them in terms of periodic 'pure tones'. It achieves this, in effect, by considering our function $f(\chi)$ to have period $l \to \infty$. As the period l gets larger, the pure-tone harmonics, having period l/n for some positive integer n, will get closer and closer to any positive real number we choose. (Recall that any real number can be approximated arbitrarily closely by rationals, for example.) What this tells us is that any pure tone of any frequency whatever is now allowed as a Fourier component. Rather than having $f(\chi)$ expressed as a discrete sum of Fourier components, we now have $f(\chi)$ expressed as a continuous sum over all frequencies, which means that $f(\chi)$ is now expressed as an *integral* (see §6.6) with respect to the frequency.

Let us see, in outline, how this works. First, recall our 'tidiest' expression for the Fourier decomposition of a periodic function $f(\chi)$, of period l, as given above:

$$F(z)=\sum \alpha_r z^r, \quad \text{where } z=e^{i\omega\chi}$$

(the angular frequency ω being given by $\omega = 2\pi/l$). Let us take the period to be initially 2π, so $\omega = 1$. Now we are going to try to increase the period by some large integer factor N (whence $l = 2\pi N$), so the frequency is reduced by the same factor (i.e. $\omega = N^{-1}$). The oscillatory wave that used to be the fundamental pure tone now becomes the Nth harmonic with respect to this new lower frequency. A pure tone that used to be an nth harmonic would now be an (nN)th harmonic. When we take the limit as N approaches infinity, it becomes inappropriate to try to keep track of a particular oscillatory component by labelling it by its 'harmonic number' (i.e. by the number n), because this number keeps changing. That is to say, it is inappropriate to label this oscillatory component by the integer r in the above sum because a fixed value of r labels a particular harmonic ($r = \pm n$ for the nth harmonic), rather than keeping track of a particular tone frequency. Instead, it is r/N that keeps track of this frequency, and we need a new variable to label this. Bearing in mind the important use that Fourier transforms are due to be put to in later chapters (see §21.11 particularly), I shall call this variable 'p' which, in the limit when N tends to infinity, stands for the *momentum*[7] of some quantum-mechanical particle whose position is measured by χ. In this limit, one may also revert to the conventional use of x in place of χ, if desired, as we shall find that χ actually does become the real part of z in the limit in the following descriptions.

For finite N, I write

$$p=\frac{r}{N}.$$

In the limit as $N \to \infty$, the parameter p becomes a continuous variable and, since the 'coefficients α_r' in our sum will then depend on the continuous real-valued parameter p rather that on the discrete integer-valued parameter r, it is better to write the dependence of the coefficients α_r on r by using the standard type of functional notation, say $g(p)$, rather than just using a suffix (e.g. g_p), as in α_r. Effectively, we shall make the replacement

$$\alpha_r \mapsto g(p)$$

in our summation $\sum \alpha_r z^r$, but we must bear in mind that, as N gets larger, the number of actual terms lying within some small range of p-values gets larger (basically in proportion to N, because we are considering fractions n/N that lie in that range). Accordingly, the quantity $g(p)$ is really a measure of density, and it must be accompanied by the differential quantity $\mathrm{d}p$ in the limit as the summation $\sum$ becomes an integral $\int$. Finally, consider the term z^r in our sum $\sum \alpha_r z^r$. We have $z = \mathrm{e}^{\mathrm{i}\omega\chi}$, with $\omega = N^{-1}$; so $z = \mathrm{e}^{\mathrm{i}\chi/N}$. Thus $z^r = \mathrm{e}^{\mathrm{i}\chi/N} = \mathrm{e}^{\mathrm{i}\chi p}$; so putting these things together, in the limit as $N \to \infty$, we get the expression

$$\sum \alpha_r z^r \to \int_{-\infty}^{\infty} g(p)\mathrm{e}^{\mathrm{i}\chi p}\,\mathrm{d}p$$

to represent our function $f(\chi)$. In fact it is usual to include a scaling factor of $(2\pi)^{-1/2}$ with the integral, for then there is the remarkable symmetry that the *inverse* relation, expressing $g(p)$ in terms of $f(\chi)$ has exactly the same form (apart from a minus sign) as that which expresses $f(\chi)$ in terms of $g(p)$:

$$f(\chi) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} g(p)\mathrm{e}^{\mathrm{i}\chi p}\,\mathrm{d}p, \quad g(p) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} f(\chi)\mathrm{e}^{-\mathrm{i}\chi p}\,\mathrm{d}\chi.$$

The functions $f(\chi)$ and $g(p)$ are called *Fourier transforms* of one another.[9.6]

9.5 Frequency splitting from the Fourier transform

A (complex) function $f(\chi)$, defined on the entire real line, is said to be of *positive frequency* if its Fourier transform $g(p)$ is zero for all $p \geqslant 0$. Thus, $f(\chi)$ is composed only of components of the form $\mathrm{e}^{\mathrm{i}\chi p}$ with $p < 0$. (Euler might well have worried—see §6.1—about such a $g(p)$, which seems to be a blatant 'gluing job' between a non-zero function for $p < 0$ and simply zero for $p > 0$. Yet this seems to be representing a perfectly respectable 'holomorphic' property of $f(\chi)$. Another way of expressing this 'positive-frequency' condition is in terms of the holomorphic extendability of $f(\chi)$, as we did before for Fourier series. Now we think of the variable χ as labelling the points on the real axis (so we can take $\chi = x$ on this axis), where on the Riemann sphere this 'real axis' (including the point '$\chi = \infty$') is now the *real circle* (see Fig. 8.9c). This circle divides the sphere into two hemispheres, the 'outside' one being that which is the lower half-plane in the standard picture of the complex plane. The condition that $f(\chi)$ be of positive frequency is now that it extend holomorphically into this outside hemisphere.

