Chapter 2 Groundwork

Most of the material in this chapter is stated without proof. This is done because the proofs entail discussions that are lengthy (in fact, they form the bulk of conventional studies in Fourier theory) and remote from the subject matter of the present work.

Omitting the proofs enables us to take the transform formulas and their known conditions as our point of departure. Since suitable notation is an important part of the work, it too is set out in this chapter.

The Fourier transform and Fourier's integral theorem

The Fourier transform of $f(x)$ is defined as

$$\int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx.$$

This integral, which is a function of s, may be written $F(s)$. Transforming $F(s)$ by the same formula, we have

$$\int_{-\infty}^{\infty} F(s)e^{-i2\pi ws}\,ds.$$

When $f(x)$ is an even function of x, that is, when $f(x) = f(-x)$, the repeated transformation yields $f(w)$, the same function we began with. This is the cyclical property of the Fourier transformation, and since the cycle is of two steps, the reciprocal property is implied: if $F(s)$ is the Fourier transform of $f(x)$, then $f(x)$ is the Fourier transform of $F(s)$.

The cyclical and reciprocal properties are imperfect, however, because when $f(x)$ is odd—that is, when $f(x) = -f(-x)$—the repeated trans ormation yields $f(-w)$. In general, whether $f(x)$ is even or odd or neither, repeated transformation yields $f(-w)$.

The customary formulas exhibiting the reversibility of the Fourier transformation are

$$F(s) = \int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx$$
$$f(x) = \int_{-\infty}^{\infty} F(s)e^{i2\pi xs}\,ds.$$

In this form, two successive transformations are made to yield the original function. The second transformation, however, is not exactly the same as the first, and where it is necessary to distinguish between these two sorts of Fourier transform, we shall say that $F(s)$ is the minus-i transform of $f(x)$ and that $f(x)$ is the plus-i transform of $F(s)$.

Writing the two successive transformations as a repeated integral, we obtain the usual statement of Fourier's integral theorem:

$$f(x) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx\right]e^{i2\pi xs}\,ds.$$

The conditions under which this is true are given in the next section, but it must be stated at once that where $f(x)$ is discontinuous the left-hand side should be replaced by $\frac{1}{2}[f(x+) + f(x-)]$, that is, by the mean of the unequal limits of $f(x)$ as x is approached from above and below.

The factor 2π appearing in the transform formulas may be lumped with s to yield the following version (system 2):

$$F(s) = \int_{-\infty}^{\infty} f(x)e^{-ixs}\,dx$$
$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(s)e^{ixs}\,ds.$$

And for the sake of symmetry, authors occasionally write (system 3):

$$F(s) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} f(x)e^{-ixs}\,dx$$
$$f(x) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} F(s)e^{ixs}\,ds.$$

All three versions are in common use, but here we shall keep the 2π in the exponent (system 1). If $f(x)$ and $F(s)$ are a transform pair in system 1, then $f(x)$ and $F(s/2\pi)$ are a transform pair in system 2, and $f[x/(2\pi)^{\frac{1}{2}}]$ and $F[s/(2\pi)^{\frac{1}{2}}]$ are a transform pair in system 3. An example of a transform pair in each of the three systems follows.

System 1		*System 2*		*System 3*	
$f(x)$	$F(s)$	$f(x)$	$F(s)$	$f(x)$	$F(s)$
$e^{-\pi x^2}$	$e^{-\pi s^2}$	$e^{-\pi x^2}$	$e^{-s^2/4\pi}$	$e^{-\frac{1}{2}x^2}$	$e^{-\frac{1}{2}s^2}$

An excellent notation which may be used as an alternative to $F(s)$ is $\bar{f}(s)$. Various advantages and disadvantages are found in both notations.

The bar notation leads to compact expression, including some convenient manuscript or blackboard forms which are not very suitable for typesetting. Consider, for example, these versions of the convolution theorem,

$$\overline{FG} = \bar{F} * \bar{G}$$
$$\overline{F * G} = \bar{F}\bar{G}$$
$$\overline{\bar{F}\bar{G}} = F * G$$
$$\overline{\bar{F} * \bar{G}} = FG,$$

which display shades of distinction not expressible at all in the capital notation. See Chapter 6 for illustrations of the freedom of expression permitted by the bar notation. A certain awkwardness sets in, however, when complex conjugates or primes representing derivatives have to be handled; this awkwardness does not afflict the capital notation. Therefore we have departed from the usual custom of adopting a single notation. In the early mathematical sections, where f and g are nearly the only symbols for functions, the capital notation is mainly used. In the physical sections, preemption of capitals such as E and H for the representation of physical quantities leads more naturally to bars.

Neither of the above two notations lends itself to symbolic statements equivalent to "the Fourier transform of exp $(-\pi x^2)$ is exp $(-\pi s^2)$"; however, we can write

$$\mathfrak{F}e^{-\pi x^2} = e^{-\pi s^2}$$

or

$$e^{-\pi x^2} \supset e^{-\pi s^2}.$$

In the first of these, $\mathfrak{F}$ can be regarded as a functional operator which converts a function into its transform. It may be applied wherever the bar and capital notations are employed, but will be found most appropriate in connection with specific functions such as those above. It lends itself to the use of affixes, (for example, $\mathfrak{F}_s$, $\mathfrak{F}_c$, ${}^2\mathfrak{F}$) and mixes with symbols for other transforms. It can also distinguish between the minus-i and plus-i transforms through the use of $\mathfrak{F}^{-1}$ for the inverse of $\mathfrak{F}$; or like the bar notation, but not the capital notation, it can remain discreetly silent. The properties of this notation make it indispensable, and it is adopted in suitable places in the sequel.

The sign $\supset$ or equivalents, several of which are in use, is not as versatile as $\mathfrak{F}$ but is useful for algebraic work. It is also often used to denote the Laplace transform.

Conditions for the existence of Fourier transforms

A circuit expert finds it obvious that every waveform has a spectrum, and the antenna designer is confident that every antenna has a radiation

pattern. It sometimes comes as a surprise to those whose acquaintance with Fourier transforms is through physical experience rather than mathematics that there are some functions without Fourier transforms. Nevertheless, we may be confident that no one can generate a waveform without a spectrum or construct an antenna without a radiation pattern.

The question of the existence of transforms may safely be ignored when the function to be transformed is an accurately specified description of a physical quantity. Physical possibility is a valid sufficient condition for the existence of a transform. Sometimes, however, it is convenient to substitute a simple mathematical expression for a physical quantity. It is very common, for example, to consider the waveforms

$\sin t$	(harmonic wave, pure alternating current)
$H(t)$	(step)
$\delta(t)$	(impulse).

It turns out that none of these three has, strictly speaking, a Fourier transform. Of course, none of them is physically possible, for a waveform $\sin t$ would have to have been switched on an infinite time ago, a step $H(t)$ would have to be maintained steady for an infinite time, and an impulse $\delta(t)$ would have to be infinitely large for an infinitely short time. However, in a given situation we can often achieve an approximation so close that any further improvement would be immaterial, and we use the simple mathematical expressions because they are less cumbersome than various slightly different but realizable functions. Nevertheless, the above functions do not have Fourier transforms; that is, the Fourier integral does not converge for all s. It is therefore of practical importance to consider the conditions for the existence of transforms.

Transforming and retransforming a single-valued function $f(x)$, we have the repeated integral

$$\int_{-\infty}^{\infty} e^{i2\pi sx} \left[\int_{-\infty}^{\infty} f(x)e^{-i2\pi sx}\,dx \right] ds.$$

This expression is equal to $f(x)$ (or to $\frac{1}{2}[f(x+) + f(x-)]$ where $f(x)$ is discontinuous), provided that

1. The integral of $|f(x)|$ from $-\infty$ to ∞ exists
2. Any discontinuities in $f(x)$ are finite

A further but less important condition is mentioned below. In physical circumstances these conditions are violated[1] when there is infinite energy, and a kind of duality between the two conditions is often noted. For instance, absolutely steady direct current which has always been

[1] Exceptions are provided by finite-energy waveforms such as $(1 + |x|)^{-\frac{1}{2}}$ and $x^{-1} \sin x$, which nevertheless do not have absolutely convergent infinite integrals.

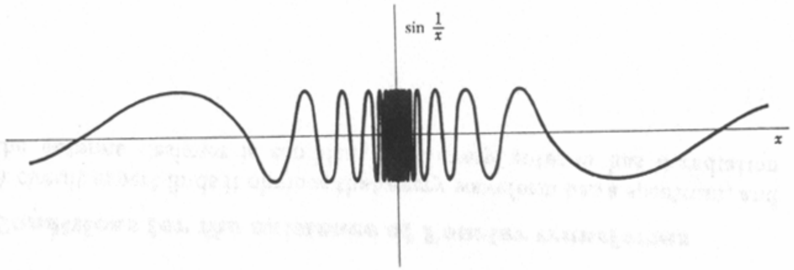


Fig. 2.1 A function with an infinite number of maxima.

flowing and always will flow represents infinite energy and violates the first condition. The distribution of energy with frequency would have to show infinite energy concentrated entirely at zero frequency and would violate the second condition. The same applies to harmonic waves.

It is sometimes stated that an infinite number of maxima and minima in a finite interval disqualifies a function from possessing a Fourier transform, the stock example being sin x^{-1} (see Fig. 2.1), which oscillates with ever-increasing frequency as x approaches zero. This kind of behavior is not important in real life, even as an approximation. We therefore record for general interest that some functions with infinite numbers of maxima and minima in a finite interval do have transforms. This is allowed for when the further condition is given as *bounded variation*.[2] Again, however, there are transformable functions with an infinite number of maxima and with unbounded variation in a finite interval, a circumstance which may be covered by requiring $f(x)$ to satisfy a Lipschitz condition.[3] Then there is a more relaxed condition used by Dini. This is a fascinating topic in Fourier theory, but it is not immediately relevant to our branch of it, which is physical applications. Furthermore, we by no means propose to abandon useful functions which do not possess Fourier transforms in the ordinary sense. On the contrary, we include them equally, by means of a generalization to Fourier transforms in the

[2] A function $f(x)$ has bounded variation over the interval $x = a$ to $x = b$ if there is a number M such that

$$|f(x_1) - f(a)| + |f(x_2) - f(x_1)| + \cdots + |f(b) - f(x_{n-1})| \leqslant M$$

for every method of subdivision $a < x_1 < x_2 < \cdots < x_{n-1} < b$. Any function having an absolutely integrable derivative will have bounded variation.