There is one issue that requires some care, however, when we compare these two definitions of 'positive frequency'. This relates to the question of

[9.6] Show (in outline) how to obtain the expression for $g(p)$ in terms of $f(\chi)$ using a limiting form of the contour integral expression $\alpha_n = (2\pi\mathrm{i})^{-1}\oint z^{-n-1}F(z)\mathrm{d}z$ of Exercise [9.2].

how we treat the point $z = \infty$, since the function $f(\chi)$ will in general have some kind of singularity there. In fact, provided that we adopt the 'hyper-functional' point of view that I shall be describing shortly (in §9.7), this singularity at $z = \infty$ presents us with no essential difficulty. With the appropriate point of view with regard to '$f(\infty)$', it turns out that the two definitions of positive frequency that I gave in the previous paragraph are in basic agreement with each other.[8]

For the interested reader, it may be helpful to examine, in terms of the Riemann sphere, some of the geometry that is involved in our limit of §9.4, taking us from Fourier series to Fourier transform. Let us return to the z-plane description that we had been considering earlier, for a function $f(\chi)$ of period 2π, where χ measures the arc length around a unit-radius circle. Suppose that we wish to change the period to values larger than 2π, in successively increasing steps, while retaining the interpretation of χ as a distance around a circle. We can achieve this by considering a sequence of larger and larger circles, but in order for the limiting procedure to make geometric sense we shall suppose that the circles are all touching each other at the starting point $\chi = 0$ (see Fig. 9.10a). For simplicity in what follows, let us choose this point to be the origin $z = 0$ (rather than $z = 1$), with all the circles lying in the lower half-plane. This makes our initial circle,

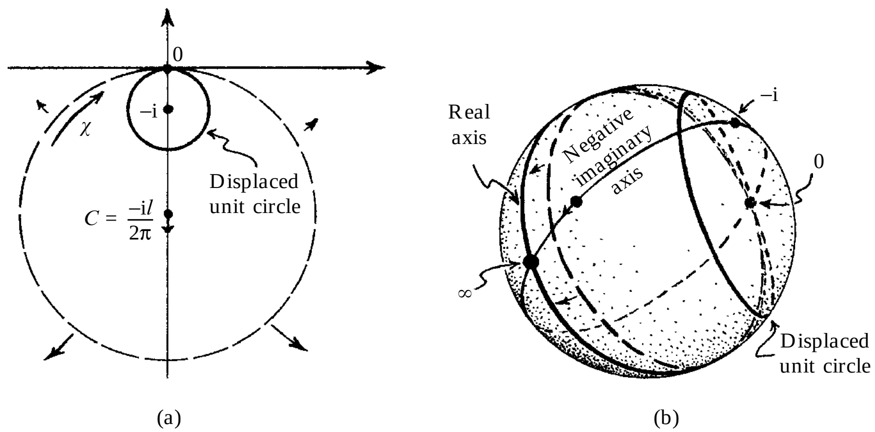

Fig. 9.10 Positive-frequency condition, as $l \to \infty$, where l is the period of $f(\chi)$. (a) Start with $l = 2\pi$, with f defined on the unit circle displaced to have its centre at $z = -\mathrm{i}$. For increasing l, the circle has radius l and centre at $C = -\mathrm{i}l/2\pi$. In each case χ measures arc length *clockwise*. Positive frequency is expressed as f being holomorphically extendible to the interior of the circle, and in the limit $l = \infty$, to the lower half-plane. (b) The same, on the Riemann sphere. For finite l, the Fourier series is obtained from a Laurent series about $z = -\mathrm{i}l/2\pi$, but on the sphere, this point is not the circle's centre, becoming the point ∞ (lying on it) in the limit $l = \infty$, where the Fourier series becomes the Fourier transform.

for period $l = 2\pi$, the unit circle centred at $z = -\mathrm{i}$, rather than at the origin. For a period $l > 2\pi$, the circle is centred at the point $C = -\mathrm{i}l/2\pi$ in the complex plane, and, in the limit as $l \to \infty$, we get the real axis itself (so $\chi = x$), the circle's 'centre' having moved off to infinity along the negative imaginary axis. In each case, we now take χ to measure arc length *clockwise* around the circle (or, in the limiting case, just positive distance along the real axis), with $\chi = 0$ at the origin. Since our circles now have a non-standard (i.e. clockwise) orientation, their 'outsides' are their *interiors* (see §9.3, Fig. 9.6), so our positive frequency condition refers to this interior. We now have the relation between χ and z expressed as[9.7]

$$z = \frac{\mathrm{i}l}{2\pi}\left(\mathrm{e}^{-\mathrm{i}\chi} - 1\right).$$

For finite l, we can express $f(\chi)$ as a Fourier series by referring to a Laurent series about the point $C = -\mathrm{i}l/2\pi$. We get the Fourier transform by taking the limit $l \to \infty$. For finite l, we obtain the condition of positive frequency as the holomorphic extendability of $f(\chi)$ into the *interior* of the relevant circle; in the limit $l \to \infty$, this becomes holomorphic extendability into the lower half-plane, in accordance with what has been stated above.