[3] A function $f(x)$ satisfies a Lipschitz condition of order α at $x = 0$ if

$$|f(h) - f(0)| \leqslant B|h|^{\beta}$$

for all $|h| < \epsilon$, where B and β are independent of h, β is positive, and α is the upper bound of all β for which finite B exists.

limit. Conditions for the existence of Fourier transforms now merely distinguish, where distinction is desired, between those transforms which are ordinary and those which are transforms in the limit.

Transforms in the limit

Although a periodic function does not have a Fourier transform, as may be verified by reference to the conditions for existence, it is nevertheless considered in physics to have a spectrum, a "line spectrum." The line spectrum may be identified with the coefficients of the Fourier series for the periodic function, or we may broaden the mathematical concept of the Fourier transform to bring it into harmony with the physical standpoint. This is what we shall do here, taking the periodic function as one example among others which we would like Fourier transform theory to embrace.

Let $P(x)$ be a periodic function of x. Then

$$\int_{-\infty}^{\infty} |P(x)|\,dx$$

does not exist, but if we modify $P(x)$ slightly by multiplication with a factor such as $\exp(-\alpha x^2)$, where α is a small positive number, then the modified version may have a transform, for

$$\int_{-\infty}^{\infty} |e^{-\alpha x^2} P(x)|\,dx$$

may exist. Of course, any infinite discontinuities in $P(x)$ will still disqualify it, but let us select $P(x)$ so that $\exp(-\alpha x^2)P(x)$ possesses a Fourier transform. Then as α approaches zero, the modifying factor for each value of x approaches unity, and the modified functions of the sequence generated as α approaches zero thus approach $P(x)$ in the limit. Since each modified function possesses a transform, a corresponding sequence of transforms is generated; now, as α approaches zero, does this sequence of transforms also approach a limit? We already know that it does not, at least not for all s; we content ourselves with saying that the sequence of regular transforms defines or constitutes an entity that shall be called a generalized function. The periodic function and the generalized function form a Fourier transform pair in the limit.

The idea of dealing with things that are not functions but are describable in terms of sequences of functions is well established in physics in connection with the impulse symbol $\delta(x)$. In this case a progression of ever-stronger and ever-narrower unit-area pulses is an appropriate sequence, and a little later in the chapter we go into this idea more fully. We use the term "generalized function" to cover impulses and their like.

Periodic functions fail to have Fourier transforms because their infinite integral is not absolutely convergent; failure may also be due to the infinite discontinuities associated with impulses. In this case we replace any impulse by a sequence of functions that do have transforms; then the sequence of corresponding transforms may approach a limit, and again we have a Fourier transform pair in the limit. As before, only one member of the pair is a generalized function involving impulses.

It may also happen that the sequence of transforms does not approach a limit. This would be so if we began with something that was both impulsive and periodic; then the members of the transform pair in the limit would both be generalized functions involving impulses.

At this point we might proceed to lay the groundwork leading to the definition of a generalized function. Instead we defer the rather severe

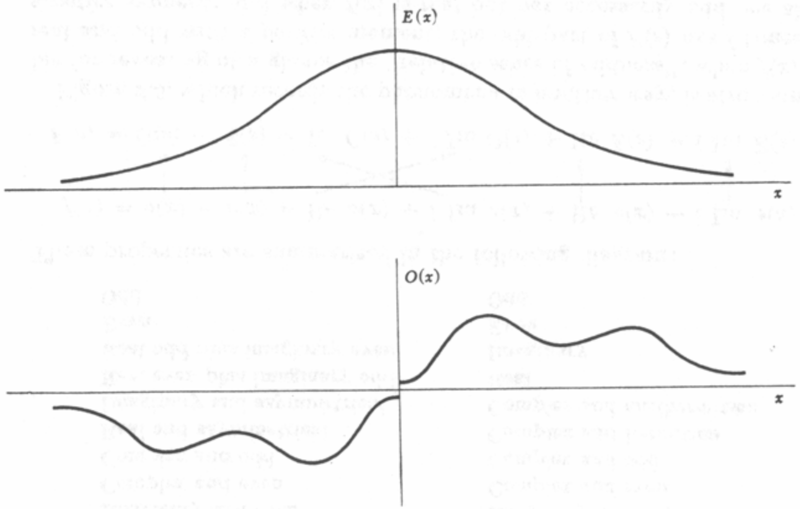


Fig. 2.2 An even function $E(x)$ and an odd function $O(x)$.

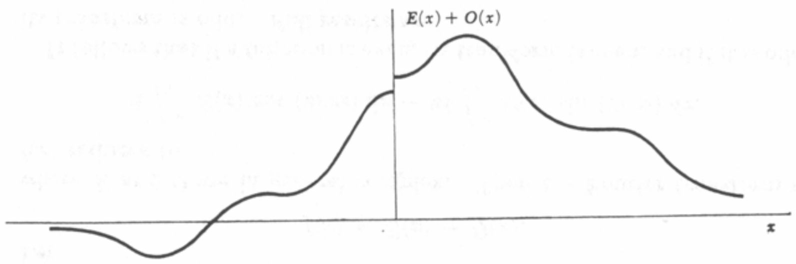


Fig. 2.3 The sum of $E(x)$ and $O(x)$.

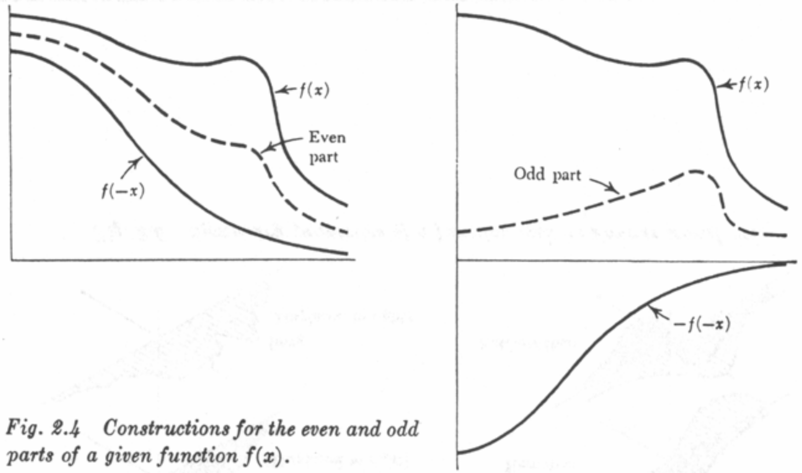


Fig. 2.4 Constructions for the even and odd parts of a given function $f(x)$.

general discussion to a much later stage, since it can be read with more profit after facility has been acquired in handling the impulse symbol $\delta(x)$.

Oddness and evenness

Symmetry properties play an important role in Fourier theory. Arguments from symmetry to show directly that certain integrals vanish, without the need of evaluating them, are familiar and perhaps often seem trivial in print. More alertness is needed, however, to ensure full exploitation in one's own reasoning of symmetry restrictions and the corresponding restrictive properties generated under Fourier transformation. Some simple terminology is recalled here.

A function $E(x)$ such that $E(-x) = E(x)$ is a symmetrical, or even, function. A function $O(x)$ such that $O(-x) = -O(x)$ is an antisymmetrical, or odd, function (see Fig. 2.2). The sum of even and odd functions is in general neither even nor odd, as illustrated in Fig. 2.3, which shows the sum of the previously chosen examples.

Any function $f(x)$ can be split unambiguously into odd and even parts. For if $f(x) = E_1(x) + O_1(x) = E_2(x) + O_2(x)$, then $E_1 - E_2 = O_2 - O_1$; but $E_1 - E_2$ is even and $O_2 - O_1$ odd, hence $E_1 - E_2$ must be zero.

The even part of a given function is the mean of the function and its reflection in the vertical axis, and the odd part is the mean of the function and its negative reflection (see Fig. 2.4). Thus

$$E(x) = \tfrac{1}{2}[f(x) + f(-x)]$$

and

$$O(x) = \tfrac{1}{2}[f(x) - f(-x)].$$

The dissociation into odd and even parts changes with changing origin of x, some functions such as $\cos x$ being convertible from fully even to fully odd by a shift of origin.

Significance of oddness and evenness

Let

$$f(x) = E(x) + O(x),$$

where E and O are in general complex. Then the Fourier transform of $f(x)$ reduces to

$$2\int_0^\infty E(x)\cos(2\pi xs)\,dx - 2i\int_0^\infty O(x)\sin(2\pi xs)\,dx.$$

It follows that if a function is even, its transform is even, and if it is odd, its transform is odd. Full results are

Real and even	Real and even
Real and odd	Imaginary and odd
Imaginary and even	Imaginary and even
Complex and even	Complex and even
Complex and odd	Complex and odd
Real and asymmetrical	Complex and hermitian
Imaginary and asymmetrical	Complex and antihermitian
Real even plus imaginary odd	Real
Real odd plus imaginary even	Imaginary
Even	Even
Odd	Odd

These properties are summarized in the following diagram:

$$f(x) = o(x) + e(x) = \text{Re}\ o(x) + i\ \text{Im}\ o(x) + \text{Re}\ e(x) + i\ \text{Im}\ e(x)$$

$$F(s) = O(s) + E(s) = \text{Re}\ O(s) + i\ \text{Im}\ O(s) + \text{Re}\ E(s) + i\ \text{Im}\ E(s).$$

Figure 2.5, which records the phenomena in another way, is also valuable for revealing at a glance the "relative sense of oddness": when $f(x)$ is real and odd with a *positive* moment, the odd part of $F(s)$ has i times a *negative* moment; and when $f(x)$ is real but not necessarily odd, we also find opposite senses of oddness. However, inverting the procedure—that is, going from $F(s)$ to $f(x)$, or taking $f(x)$ to be imaginary—produces the same sense of oddness.

Real even functions play a special part in this work because both they and their transforms may easily be graphed. Imaginary odd, real odd, and imaginary even functions are also important in this respect.

Another special kind of symmetry is possessed by a function $f(x)$ whose

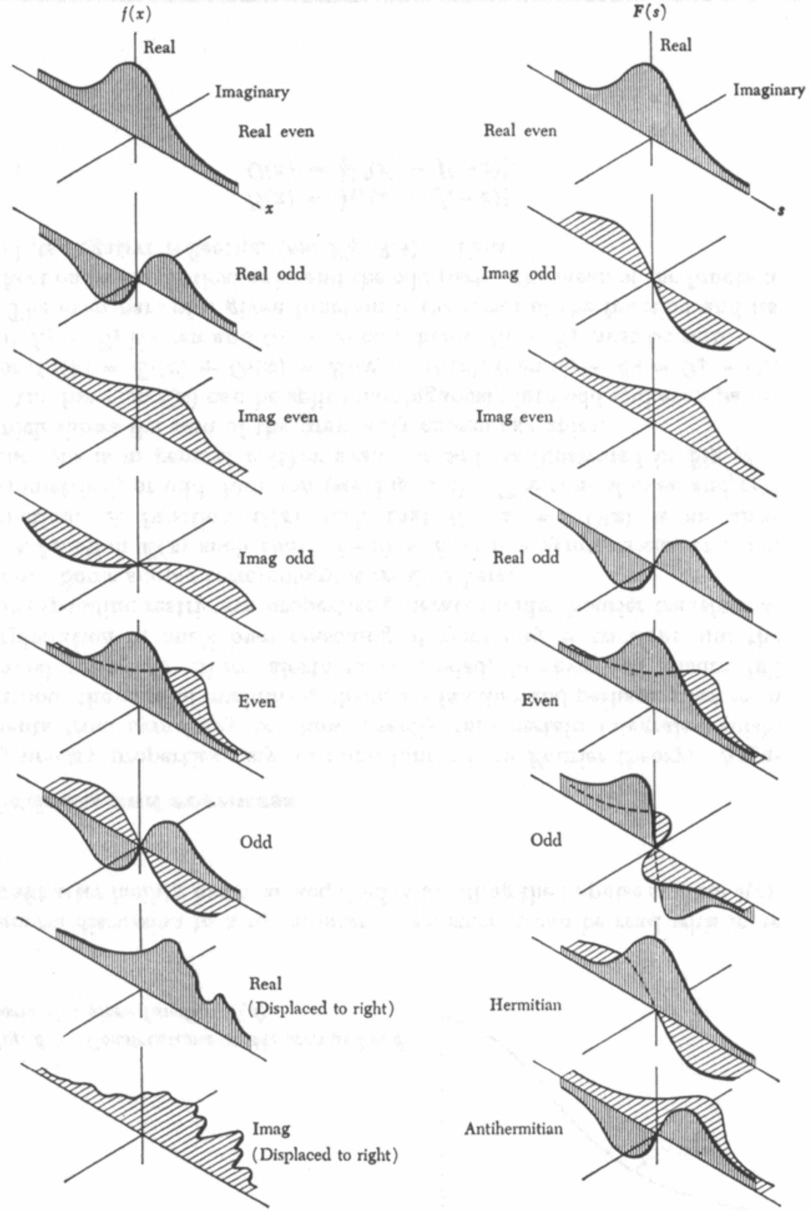


Fig. 2.5 Symmetry properties of a function and its Fourier transform.

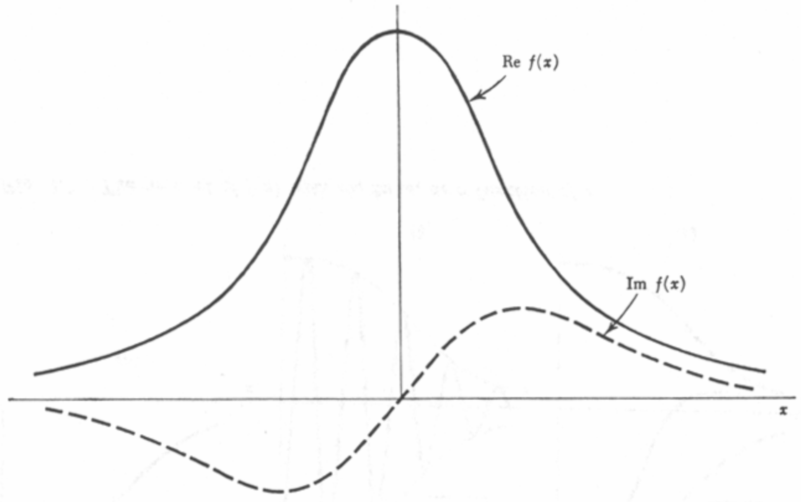


Fig. 2.6 Hermitian functions have their real part even and their imaginary part odd. Their Fourier transform is pure real.

real part is even and imaginary part odd. Such a function will be described as hermitian (see Fig. 2.6); it is often succinctly defined by the property

$$f(x) = f^*(-x),$$

and as mentioned above its Fourier transform is real. As an example of algebraic procedure for handling matters of this kind, consider that

$$f(x) = E + O + i\hat{E} + i\hat{O}.$$

Then
$$f(-x) = E - O + i\hat{E} - i\hat{O}$$
and
$$f^*(-x) = E - O - i\hat{E} + i\hat{O}.$$

If we now require that $f(x) = f^*(-x)$ we must have $O = 0$ and $\hat{E} = 0$. Hence $f(x) = E + i\hat{O}$.

Complex conjugates

The Fourier transform of the complex conjugate of a function $f(x)$ is $F^*(-s)$, that is, the *reflection* of the conjugate of the transform. Special cases of this may be summarized as follows:

$$\text{If } f(x) \text{ is } \begin{cases} \text{real} \\ \text{imaginary} \\ \text{even} \\ \text{odd} \end{cases} \text{ the transform of } f^*(x) \text{ is } \begin{cases} F(s) \\ -F(s) \\ F^*(s) \\ -F^*(s) \end{cases} = F^*(-s).$$

Related statements are tabulated for reference.

$$\begin{aligned} f(x) &\supset F(s) \\ f^*(x) &\supset F^*(-s) \\ f^*(-x) &\supset F^*(s) \\ f(-x) &\supset F(-s) \\ 2\,\mathrm{Re}\, f(x) &\supset F(s) + F^*(-s) \\ 2\,\mathrm{Im}\, f(x) &\supset F(s) - F^*(-s) \\ f(x) + f^*(-x) &\supset 2\,\mathrm{Re}\, F(s) \\ f(x) - f^*(-x) &\supset 2\,\mathrm{Im}\, F(s) \end{aligned}$$

Cosine and sine transforms

The cosine transform of a function $f(x)$ is defined as

$$2\int_0^\infty f(x)\cos 2\pi sx\, dx.$$

The cosine transform is the same as the Fourier transform if $f(x)$ is an even function. In general the even part of the Fourier transform of $f(x)$ is the cosine transform of the even part of $f(x)$.

It will be noted that the cosine transform, as defined, takes no account of $f(x)$ to the left of the origin.

Let $F_c(s)$ represent the cosine transform of $f(x)$. Then the cosine transformation and the reverse transformation by which $f(x)$ is obtained from $F_c(s)$ are identical. Thus

$$F_c(s) = 2\int_0^\infty f(x)\cos 2\pi sx\, dx$$
$$f(x) = 2\int_0^\infty F_c(s)\cos 2\pi sx\, ds.$$

The sine transform of $f(x)$ is defined by

$$F_s(s) = 2\int_0^\infty f(x)\sin 2\pi sx\, dx.$$

This transformation is also identical with its reverse; thus

$$f(x) = 2\int_0^\infty F_s(s)\sin 2\pi sx\, dx.$$

We may say that i times the odd part of the Fourier transform of $f(x)$ is the sine transform of the odd part of $f(x)$.

Combining the sine and cosine transforms of the even and odd parts leads to the Fourier transform of the whole of $f(x)$:

$$\mathfrak{F}f(x) = \mathfrak{F}_c E(x) - i\mathfrak{F}_s O(x),$$

where the operators $\mathfrak{F}$, $\mathfrak{F}_c$, and $\mathfrak{F}_s$ stand for the minus-i Fourier, cosine, and sine transformations, respectively. It will be clear that the terms on the right are the even and odd parts of $F(s)$, not the real and imaginary parts.

If $f(x)$ is zero to the left of the origin, then

$$F(s) = \tfrac{1}{2}F_c(s) - \tfrac{1}{2}iF_s(s),$$

or, to restate this property for any $f(x)$,

$$\tfrac{1}{2}F_c(s) - \tfrac{1}{2}iF_s(s) = \mathfrak{F}f(x)\,H(x),$$

where $H(x)$ is the unit step function (unity when x is positive and zero when x is negative).

In this text the terms "cosine transform" and "sine transform" will be avoided, but they are commonly used elsewhere. The definitions other than the above which may, however, be encountered are

$$\int_0^\infty f(x) \begin{matrix}\cos\\ \sin\end{matrix}\, sx\,dx$$

and
$$\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty f(x) \begin{matrix}\cos\\ \sin\end{matrix}\, sx\,dx.$$

Some advantage would accrue from the adoption of

$$\int_{-\infty}^{\infty} f(x) \cos 2\pi sx\,dx.$$

Interpretation of the formulas

Habitués in Fourier analysis undoubtedly are conscious of graphical interpretations of the Fourier integral. Since the integral contains a complex factor, probably the simpler cosine and sine versions are more often pictured. Thus, given $f(x)$, we picture $f(x) \cos 2\pi sx$ as an oscillation (see Fig. 2.7*a*), lying within the envelope $f(x)$ and $-f(x)$. Twice the

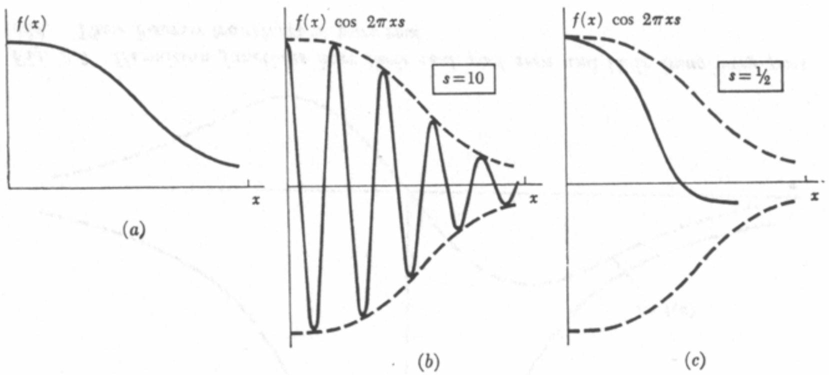


Fig. 2.7 The product of $f(x)$ with $\cos 2\pi xs$, as a function of x.

area under $f(x) \cos 2\pi sx$ is then $F_c(s)$, for $F_c(s) = 2\int_0^\infty f(x) \cos 2\pi sx\,dx$. In Fig. 2.7*b* this area is virtually zero, but a rather high value of s is implied. Figure 2.7*c* is for a low value of s.