What happens to the Laurent series in the limit $l \to \infty$? We shall need to look at the Riemann sphere to understand what happens in this limit. For each finite value of l, the point $C(= \mathrm{i}l/2\pi)$ is the centre of the χ-circle, but, on the Riemann sphere, the point C need be nothing like the centre of the circle. As l increases, C moves out along the circle on the Riemann sphere which represents the imaginary axis (see Fig. 9.10b), and the point $C(= -\mathrm{i}l/2\pi)$ looks less and less like the centre of the circle. Finally, when the limit $l = \infty$ is reached, C becomes the point $z = \infty$ on the Riemann sphere. But when $C = \infty$, we find that it actually lies *on* the circle which it is supposed to be the centre of! (This circle is, of course, now the real axis.) Thus, there is something peculiar (or 'singular') about the taking of a power series about this point—which is to be expected, of course, because we do not get a sum of individual terms any more, but a continuous integral.

9.6 What kind of function is appropriate?

Let us now return to the question posed at the beginning of this chapter, concerning the type of 'function' that is appropriate to use. We can raise

[9.7] Derive this expression.

the following issue: what kind of functions can we represent as Fourier transforms? It would seem to be inappropriate to restrict attention only to analytic (i.e to $C^{\omega}-$) functions because, as we saw above, the Fourier transform $g(p)$ of a positive-frequency function $f(\chi)$—which can certainly be analytic—is a distinctly non-analytic 'gluing job' of a non-zero function to the zero function. The relation between a function and its Fourier transform is symmetrical, so it seems unreasonable to adopt such different standards for each. As a further point, it was noted above that the behaviour of $f(\chi)$ at the point $\chi = \infty$ is relevant to the issue of its positive/negative-frequency splitting, but only in very special circumstances would $f(\chi)$ actually be analytic (C^{ω}) at ∞ (since this would require a precise matching between the behaviour of $f(\chi)$ as $\chi \to +\infty$ and as $\chi \to -\infty$). In addition to all this, there is our initial *physical* motivation, referred to earlier, for studying Fourier transforms, namely that they allow us to treat signals which can transmit 'unexpected' (non-analytic) messages. Thus, we must return to the question which confronted us at the beginning of this chapter: what kind of function should we accept as being an 'honest' function?

We recall that, on the one hand, Euler and his contemporaries might indeed have probably settled for a holomorphic (or analytic) function as being the kind of thing that they had in mind for a respectable 'function'; yet, on the other hand, such functions seem unreasonably restrictive for many kinds of mathematical and physical problem, including those concerned with wave propagation, so a more general notion is needed. Is one of these points of view more 'correct' than the other? There is probably a strong prevailing opinion that supporters of the first viewpoint are 'old-fashioned', and that modern concepts lean heavily towards the second, so that holomorphic or analytic functions are just very special cases of the general notion of a 'function'. But is this necessarily the 'right' attitude to take? Let us try to put ourselves into an 18th-century frame of mind.

Enter Joseph Fourier early in the 19th century. Those who belonged to the 'analytic' ('Eulerian') school of thought would have received a nasty shock when Fourier showed that certain periodic functions, such as the square wave or saw tooth depicted in Fig. 9.11, have perfectly reasonable-looking Fourier representations! Fourier encountered a great deal of opposition from the mathematical establishment at the time. Many were reluctant to accept his conclusions. How could there be a 'formula' for the square-wave function, for example? Yet, as Fourier showed, the series

$$s(\chi) = \sin \chi + \tfrac{1}{3}\sin 3\chi + \tfrac{1}{5}\sin 5\chi + \tfrac{1}{7}\sin 7\chi + \cdots$$

actually sums to a square wave, taking this wave to oscillate between the constant values $\frac{1}{4}\pi$ and $-\frac{1}{4}\pi$ in the half-period π (see Fig. 9.12).

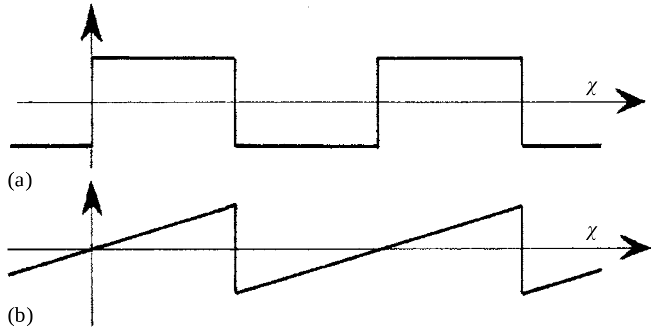

Fig. 9.11 Discontinuous periodic functions (with perfectly reasonable-looking Fourier representations): (a) Square wave (b) Saw tooth.