The Fourier integral is thus visualized for discrete values of s. The

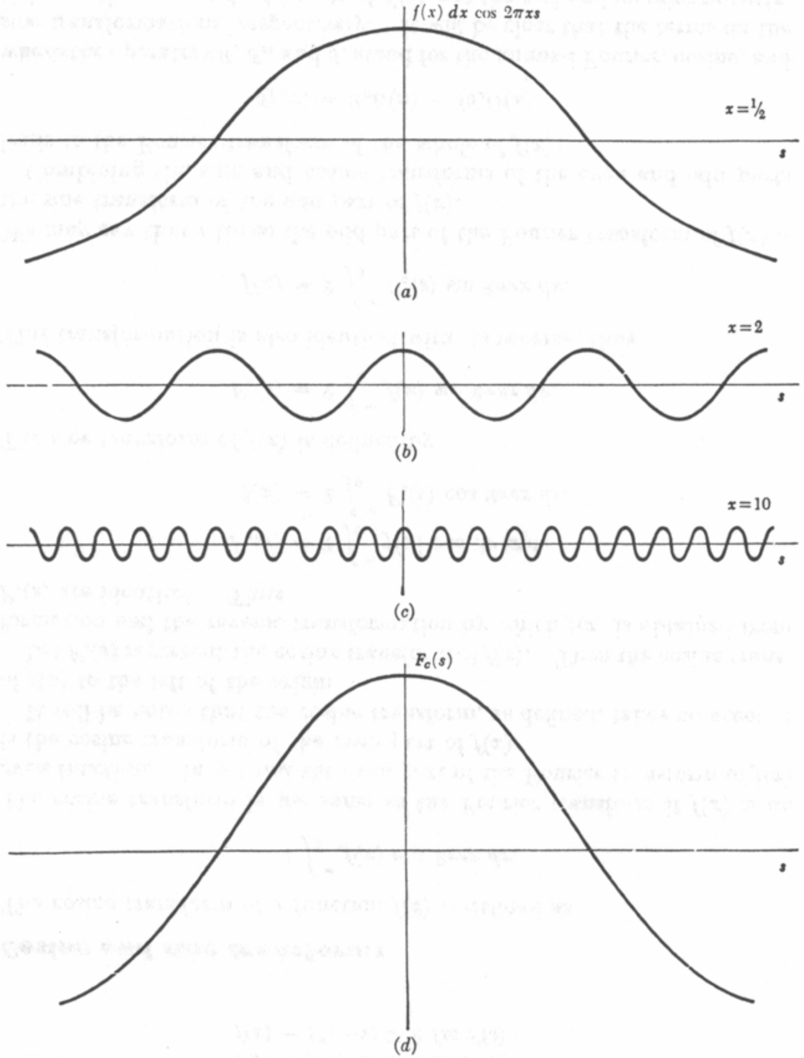


Fig. 2.8 The product of $f(x)\,dx$ with $\cos 2\pi xs$, as a function of s.

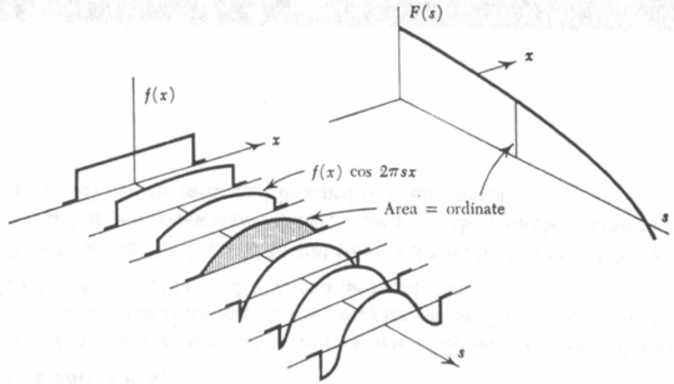


Fig. 2.9 The surface $f(x)$ cos $2\pi sx$ shown sliced in one of two possible ways.

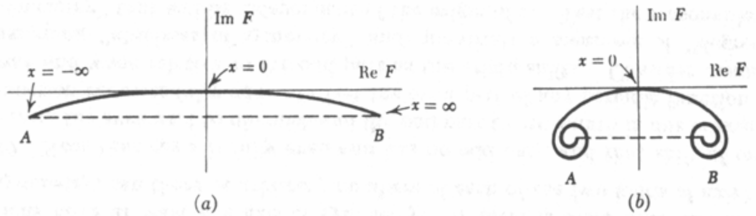


Fig. 2.10 The Fourier integral on the complex plane of $F(s)$ when s is (a) small, (b) large.

interpretation of s is important: s characterizes the frequency of the cosinusoid and is equal to *the number of cycles per unit of x.*

As an exercise in this approach to the matter, contemplate the graphical interpretation of the algebraic statements

$$F_c(s) \begin{cases} \to 0 & s \to \infty \\ = 2\int_0^\infty f(x)\,dx & s = 0. \end{cases}$$

The sine transform may be pictured in the same way, and the complex transform may be pictured as a combination of even and odd parts.

A complementary and equally familiar picture results when the integrand $f(x)$ cos $2\pi xs\,dx$ is regarded as a cosinusoid of amplitude $f(x)\,dx$ and frequency x; that is, it is regarded as a function of s, as in Fig. 2.8*a*. The same thing is shown in Fig. 2.8*b* and Fig. 2.8*c* for other discrete values of x. The summation of such curves for all values of x gives $F_c(s)$. A feeling for this approach to the transform formula is engendered in students of Fourier series by exercises in graphical addition of (co)sinusoids increasing arithmetically in frequency.

Each of the foregoing points of view has dual aspects, according as one ponders the analysis of a function into components or its synthesis from components. The curious fact that whether you analyze or synthesize you do the same thing simply reflects the reciprocal property of the Fourier transform. Figure 2.9 illustrates the first view of the matter. If we visualize the surface represented by the slices shown for particular values of s, and then imagine it to be sliced for particular values of x, we perceive the second view.

In a further point of view we think on the complex plane (see Fig. 2.10), taking s to be fixed. The vector $f(x)\,dx$ is rotated through an angle $2\pi sx$ by the factor exp $(-i2\pi sx)$. As $x \to \pm\infty$, the integrand $f(x)\,dx$ exp $(-i2\pi sx)$ shrinks in amplitude and rotates indefinitely in angle, causing the integral to spiral into two limiting points, A and B. The vector AB represents the infinite integral $F(s)$ as in Fig. 2.10*a*. In Fig. 2.10*b* the more rapid coiling-up for a larger value of s is shown. The behavior of $F(s)$ as $s \to \infty$ and when $s = 0$ is readily perceived.

This kind of diagram, of which Cornu's spiral is an example, is familiar from optical diffraction. It is known as a useful tool both for qualitative thinking and for numerical work in optics. It arises in the propagation of radio waves, and neatly summarizes the behavior of radio echoes reflected from ionized meteor trails as they form. Probably it would be illuminating in fields where its use is not customary.

Problems

1 What condition must $F(s)$ satisfy in order that $f(x) \to 0$ as $x \to \pm\infty$?

2 Prove that $|F(s)|^2$ is an even function if $f(x)$ is real.

3 The Fourier transform in the limit of sgn x is $(i\pi s)^{-1}$. What conditions for the existence of Fourier transforms are violated by these two functions? (sgn x equals 1 when x is positive and -1 when x is negative.)

4 Show that all periodic functions violate a condition for the existence of a Fourier transform.

5 Verify that the function cos x violates one of the conditions for existence of a Fourier transform. Prove that exp $(-\alpha x^2)$ cos x meets this condition for any positive value of α.

6 Give the odd and even parts of $H(x)$, e^{ix}, $e^{-x}H(x)$, where $H(x)$ is unity for positive x and zero for negative x.

7 Graph the odd and even parts of $[1 + (x - 1)^2]^{-1}$.

8 Show that the even part of the product of two functions is equal to the product of the odd parts plus the product of the even parts.

9 Investigate the relationship of $\mathfrak{F}\mathfrak{F}f$ to f when f is neither even nor odd.

10 Show that $\mathfrak{FFFF}f = f$.

11 It is asserted that the odd part of log x is a constant. Could this be correct?

12 Is an odd function of an odd function an odd function? What can be said about odd functions of even functions and even functions of odd functions?

13 Prove that the Fourier transform of a real odd function is imaginary and odd. Does it matter whether the transform is the plus-i or minus-i type?

14 An antihermitian function is one for which $f(x) = -f^*(-x)$. Prove that its real part is odd and its imaginary part even and thus that its Fourier transform is imaginary.

15 Point out the fallacy in the following reasoning. "Let $f(x)$ be an odd function. Then the value of $f(-a)$ must be $-f(a)$; but this is not the same as $f(a)$. Therefore an odd function cannot be even."

16 Let the odd and even parts of a function $f(x)$ be $o(x)$ and $e(x)$. Show that, irrespective of shifts of the origin of x,

$$\int_{-\infty}^{\infty} |o(x)|^2\,dx + \int_{-\infty}^{\infty} |e(x)|^2\,dx = \text{const.}$$

17 Note that the odd and even parts into which a function is analyzed depend upon the choice of the origin of abscissas. Yet the sum of the integrals of the squares of the odd and even parts is a constant that is independent of the choice of origin. What is the constant?

18 Let axes of symmetry of a real function $f(x)$ be defined by values of a such that if o and e are the odd and even parts of $f(x - a)$, then

$$\frac{\int o^2\,dx - \int e^2\,dx}{\int o^2\,dx + \int e^2\,dx}$$

has a maximum or minimum with respect to variation of a. Show that all functions have at least one axis of symmetry. If there is more than one axis of symmetry, can there be arbitrary numbers of each of the two kinds of axis?

19 Note that cos x is fully even and has no odd part, and that shift of origin causes the even part to diminish and the odd part to grow until in due course the function becomes fully odd. In fact the even part of any periodic function will wax and wane relative to the odd part as the origin shifts. Consider means of assigning "abscissas of symmetry" and quantitative measures of "degree of symmetry" that will be independent of the origin of x. Test the reasonableness of your conclusions—for example, on the functions of period 2 which in the range $-1 < x < 1$ are given by $\Lambda(x)$, $\Lambda(x - \frac{1}{2})$, $\Lambda(x - \frac{1}{4})$. See Chapter 4 for triangle-function notation $\Lambda(x)$.

20 The function $f(x)$ is equal to unity when x lies between $-\frac{1}{2}$ and $\frac{1}{2}$ and is zero outside. Draw accurate loci on the complex plane of $F(s)$ from which values of $F(0)$, $F(\frac{1}{2})$, $F(1)$, $F(1\frac{1}{2})$, and $F(2)$ can be measured.

21 The function $f(x)$ is equal to 100 when x differs by less than 0.01 from 1, 2, 3, 4, or 5 and is zero elsewhere. Draw a locus on the complex plane of $F(s)$ from which $F(0.05)$ can be measured in amplitude and phase.

Bibliography

Fourier integral The following are standard texts on the Fourier integral.

Beurling, A.: "Sur les intégrales de Fourier absolument convergentes et leur application à une transformation fonctionnelle" (Neuvième congrès des mathématiciens scandinaves), Helsingfors, Finland, 1939.

Bochner, S.: "Vorlesungen über Fouriersche Integrale," Chelsea Publishing Company, New York, 1948.

Bochner, S., and K. Chandrasekharan: "Fourier Transforms," Princeton University Press, Princeton, N.J., 1949.

Carleman, T.: "L'Intégrale de Fourier et questions qui s'y rattachent," Almqvist and Wiksells, Uppsala, Sweden, 1944.

Carslaw, H. S.: "Introduction to the Theory of Fourier's Series and Integrals," Dover Publications, New York, 1930.

Lighthill, M. J.: "An Introduction to Fourier Analysis and Generalised Functions," Cambridge University Press, Cambridge, England, 1958.

Paley, R. E. A. C., and N. Wiener: "Fourier Transforms in the Complex Domain," vol. 19, American Mathematical Society, Colloquium Publications, New York, 1934.

Sneddon, I. N.: "Fourier Transforms," McGraw-Hill Book Company, New York, 1951.

Titchmarsh, E. C.: "Introduction to the Theory of Fourier Integrals," Oxford University Press, Oxford, England, 1937.

Wiener, N.: "The Fourier Integral and Certain of Its Applications," Cambridge University Press, Cambridge, England, 1933.

Tables of Fourier integrals Since tables of the Laplace transform, properly interpreted, are sources of Fourier integrals, some are included here. The following are the most lengthy compilations.

Campbell, G. A., and R. M. Foster: "Fourier Integrals for Practical Applications," Van Nostrand Company, Princeton, N.J., 1948.

Doetsch, G., H. Kniess, and D. Voelker: "Tabellen zur Laplace Transformation," Springer-Verlag, Berlin, 1947.

Erdélyi, A.: "Tables of Integral Transforms," vol. 1, McGraw-Hill Book Company, New York, 1954.

McLachlan, N. W., and P. Humbert: "Formulaire pour le calcul symbolique," 2d ed., Gauthier-Villars, Paris, 1950.

Tables of the one-sided Laplace transformation (lower limit of integration zero) are necessarily sources of Fourier transforms only of functions that are zero for negative arguments. A table of the double-sided Laplace transform is given in the following work.

Van der Pol, B., and H. Bremmer: "Operational Calculus Based on the Two-sided Laplace Integral," 2d ed., Cambridge University Press, Cambridge, England. 1955.

Chapter 3 Convolution

The *idea* of the convolution of two functions occurs widely, as witnessed by the multiplicity of its aliases. The *word* "convolution" is coming into more general use as awareness of its oneness spreads into various branches of science. The German term *Faltung* is widely used, as is the term "composition product," adapted from the French. Terms encountered in special fields include superposition integral, Duhamel integral, Borel's theorem, (weighted) running mean, cross-correlation function, smoothing, blurring, scanning, and smearing.

As some of these last terms indicate, convolution describes the action of an observing instrument when it takes a weighted mean of some physical quantity over a narrow range of some variable. When, as very often happens, the form of the weighting function does not change appreciably as the central value of the variable changes, the observed quantity is a value of the convolution of the distribution of the desired quantity with the weighting function, rather than a value of the desired quantity itself. All physical observations are limited in this way by the resolving power of instruments, and for this reason alone convolution is ubiquitous. Later we show that the appearance of convolution is coterminous with linearity plus time or space invariance, and also with sinusoidal response to sinusoidal stimulus.

Not only is convolution widely significant as a physical concept, but because of a powerful theorem encountered below, it also offers an advantageous starting point for theoretical developments. Conversely, because of its adaptability to computing, it is an advantageous terminal point for numerical work.

The convolution of two functions $f(x)$ and $g(x)$ is

$$\int_{-\infty}^{\infty} f(u)g(x-u)\,du,$$

or briefly,

$$f(x) * g(x).$$

The convolution itself is also a function of x, let us say $h(x)$.

Various ways of looking at the convolution integral suggest themselves. For example, suppose that $g(x)$ is given. Then for every function $f(x)$ for which the integral exists there will be an $h(x)$. Following Volterra, we may say that $h(x)$ is a *functional* of the function $f(x)$. Note that to calculate $h(x_1)$ we need to know $f(x)$ for a whole range of x, whereas to calculate a *function* of the function $f(x)$ at $x = x_1$, we need only know $f(x_1)$.

In Fig. 3.1 the product $f(u)g(x - u)$ is shown shaded, and the ordinate $h(x)$ is equal to the shaded area.

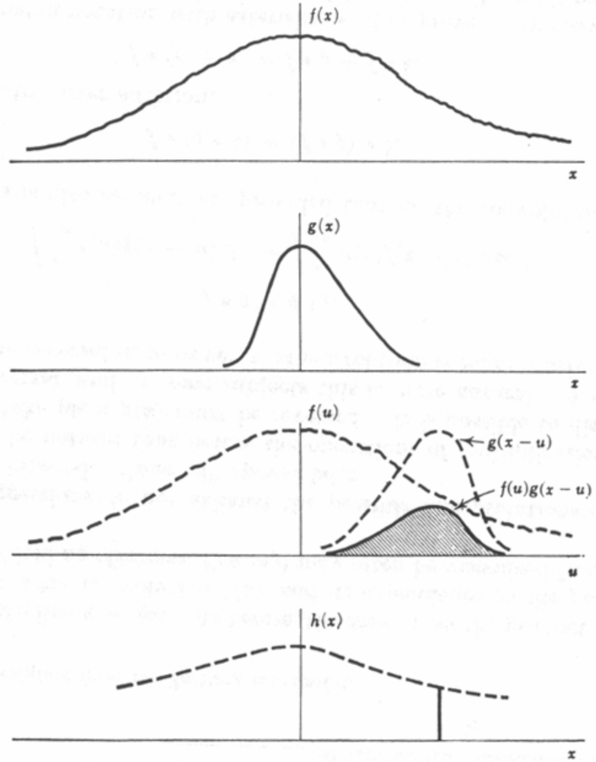


Fig. 3.1 The convolution integral $h(x) = f(x) * g(x)$ represented by a shaded area.

A second example with a different $f(x)$ but the same $g(x)$ is shown in Fig. 3.2 to illustrate the general features relating the functional $h(x)$ to $f(x)$. It will be seen that $h(x)$ is smoother in detail than $f(x)$, is more spread out, and has less total variation.

In another approach, $f(x)$ is resolved into infinitesimal columns (see Fig. 3.3). Each column is regarded as melted out into heaps having the form of $g(x)$ but centered at its original value of x. Just two of these melted-down columns are shown in the figure, but all are to be so pictured. Then $h(x)$ is equal to the sum of the contributions at the point x made by all the heaps; that is,

$$h(x) = \int_{-\infty}^{\infty} f(x_1)g(x - x_1)\,dx_1.$$

A further view is illustrated in Fig. 3.4, where $g(u)$ is shown folded back on

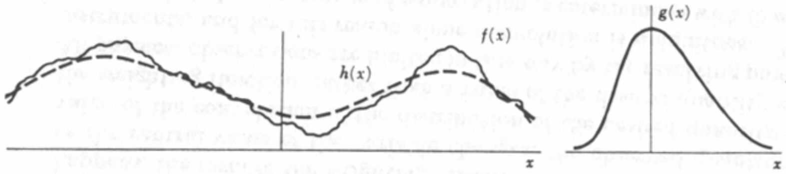


Fig. 3.2 Illustrating the smoothing effect of convolution ($h = f * g$).