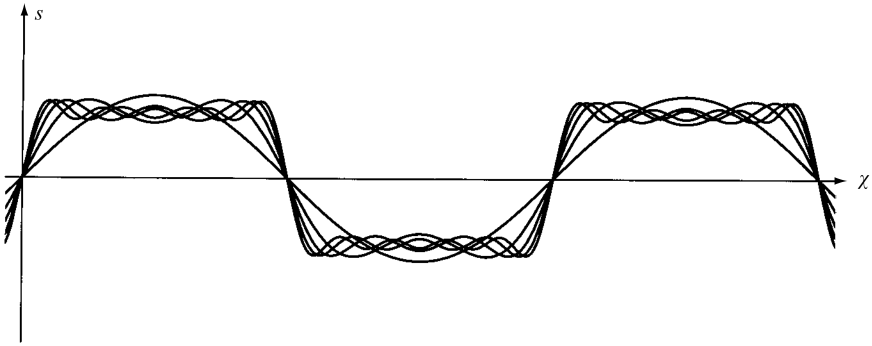

Fig. 9.12 Partial sums of the Fourier series $s(\chi) = \sin\chi + \frac{1}{3}\sin 3\chi + \frac{1}{5}\sin 5\chi + \frac{1}{7}\sin 7\chi + \frac{1}{9}\sin 9\chi + \cdots$, converging to a square wave (like that of Fig. 9.11a).

Let us consider the Laurent-series description for this, as given above. We have the rather elegant-looking expression[9.8]

$$2\mathrm{i}s(\chi) = \cdots - \tfrac{1}{5}z^{-5} - \tfrac{1}{3}z^{-3} - z^{-1} + z + \tfrac{1}{3}z^{3} + \tfrac{1}{5}z^{5} + \cdots,$$

where $z = \mathrm{e}^{i\chi}$. In fact this is an example where the annulus of convergence shrinks down to the unit circle—with no actual open region left. However, we can still make sense of things in terms of holomorphic functions if we split the Laurent series into two halves, one with the positive powers, giving an ordinary power series in z, and one with the negative powers, giving a power series in z^{-1}. In fact, these are well-known series, and can be summed explicitly:[9.9]

[9.8] Show this.

[9.9] Do this, by taking advantage of a power series expansion for log z taken about $z = 1$, given towards the end of §7.4.

$$S^- = z + \tfrac{1}{3}z^3 + \tfrac{1}{5}z^5 + \cdots = \tfrac{1}{2}\log\left(\frac{1+z}{1-z}\right)$$

and

$$S^+ = \cdots - \tfrac{1}{5}z^{-5} - \tfrac{1}{3}z^{-3} - z^{-1} = -\tfrac{1}{2}\log\left(\frac{1+z^{-1}}{1-z^{-1}}\right),$$

giving $2\mathrm{i}s(\chi) = S^- + S^+$. A little rearrangement of these expressions leads to the conclusion that S^- and $-S^+$ differ only by $\pm\frac{1}{2}\mathrm{i}\pi$, telling us that $s(\chi) = \pm\frac{1}{4}\pi$.[9.10] But we need to look a little more closely to see why we actually get a square wave oscillating between these alternative values.

It is a little easier to appreciate what is going on if we apply the transformation $t = (z-1)/(\mathrm{i}z+\mathrm{i})$, given in §8.3, which takes the interior of the unit circle in the z-plane to the upper half-t-plane (as illustrated in Fig. 8.10). In terms of t, the quantity S^- now refers to this upper half-plane and S^+ to the lower half-plane, and we find (with possible $2\pi\mathrm{i}$ ambiguities in the logarithms)

$$S^- = -\tfrac{1}{2}\log t + \tfrac{1}{2}\log \mathrm{i}, \quad S^+ = \tfrac{1}{2}\log t + \tfrac{1}{2}\log \mathrm{i}.$$

Following the logarithms continuously from the respective starting points $t = \mathrm{i}$ (where $S^- = 0$) and $t = -\mathrm{i}$ (where $S^+ = 0$), we find that along the positive real t-axis we have $S^- + S^+ = +\frac{1}{2}\mathrm{i}\pi$, whereas along the negative real t-axis we have $S^- + S^+ = -\frac{1}{2}\mathrm{i}\pi$.[9.11] From this we deduce that along the top half of the unit circle in the z-plane we have $s(\chi) = +\frac{1}{4}\pi$, whereas along the bottom half we have $s(\chi) = -\frac{1}{4}\pi$. This shows that the Fourier series indeed sums to the square wave, just as Fourier had asserted.