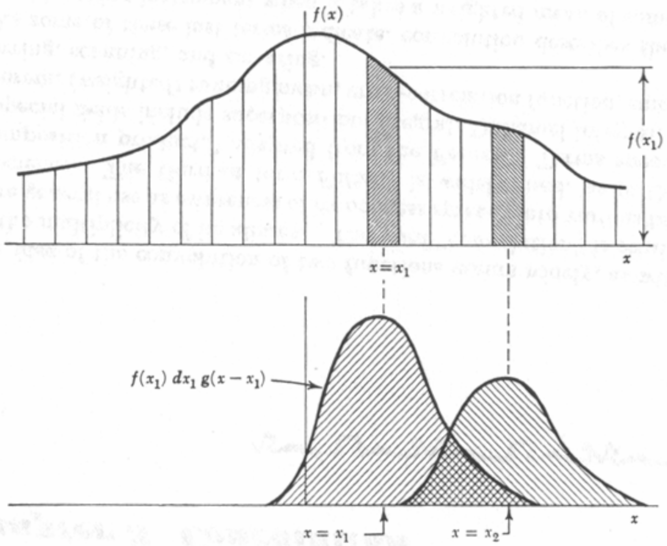


Fig. 3.3 The convolution integral regarded as a superposition of characteristic contributions.

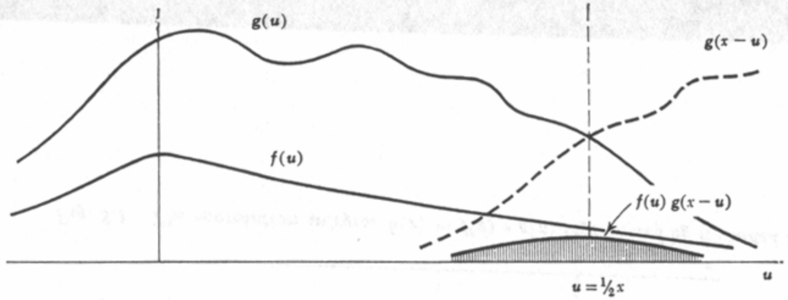


Fig. 3.4 Convolution from the Faltung *standpoint.*

itself about the line $u = \frac{1}{2}x$. As before, the area under the product curve $f(u)g(x - u)$ is the convolution $h(x)$, and its dependence on the position of the line of folding (German *Faltung*) may often be visualized from this standpoint.

These suggestions do not exhaust the possible interpretations of the convolution integral; others will appear later.

It should be noticed that before the operations of multiplication and integration take place $g(x)$ must be reversed. It is possible to dispense with the reversal, and in some subjects this is more natural. A consequence of the reversal is, however, that convolution is commutative; that is,

$$f * g = g * f,$$

or

$$\int_{-\infty}^{\infty} f(u)g(x - u)\,du = \int_{-\infty}^{\infty} g(u)f(x - u)\,du.$$

Convolution is also associative (provided that all the convolution integrals exist),

$$f * (g * h) = (f * g) * h,$$

and distributive over addition,

$$f * (g + h) = f * g + f * h.$$

The abbreviated notation with asterisks (*) thus proves very convenient in formal manipulation, since the asterisks behave like multiplication signs.

Examples of convolution

Consider the truncated exponential function

$$E(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x < 0. \end{cases}$$

We shall calculate the convolution between two such truncated exponentials with different (positive) decay constants. Thus

$$\begin{aligned} aE(\alpha x) * bE(\beta x) &= ab \int_{-\infty}^{\infty} E(\alpha u)E(\beta x - \beta u)\,du \\ &= abE(\beta x) \int_0^x E(\alpha u - \beta u)\,du \\ &= abE(\beta x)\,\frac{E(\alpha x - \beta x) - 1}{\beta - \alpha} \\ &= ab\,\frac{E(\alpha x) - E(\beta x)}{\beta - \alpha}. \end{aligned}$$

The result is thus the difference of two truncated exponentials, each with the same amplitude, as illustrated in Fig. 3.5. This function occurs commonly; for instance, it describes the concentration of a radioactive isotope which decays with a constant α while simultaneously being replenished as the decay product of a parent isotope which decays with a constant β. From the commutative property of convolution we see that the result is the same if α and β are interchanged, and as $t \to \infty$ one of the terms dies out, leaving a simple exponential with constant α or β, whichever describes the slower decay. As a special case we calculate $E(\alpha x) * E(\alpha x)$ by taking the limit as $\beta - \alpha \to 0$. Thus we find

$$\begin{aligned} E(\alpha x) * E(\alpha x) &= \lim_{\beta - \alpha \to 0} \frac{E(\alpha x) - E(\beta x)}{\beta - \alpha} \\ &= -\frac{d}{d\alpha} E(\alpha x) \\ &= xE(\alpha x). \end{aligned}$$

This particular function describes the response of a critically damped resonator, such as a dead-beat galvanometer, to an impulsive disturbance.

As a further example consider $E(-\alpha x) * E(\beta x)$. This gives an entirely

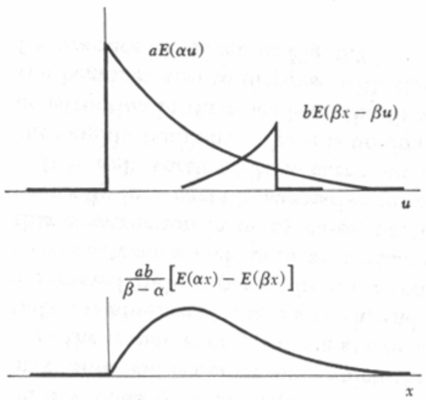


Fig. 3.5 Convolution of two truncated exponentials.

different type of result. Thus

$$\begin{aligned} E(-\alpha x) * E(\beta x) &= \int_{-\infty}^{\infty} E(-\alpha x + \alpha u)E(\beta u)\,du \\ &= \begin{cases} \int_x^{\infty} e^{-\alpha u + \alpha x} e^{-\beta u}\,du & x > 0 \\ \int_0^{\infty} e^{-\alpha u + \alpha x} e^{-\beta u}\,du & x < 0 \end{cases} \\ &= \frac{E(-\alpha x) + E(\beta x)}{\alpha + \beta}. \end{aligned}$$

Here we have a function that is peaked at the origin and dies away with a constant α to the left and with a constant β to the right.

In calculations of this kind care is required in fixing the limits of the integrals and checking signs, because the ordinary sort of algebraic error can make a radical change in the result. The following graphical construction for convolution is useful as a check. Plot one of the functions entering into the convolution backward on a movable piece of paper as shown in Fig. 3.6, and slide it along in the direction of the axis of abscissas.

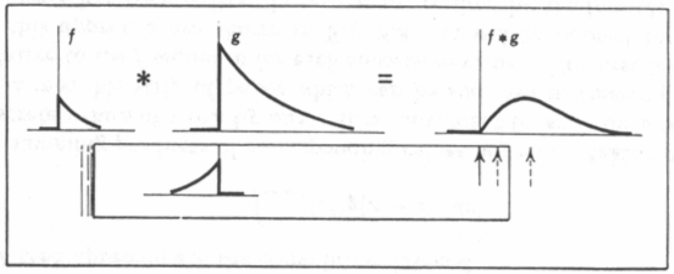


Fig. 3.6 Graphical construction for convolution. The movable piece of paper has a graph of one of the functions plotted backward.

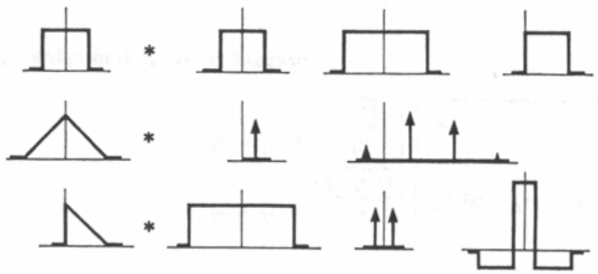


Fig. 3.7 Examples for practicing graphical convolution. The arrows represent impulses.

When the movable piece is to the left of the position shown, the product of g with f reversed is zero. By marking an arrow in some convenient position, we can keep track of this. Then suddenly, at the position shown, the integral of the product begins to assume nonzero values. By moving the paper a little farther along, as indicated by a broken outline, we find that the convolution will be positive and increasing from zero approximately linearly with displacement. Farther along still, we see that a maximum will occur beyond which the convolution dies away.

In the second example given above, two oppositely directed exponential tails, the absence of a zero stretch, and the presence of a cusp at the origin are immediately apparent from the construction. In addition to qualitative conclusions such as these, certain quantitative results are yielded by this construction in many cases, particularly where the functions break naturally into parts in successive ranges of the abscissa.

It is well worth while to carry out this moving construction until it is thoroughly familiar. There is no doubt that experts do this geometrical construction in their heads all the time, and a little practice soon enables the beginner also to dispense with the actual piece of paper. Examples for practice are given in Fig. 3.7.

Serial products

Consider two polynomials

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots$$

and

$$b_0 + b_1x + b_2x^2 + b_3x^3 + \cdots.$$

Their product is

$$a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + (a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0)x^3 + \cdots,$$

which we may call

$$c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots,$$

where
$$\begin{aligned} c_0 &= a_0b_0 \\ c_1 &= a_0b_1 + a_1b_0 \\ c_2 &= a_0b_2 + a_1b_1 + a_2b_0 \\ c_3 &= a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0. \end{aligned}$$

This elementary observation has an important connection with convolution. Suppose that two functions f and g are given, and that it is required to calculate their convolution numerically. We form a sequence of values of f at short regular intervals of width w,

$$\{f_0\ f_1\ f_2\ f_3\ \ldots\ f_m\},$$

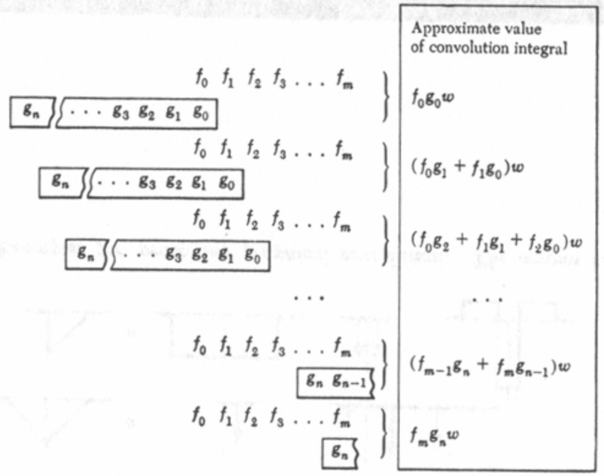


Fig. 3.8 Explaining the serial product.

and a corresponding sequence of values of g,

$$\{g_0\ g_1\ g_2\ g_3\ \ldots\ g_n\}.$$

We then approximate the convolution integral

$$\int_{-\infty}^{\infty} f(x')g(x - x')\,dx'$$

by summing products of corresponding values of f and g, taking different discrete values of x one by one. It is convenient to write the g sequence on a movable strip of paper which can be slid into successive positions relative to the f sequence for each successive value. The first few stages of this approach are shown in Fig. 3.8. It will be noticed that the g sequence has been written in reverse as required by the formula. Since $f * g = g * f$, the f sequence could have been written in reverse, in which case *it* would have been written on the movable strip.

It will be seen that this procedure generates the same expressions that occur in the multiplication of series, and we therefore introduce the term "serial product" to describe the sequence of numbers

$$\{f_0g_0\quad f_0g_1 + f_1g_0\quad f_0g_2 + f_1g_1 + f_2g_0\quad \ldots\}$$

derived from the two sequences

$$\{f_0\ f_1\ f_2\ f_3\ \ldots\} \quad \text{and} \quad \{g_0\ g_1\ g_2\ g_3\ \ldots\}.$$

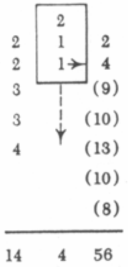


Fig. 3.9 Calculating a serial product by hand.

We transfer the asterisk notation to represent this relationship between the three sequences as follows:

$$\{f_0\ f_1\ .\ .\ .\ f_m\} * \{g_0\ g_1\ .\ .\ .\ g_n\} = \{f_0g_0\quad f_0g_1 + f_1g_0\quad .\ .\ .\ f_mg_n\}.$$

Alternatively, we may define the $(i + 1)$th term of the serial product of $\{f_i\}$ and $\{g_i\}$ to be

$$\sum_j f_j g_{i-j}.$$

In practice the calculation of serial products is an entirely feasible procedure (for example, on a hand calculator by allowing successive products to accumulate in the product register). The two sequences are most conveniently written in vertical columns, and the answers are written opposite an arrow marked in a convenient place on the movable strip. Figure 3.9 shows an early stage in the calculation of $\{2\ 2\ 3\ 3\ 4\} * \{1\ 1\ 2\}$. The value of 4, shown opposite the arrow, has just been calculated, and the values still to be calculated as the movable strip is taken downward are shown in parentheses. Note that the sequence on the *moving* strip has been written in *reverse* (upward).

It will be seen that the serial product is a longer sequence than either of the component sequences, the number of terms being one less than the sum of the numbers of terms in the components:

$$\{2\ 2\ 3\ 3\ 4\} * \{1\ 1\ 2\} = \{2\ 4\ 9\ 10\ 13\ 10\ 8\}.$$

Furthermore the sum of the terms of the serial product is the product of the sums of the component sequences, a fact which allows a very valuable check on numerical work. As a special case, if the sum of one of the component sequences is unity, then the sum of the serial product will be the same as the sum of the other component. These properties are analogous to properties of the convolution integral.

A semi-infinite sequence is one such as

$$\{f_0\ f_1\ f_2\ .\ .\ .\},$$

which has an end member in one direction but runs on without end in the other. If we take the serial product of such a sequence with a finite sequence, the result is also semi-infinite; for example,

$$\{1\ 1\} * \{1\ 2\ 3\ 4\ .\ .\ .\} = \{1\ 3\ 5\ 7\ 9\ .\ .\ .\}.$$

A semi-infinite sequence may or may not have a finite sum, but this does not lead to problems in defining the serial product because each member is the sum of only a finite number of terms. Of course, the numerical check, according to which the sum of the members of the serial product equals the product of the sums, breaks down if any one of the sums does not exist.

The serial product of two semi-infinite sequences presents no problems when both run on indefinitely in the same direction; thus

$$\{1\ 2\ 3\ 4\ .\ .\ .\} * \{1\ 2\ 3\ 4\ .\ .\ .\} = \{1\ 4\ 10\ 20\ .\ .\ .\},$$

but if they run on in opposite directions then each member of the serial product is the sum of an infinite number of terms. This sum may very well not exist; thus

$$\{.\ .\ .\ 4\ 3\ 2\ 1\} * \{1\ 2\ 3\ 4\ .\ .\ .\}$$

does not lead to convergent series.

Two-sided sequences are often convenient to deal with and call for no special comment other than that one must specify the origin explicitly, for example, by an arrow as in

$$\{.\ .\ .\ 0.1\ 0.2\ 0.4\ \underset{\uparrow}{0.9}\ 0.8\ 0.7\ 0.6\ .\ .\ .\}.$$

The sequence

$$\{.\ .\ .\ 0\ 0\ 0\ \underset{\uparrow}{1}\ 0\ 0\ 0\ .\ .\ .\} = \{J\}$$

plays an important role analogous to that of the impulse symbol $\delta(x)$. It has the property that

$$\{J\} * \{f\} = \{f\}$$

for all sequences $\{f\}$. This property is, of course, also possessed by the one-member sequence $\{1\}$ and by other sequences such as $\{1\ 0\ 0\}$.

Serial multiplication by the sequence

$$\{.\ .\ .\ 0\ 0\ 0\ 1\ -1\ 0\ 0\ 0\ .\ .\ .\}\qquad \text{or}\qquad \{1\ -1\}$$

is equivalent to taking the first finite difference; that is,

$$\{1\ -1\} * \{.\ .\ .\ f_{-2}\ f_{-1}\ f_0\ f_1\ f_2\ .\ .\ .\}$$
$$= \{.\ .\ .\ f_{-1} - f_{-2}\quad f_0 - f_{-1}\quad f_1 - f_0\quad f_2 - f_1\quad .\ .\ .\}.$$

Where it is important to distinguish between the central difference

$f_{n+\frac{1}{2}} - f_{n-\frac{1}{2}}$ and the forward difference $f_{n+1} - f_n$, it is necessary to specify the origins of the sequences appropriately. Apart from a shift in the indexing of the terms, however, the resulting sequence is the same.

The sequence

$$\left\{\frac{1}{n}\ \frac{1}{n}\ \ldots\ (n \text{ terms})\ \ldots\ \frac{1}{n}\right\}$$

generates the running mean over n terms, a familiar process for smoothing sequences of meteorological data. It may be written, as in the case of weekly running means of daily values, as

$$\frac{1}{7}\{1\ 1\ 1\ 1\ 1\ 1\ 1\}.$$

The sequence $\{1\ 1\ 1\ 1\ 1\ 1\ 1\}$ itself generates running sums, in this case over seven terms.

The semi-infinite sequence

$$\{1\ 1\ 1\ 1\ \ldots\}$$

enters into a serial product with another sequence to generate a familiar result. Thus if $\{f\}$ is a sequence and S_n is the sum of the first n terms, then the sequence $\{S\}$ defined by

$$\{S\} = \{S_1\ S_2\ S_3\ \ldots\}$$

may be expressed in the form

$$\{1\ 1\ 1\ 1\ \ldots\} * \{f\} = \{S\}.$$

Since each term of the sequence $\{f\}$ is the difference between two successive terms of $\{S\}$, a converse relationship can also be written:

$$\{1\ -1\} * \{S\} = \{f\}.$$

Substituting in the previous equation, we have

$$\{1\ 1\ 1\ 1\ \ldots\} * \{1\ -1\} * \{S\} = \{S\};$$

in other words the two operations neutralize each other. We may express this reciprocal relationship by writing

$$\{1\ 1\ 1\ 1\ \ldots\} * \{1\ -1\} = \{1\ 0\ 0\ 0\ \ldots\}$$

or, in a two-sided version,

$$\{\ldots\ 0\ 0\ 0\ \underset{\uparrow}{1}\ 1\ 1\ 1\ \ldots\} * \{\ldots\ 0\ 0\ 0\ \underset{\uparrow}{1}\ -1\ 0\ 0\ \ldots\} = \{\ldots\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ \ldots\}.$$

Another example of a reciprocal pair of sequences is

$$\{1\ 2\ 3\ 4\ \ldots\} \quad \text{and} \quad \{1\ -2\ 1\}.$$

This is an important relationship because the reciprocal sequence represents the solution to the problem of finding $\{g\}$ when $\{f\}$ and $\{h\}$ are given and it is known that

$$\{f\} * \{g\} = \{h\}.$$

The answer is

$$\{g\} = \{f\}^{-1} * \{h\},$$

where $\{f\}^{-1}$ is the reciprocal of $\{f\}$, that is, the sequence with the property that

$$\{f\}^{-1} * \{f\} = \{\ldots\ 0\ 0\ 1\ 0\ 0\ \ldots\}.$$

Inversion of serial multiplication If $\{f\} * \{g\} = \{h\}$, then $\{h\}$ is called the serial product of $\{f\}$ and $\{g\}$, because the sequence $\{h\}$ comprises the coefficients of the polynomial which is the product of the polynomials represented by $\{f\}$ and $\{g\}$.

Conversely, the process of finding $\{g\}$ when $\{f\}$ and $\{h\}$ are given may be called serial division, and in fact one could carry out the solution of such a problem by actually doing long division of polynomials.

For example, to solve the problem

$$\{1\ 1\} * \{g_0\ g_1\ g_2\ \ldots\} = \{1\ 3\ 3\ 1\}$$

one could write

$$g_0 + g_1x + g_2x^2 + \ldots = \frac{1 + 3x + 3x^2 + x^3}{1 + x}.$$

The long division would then proceed as follows:

$$\begin{array}{r|l} & 1 + 2x + x^2 \\ \hline 1 + x & 1 + 3x + 3x^2 + x^3 \\ & \underline{1 + x} \\ & 2x + 3x^2 \\ & \underline{2x + 2x^2} \\ & x^2 + x^3 \\ & \underline{x^2 + x^3} \end{array}$$

Thus $\quad g_0 + g_1x + g_2x^2 + \ldots = 1 + 2x + x^2$

or $\quad \{g_0\ g_1\ g_2\ \ldots\} = \{1\ 2\ 1\}.$

This, of course, is recognizable immediately as the correct answer.