What is the moral to be drawn from this example? We have seen that a particular (periodic) function that is not even continuous, let alone differentiable (in this case being a C^{-1}-function), can be represented as a perfectly sensible-looking Fourier series. Equivalently, when we think of the function as being defined on the unit circle, it can be represented as a reasonable-appearing Laurent series, although it is one for which the annulus of convergence has, in effect, shrunk down to the unit circle itself. The positive and the negative half of this Laurent series each sums to a perfectly good holomorphic function on half of the Riemann sphere. One is defined on one side of the unit circle, and the other is defined on the other side. We can think of the 'sum' of these two functions as giving the required square wave on the unit circle itself. It is because of the existence of branch singularities at the two points $z = \pm 1$ on

[9.10] Show this (assuming that $|s(\chi)| < 3\pi/2$).

[9.11] Show this.

the unit circle that the sum can 'jump' from one side to the other, giving the square wave that arises in this sum. These branch singularities also prevent the power series on the two sides from converging beyond the unit circle.

9.7 Hyperfunctions

This example is only a very special case, but it illustrates what we must do in general. Let us ask what is the most general type of function that can be defined on the unit circle (on the Riemann sphere) and represented as a 'sum' of some holomorphic function F^+ on the open region lying to one side of the circle and of another holomorphic function F^- on the open region lying to the other side, just as in the example that we have been considering. We shall find that the answer to this question leads us directly to an exotic but important notion referred to as a 'hyperfunction'.

In fact, it turns out to be more illuminating to think of f as being the 'difference' between F^- and $-F^+$. One reason for this is that, in the most general cases, there may be no analytic extension of either F^- or F^+ to the actual unit circle, so it is not clear what such a 'sum' could mean on the circle itself. However, we can think of the *difference* between F^- and $-F^+$ as representing the 'jump' between these two functions as their regions of definition come together at the unit circle.

This idea of a 'jump' between a holomorphic function on one side of a curve in the complex plane and another holomorphic function on the other—where neither holomorphic function need extend holomorphically over the curve itself—actually provides us with a new concept of a 'function' defined on the curve. This is, in effect, the definition of a *hyperfunction* on an (analytic) curve. It is a wonderful notion put forward by the Japanese mathematician Mikio Sato in 1958,[9] although, as we shall shortly be seeing, Sato's actual definition is considerably more elegant than just this.[10]

We do not need to think of a closed curve, like the entire unit circle, for the definition of a hyperfunction, but we can consider some part of a curve. Indeed, it is more usual to consider hyperfunctions as defined on some segment γ of the real line. We shall take γ to be the segment of the real line between a and b, where a and b are real numbers with $a < b$. A hyperfunction defined on γ is then the *jump* across γ, starting from a holomorphic function f on an open set $\mathcal{R}^-$ (having γ as its upper boundary) to a holomorphic function g on an open set $\mathcal{R}^+$ (having γ as its lower boundary) see Fig. 9.13.

Simply to refer to a 'jump' in this way does not give us much idea of what to do with such a thing (and it is not yet very mathematically precise). Sato's elegant resolution of these issues is to proceed in a rather

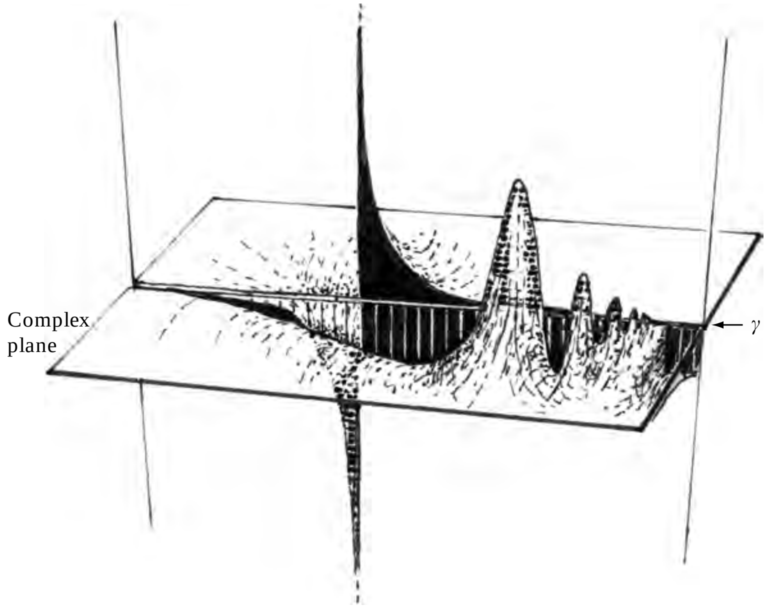

Fig. 9.13 A hyperfunction on a segment γ of the real axis expresses the 'jump' from a holomorphic function on one side of γ to one on the other.

formally algebraic way, which is actually extrordinarily simple. We merely represent this jump as the *pair* (f, g) of these holomorphic functions, but where we say that such a pair (f, g) is *equivalent* to another such pair (f_0, g_0) if the latter is obtained from the former by adding to both f and g the same holomorphic function h, where h is defined on the combined (open) region $\mathcal{R}$, which consists of $\mathcal{R}^-$ and $\mathcal{R}^+$ joined together along the curve segment γ; see Fig. 9.14. We can say

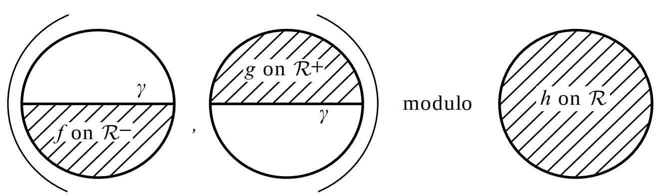

Fig. 9.14 A hyperfunction, on a segment γ of the real axis, is provided by a pair of holomorphic functions (f, g), with f defined on some open region $\mathcal{R}^-$, extending downwards from γ and g on an open region $\mathcal{R}^+$, extending upwards from γ. The actual hyperfunction h, on γ, is (f, g) *modulo* quantities $(f + h, g + h)$, where h is holomorphic on the union $\mathcal{R}$ of $\mathcal{R}^-$, γ, and $\mathcal{R}^+$.

$$(f, g) \quad \text{is equivalent to} \quad (f+h, g+h),$$

where the holomorphic functions f and g are defined on $\mathcal{R}^-$ and $\mathcal{R}^+$, respectively, and where h is an arbitrary holomorphic function on the combined region $\mathcal{R}$. Either of the above displayed expressions can be used to represent the same hyperfunction. The hyperfunction itself would be mathematically referred to as the *equivalence class* of such pairs, 'reduced modulo'[11] the holomorphic functions h defined on $\mathcal{R}$. The reader may recall the notion of 'equivalence class' referred to in the Preface, in connection with the definition of a fraction. This is the same general idea—and no less confusing. The essential point here is that adding h does not affect the 'jump' between f and g, but h can change f and g in ways that are irrelevant to this jump. (For example, h can change how these functions happen to continue away from γ into the open regions $\mathcal{R}^-$ and $\mathcal{R}^+$.) Thus, the jump itself is neatly represented *as* this equivalence class.

The reader may be genuinely disturbed that this slick definition seems to depend crucially on our arbitrary choices of open regions $\mathcal{R}^-$ and $\mathcal{R}^+$, restricted merely by their being joined along their common boundary line γ. Remarkably, however, the definition of a hyperfunction does *not* depend on this choice. According to an astonishing theorem, known as the *excision theorem*, this notion of hyperfunction is actually quite independent of the particular choices of $\mathcal{R}^-$ and $\mathcal{R}^+$; see top three examples of Fig. 9.15.

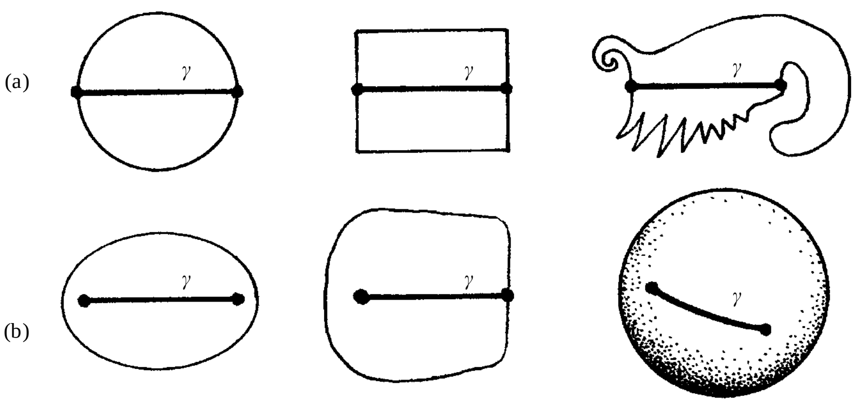

Fig. 9.15 The excision theorem tells us that the notion of a hyperfunction is independent of the choice of open region $\mathcal{R}$, so long as $\mathcal{R}$ contains the given curve γ. (a) The region $\mathcal{R} - \bar{\gamma}$ may consist of two separate pieces (so we get two distinct holomorphic functions f and g, as in Fig. 9.14) or (b) the region $\mathcal{R} - \bar{\gamma}$ may be a single connected piece, in which case f and g are simply two parts of the same holomorphic function.

In fact, the excision theorem gives us more than even this. We do not require that our open region $\mathcal{R}$ be divided into two (namely into $\mathcal{R}^-$ and $\mathcal{R}^+$) by the removal of γ. All we need is that the open region $\mathcal{R}$, in the complex plane, must contain the open[12] segment γ. It may be that $\mathcal{R} - \gamma$ (i.e. what is left of $\mathcal{R}$ when γ is removed from it[13]) consists of two separate pieces, just as we have been considering up to this point, but more generally the removal of γ from $\mathcal{R}$ may leave us with a single connected region, as illustrated in the bottom three examples of Fig. 9.15. In these cases, we must also remove any internal end-point a or b, of γ, so that we are left with an open set, which I refer to as $\mathcal{R} - \bar{\gamma}$. In this more general case, our hyperfunctions are defined as 'holomorphic functions on $\mathcal{R}$, reduced modulo holomorphic functions on $\mathcal{R} - \bar{\gamma}$'. It is quite remarkable that this very liberal choice of $\mathcal{R}$ makes no difference to the class of 'hyperfunctions' that is thereby defined.[9.12] The case when a and b both lie within $\mathcal{R}$ is useful for integrals of hyperfunctions, since then a closed contour in $\mathcal{R} - \bar{\gamma}$ can be used.

All this applies also to our previous case of a circle on the Riemann sphere. Here, there is some advantage in taking $\mathcal{R}$ to be the entire Riemann sphere, because then the functions that we have to 'mod out by' are the holomorphic functions that are global on the entire Riemann sphere, and there is a theorem which tells us that these functions are just constants. (These are actually the 'constants' α_0 that we chose not to worry about in §9.2.) Thus, modulo constants, a hyperfunction defined on a circle on the Riemann sphere is specified simply by one holomorphic function on the entire region on one side of the circle and another function on the other side. This gives the splitting of an arbitrary hyperfunction on the circle uniquely (modulo constants) into its positive- and negative-frequency parts.

Let us end by considering some basic properties of hyperfunctions. I shall use the notation $(\!| f, g |\!)$ to denote the hyperfunction specified by the pair f and g defined holomorphically on $\mathcal{R}^-$ and $\mathcal{R}^+$, respectively (where I am reverting to the case where γ divides $\mathcal{R}$ into $\mathcal{R}^-$ and $\mathcal{R}^+$. Thus, if we have two different representations $(\!| f, g |\!)$ and $(\!| f_0, g_0 |\!)$ of the same hyperfunction, that is, $(\!| f, g |\!) = (\!| f_0, g_0 |\!)$, then $f - f_0$ and $g - g_0$ are both the *same* holomorphic function h defined on $\mathcal{R}$, but restricted to $\mathcal{R}^-$ and $\mathcal{R}^+$ respectively. It is then straightforward to express the *sum* of two hyperfunctions, the *derivative* of a hyperfunction, and the *product* of a hyperfunction with an analytic function q defined on γ:

[9.12] Why does 'holomorphic functions on $\mathcal{R}$, reduced modulo holomorphic functions on $\mathcal{R} - \bar{\gamma}$' become the definition of a hyperfunction that we had previously, when $\mathcal{R} - \bar{\gamma}$ splits into $\mathcal{R}^-$ and $\mathcal{R}^+$?

$$\llparenthesis f, g \rrparenthesis + \llparenthesis f_1, g_1 \rrparenthesis = \llparenthesis f + f_1, g + g_1 \rrparenthesis,$$
$$\frac{\mathrm{d}\llparenthesis f, g \rrparenthesis}{\mathrm{d}z} = \left(\!\!\left| \frac{\mathrm{d}f}{\mathrm{d}z}, \frac{\mathrm{d}g}{\mathrm{d}z} \right|\!\!\right),$$
$$q\llparenthesis f, g \rrparenthesis = = \llparenthesis qf, qg \rrparenthesis.$$

where, in the last expression, the analytic function q is extended holomorphically into a neighbourhood[14] of γ.[9.13] We can represent q itself as a hyperfunction by $q = \llparenthesis q, 0 \rrparenthesis = \llparenthesis 0, -q \rrparenthesis$, but there is no general product defined between two hyperfunctions. The lack of a product is not the fault of the hyperfunction approach to generalized functions. It is there with all approaches.[15] The fact that the Dirac delta function (referred to in §6.6; also see below) cannot be squared, for example, causes many quantum field theorists no end of trouble.

Some simple examples of hyperfunctional representations, in the case when $\gamma = \mathbb{R}$, and $\mathcal{R}^-$ and $\mathcal{R}^+$ are the upper and lower open complex half-planes, are the Heaviside step funtion $\theta(x)$ and the Dirac (-Heaviside) delta function $\delta(x)(= \mathrm{d}\theta(x)/\mathrm{d}x)$ (see §§6.1,6):

$$\theta(x) = \left(\!\!\left| \frac{1}{2\pi\mathrm{i}}\log z, \frac{1}{2\pi\mathrm{i}}\log z - 1 \right|\!\!\right),$$
$$\delta(x) = \left(\!\!\left| \frac{1}{2\pi\mathrm{i}z}, \frac{1}{2\pi\mathrm{i}z} \right|\!\!\right),$$

where we take the branch of the logarithm for which $\log 1 = 0$. The integral of the hyperfunction $\llparenthesis f, g \rrparenthesis$ over the entire real line can be expressed as the integral of f along a contour just below the real line minus the integral of g along a contour just above the real line (assuming these converge), both from left to right.[9.14] Note that the hyperfunction can be non-trivial even when f and g are analytic continuations of the same function.

How general are hyperfunctions? They certainly include all analytic functions. They also include discontinuous functions like $\theta(x)$ and the square wave (as our discussions above show), or other C^{-1}-functions obtained by adding such things together. In fact all C^{-1}-functions are examples of hyperfunctions. Moreover, since we can differentiate a hyperfunction to obtain another hyperfunction, and any C^{-2}-function can be obtained as the derivative of some C^{-1}-function, it follows that all C^{-2}-functions are also hyperfunctions. We have seen that this includes the

---

[9.13] There is a small subtlety here. Sort it out. *Hint*: Think carefully about the domains of definition.

[9.14] Check the standard property of the delta function that $\int q(x)\delta(x)\mathrm{d}x = q(0)$, in the case when $q(x)$ is analytic.

Dirac delta function. We can differentiate again, and then again. Indeed, any C^{-n}-function is a hyperfunction for any integer n whatever. What about the $C^{-\infty}$-functions, referred to as *distributions* (see §6.6). Yes, these also are all hyperfunctions.

The normal definition of a distribution[16] is as an element of what is called the *dual* space of the C^{∞}-smooth functions. The concept of a 'dual space' will be discussed in §12.3 (and §13.6). In fact, the dual (in an appropriate sense) of the space of C^{n}-functions is the space of C^{-2-n}-functions for any integer n, and this applies also to $n = \infty$, if we write $-2-\infty = -\infty$ and $-2+\infty = \infty$. Accordingly, the $C^{-\infty}$-functions are indeed dual to the C^{∞}-functions. What about the dual ($C^{-\omega}$) of the C^{ω}-functions? Indeed; with the appropriate definition of 'dual', these $C^{-\omega}$-functions are precisely the hyperfunctions!

We have come full circle. In trying to generalize the notion of 'function' as far as we can away from the apparently very restrictive notion of an 'analytic' or 'holomorphic' function—the type of function that would have made Euler happy—we have come round to the extremely general and flexible notion of a *hyperfunction*. But hyperfunctions are themselves defined, in a basically very simple way, in terms of the these very same 'Eulerian' holomorphic functions that we thought we had reluctantly abandoned. In my view, this is one of the supreme magical achievements of complex numbers.[16] If only Euler had been alive to appreciate this wondrous fact!

Notes

Section 9.1

9.1. I am using the greek letter χ ('chi') here, rather than an ordinary x, which might have seemed more natural, only because we need to distinguish this variable from the real part x of the complex number z, which will play an important part in what follows.

9.2. There is no requirement that $f(\chi)$ be real for real values of χ, that is, for the a_n,b_n, and c to be real numbers. It is perfectly legitimate to have complex functions of real variables. The condition that $f(\chi)$ be real is that α_{-n} be the complex conjugate of α_n. Complex conjugates will be discussed in §10.1.

Section 9.2

9.3. The odd-looking notational anomaly of using 'F^-' for the part of the series with positive powers and 'F^+' for the part with negative powers springs ultimately from a perhaps unfortunate sign convention that has become almost universal in the quantum-mechanical literature (see §§21.2,3 and §24.3). I apologize for this, but there is nothing that I can reasonably do about it!

9.4. It is a general principle that, for any C^{ω}-function f, defined on a real domain $\mathcal{R}$, it is possible to 'complexify' $\mathcal{R}$ to a slightly extended complex domain $\mathbb{C}\mathcal{R}$, called a 'complex thickening' of $\mathcal{R}$, containing $\mathcal{R}$ in its interior, such that f extends uniquely to a holomorphic function defined on $\mathbb{C}\mathcal{R}$.

9.5. See e.g. Bailey *et al.* (1982).

Section 9.4

9.6. On the other hand, it is usual to impose some requirement that $f(\chi)$ behaves 'reasonably' as χ tends to positive or negative infinity. This will not be of particular concern for us here and, in any case, with the approach that I am adopting, the normal requirements would be unnecessarily restrictive.

9.7. In quantum mechanics, there is also a constant quantity $\hbar$ introduced to fix the scaling of p appropriately, in relation to x (see §§21.2,11), but for the moment I am keeping things simple by taking $\hbar = 1$. In fact, $\hbar$ is Dirac's form of Planck's constant (i.e. $h/2\pi$, where h is Planck's original 'quantum of action'). The choice $\hbar = 1$ can always be made, by defining our basic units in a suitable way. See §27.10.

Section 9.5

9.8. See Bailey *et al.* (1982).

Section 9.7

9.9. See Sato (1958, 1959, 1960).

9.10. See also Bremermann (1965), although the term 'hyperfunction' is not used explicitly in this work.

9.11. Another aspect of the notion 'modulo' will be discussed in §16.1 (and compare Note 3.17).

9.12. Here 'open segment' simply refers to the fact that the actual end-points a and b are not included in γ, so that 'containing' γ does not imply the containing of a and b within $\mathcal{R}$.

9.13. This 'difference' between sets $\mathcal{R}, \gamma$ is also commonly written $\mathcal{R} \setminus \gamma$.

9.14. The technical definition of 'neighbourhood of' is 'open set containing'.

9.15. For the more standard ('distribution') approach to the idea of 'generalized function', see Schwartz (1966); Friedlander (1982); Gel'fand and Shilov (1964); Trèves (1967); for an alternative proposal, useful in 'nonlinear' contexts, and which shifts the 'product existence problem to a non-uniqueness problem—see Colombeau (1983, 1985) and Grosser *et al.* (2001).

9.16. There are also important interconnections between hyperfunctions and the holomorphic sheaf cohomology that will be discussed in §33.9. Such ideas play important roles in the theory of hyperfunctions on higher-dimensional surfaces, see Sato (1959, 1960) and Harvey (1966).