However, it is not necessary to organize the calculation at such length. A convenient way of doing the job on a hand calculator may be shown by going back to an example in the previous section, where $\{2\ 2\ 3\ 3\ 4\} *$

$\{1\ 1\ 2\}$ was calculated. The situation at the end of the calculation is as shown in Fig. 3.10*a*, the serial product being in the right-hand column. Now, suppose that the left-hand column $\{f\}$ and the right-hand column $\{h\}$ were given. To find $\{g\}$, write the sequence upward on the movable strip of paper, and place it in position for beginning the calculation of the serial product (see Fig. 3.10*b*). Clearly g_0 is immediately deducible, for it has to be such that when it is multiplied by the 2 on the left the result is 2. Hence in the space labeled g_0 one may write 1 and then move the paper down to the next position (see Fig. 3.10*c*). Then, when g_1 is multiplied by 2 and added to the product of 2 and 1, the result must be 4. This gives g_1, and so on.

With a hand calculator one forms the products of the numbers in the left-hand column with such numbers on the moving strip as are already known and accumulates the sum of the products. Subtract this sum from the current value of $\{h\}$ opposite the arrow and divide by the first member of $\{f\}$ at the top of the left column to obtain the next value of $\{g\}$. Enter this value on the paper strip, which may then be moved down one more place. The process is started by entering the first value of $\{g\}$, which is h_0/f_0. This inversion of serial multiplication is as easy as the direct process and takes very little longer.

Compared with long division, this method is superior, since it involves writing down no numbers other than the data and the answer. Furthermore it is as readily applicable when the coefficients are large numbers or fractions. Long division might, however, be considered for short sequences of small integers.

Table 3.1 lists a few common sequences and their inverses. Most of

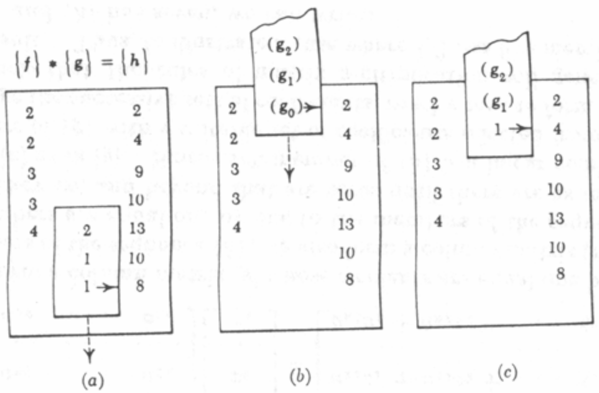


Fig. 3.10 (*a*) *Completion of serial multiplication of* $\{2\ 2\ 3\ 3\ 4\}$ by $\{1\ 1\ 2\}$; (*b*) *and* (*c*) *first steps in the inversion of the process.*

Table 3.1 Some sequences and their inverses

Sequence	*Inverse*
$\{1\ 1\}$	$\{1\ -1\ 1\ -1\ 1\ -1\ \ldots\}$
$\{1\ -1\}$	$\{1\ 1\ 1\ 1\ 1\ 1\ \ldots\}$
$\{1\ 0\ 1\}$	$\{1\ 0\ -1\ 0\ 1\ 0\ \ldots\}$
$\{1\ 0\ -1\}$	$\{1\ 0\ 1\ 0\ 1\ 0\ \ldots\}$
$\{1\ 1\ 1\}$	$\{1\ -1\ 0\ 1\ -1\ 0\ \ldots\}$
$\{1\ 2\ 1\}$	$\{1\ -2\ 3\ -4\ 5\ -6\ \ldots\}$
$\{n+1\} = \{1\ 2\ 3\ 4\ \ldots\}$	$\{1\ -1\}^{*2} = \{1\ -2\ 1\}$
$\{(n+1)^2\} = \{1\ 4\ 9\ 16\ \ldots\}$	$\{1\ -1\}^{*3} * \{1\ 1\}^{-1}$
$\{(n+1)^3\} = \{1\ 8\ 27\ \ldots\}$	$\{1\ -1)^{*4} * \{1\ 4\ 1\}^{-1}$
$\{1\ -a\}$	$\{1\ a\ a^2\ a^3\ \ldots\}$
$\{1\ -a\}^{*2}$	$\{1\ 2a\ 3a^2\ 4a^3\ \ldots\}$
$\{e^{-\alpha n}\} = \{1\ e^{-\alpha}\ e^{-2\alpha}\ \ldots\}$	$\{1\ -e^{-\alpha}\}$
$\{1 - e^{-\alpha(n+1)}\}$	$\{1\ -1\} * \{1\ -e^{-\alpha}\}/(1 - e^{-\alpha})$
$\{\sin \omega(n+1)\}$	$\{1\ -2\cos\omega\ 1\}/\sin\omega$
$\{\cos \omega n\}$	$\{1\ -2\cos\omega\ 1\} * \{1\ -\cos\omega\}^{-1}$
$\{e^{-\alpha(n+1)} \sin \omega(n+1)\}$	$\{1\ -2e^{-\alpha}\cos\omega\ e^{-2\alpha}\}/e^{-\alpha}\sin\omega$
$\{e^{-\alpha n}\cos\omega n\}$	$\{1\ -2e^{-\alpha}\cos\omega\ e^{-2\alpha}\} * \{1\ -e^{-\alpha}\cos\omega\}^{-1}$

the entries may be verified by serial multiplication of corresponding entries, whereupon the result will be found to be $\{1\ 0\ 0\ 0\ \ldots\}$. The table is precisely equivalent to a list of polynomial pairs beginning as follows:

$$\begin{array}{ll} 1+x & 1-x+x^2-x^3+\ldots \\ 1-x & 1+x+x^2+x^3+\ldots \\ 1+x^2 & 1-x^2+x^4+\ldots \\ \cdots & \cdots \end{array}$$

In this case the property of corresponding entries would be that their product was unity.

Some of the later entries in Table 3.1 represent exponential and other variations reminiscent of the natural behavior of circuits, and will prove more difficult, though not impossible, to verify at this stage. They are readily derivable by transform methods.

The serial product in matrix notation Let the sequence $\{h\}$ be the serial product of two sequences $\{f\}$ and $\{g\}$, where

$$\begin{aligned} \{f\} &= \{f_0\ f_1\ f_2\ \ldots\ f_m\} \\ \{g\} &= \{g_0\ g_1\ g_2\ \ldots\ g_n\} \\ \{h\} &= \{h_0\ h_1\ h_2\ \ldots\ h_{m+n}\}. \end{aligned}$$

Evidently all three sequences can be expressed as single-row or single-column matrices, but since the sequences have different numbers of members, matrix notation might not immediately spring to mind. The relationship between $\{f\}$, $\{g\}$, and $\{h\}$ can, however, be expressed in terms of matrix multiplication as follows.

First we recall the special case of a matrix product where the first factor has n rows and n columns, and the second has n rows and one column only. The product is a single-column matrix of n elements. Thus

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \qquad \cdots \qquad \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix}$$

Now we form a column matrix $[y]$ whose elements are equal one by one to the members of the sequence $\{h\}$; we also form a column matrix $[x]$, whose early members are equal one by one to the members of the sequence $\{g\}$ as far as they go, and beyond that are zeros until there are as many elements in $[x]$ as in $[y]$. Since each member of $\{h\}$ is a linear combination of members of $\{g\}$ with a suitable set of coefficients selected from $\{f\}$, we can arrange the successive sets of coefficients, row by row, to form a square matrix such that the rules of matrix multiplication will generate the desired result. Thus, to illustrate a case where $\{f\}$ has five members, $\{g\}$ has three, and $\{h\}$ has seven, we can write:

$$\begin{bmatrix} f_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ f_1 & f_0 & 0 & 0 & 0 & 0 & 0 \\ f_2 & f_1 & f_0 & 0 & 0 & 0 & 0 \\ f_3 & f_2 & f_1 & f_0 & 0 & 0 & 0 \\ f_4 & f_3 & f_2 & f_1 & f_0 & 0 & 0 \\ 0 & f_4 & f_3 & f_2 & f_1 & f_0 & 0 \\ 0 & 0 & f_4 & f_3 & f_2 & f_1 & f_0 \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} f_0g_0 \\ f_1g_0 + f_0g_1 \\ f_2g_0 + f_1g_1 + f_0g_2 \\ f_3g_0 + f_2g_1 + f_1g_2 \\ f_4g_0 + f_3g_1 + f_2g_2 \\ f_4g_1 + f_3g_2 \\ f_4g_2 \end{bmatrix} = \begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \end{bmatrix}$$

The elements in the southeast quadrant of the f matrix could be replaced by zeros without affecting the result.

By listing parts of the sequence $\{f\}$ row by row with progressive shift as required, we have succeeded in forcing the serial product into matrix notation. This may appear to be a labored and awkward exercise, and one that sacrifices the elegant commutative property. In fact, however, matrix representation for serial products is widely used in control-system engineering and elsewhere because of the rich possibilities offered by the theory of matrices, especially infinite matrices.

Sequences as vectors Just as the short sequence $\{x_1\ x_2\ x_3\}$ may be regarded as the representation of a certain vector in three-dimensional space in terms of three orthogonal components, so a sequence of n members may be regarded as representing a vector in n-dimensional space. As a rule we do not expend much effort trying to visualize the n mutually perpendicular axes along which the vector is to be resolved, but simply handle the members of the sequence according to algebraic rules which are natural extensions of those for handling vector components in three-dimensional space. Nevertheless, we borrow a whole vocabulary of terms and expressions from geometry, which lends a certain picturesqueness to linear algebra.

If the relationship between two sequences $\{x\}$ and $\{y\}$ is

$$\{a\} * \{x\} = \{y\},$$

where $\{a\}$ is some other sequence, then an equivalent statement is

$$T\mathbf{X} = \mathbf{Y},$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the vectors whose components are $\{x\}$ and $\{y\}$, and T is an operator representing the transformation that converts $\mathbf{X}$ to $\mathbf{Y}$. The transformation consists of forming linear combinations of the components of $\mathbf{X}$ in accordance with the rules that are so concisely expressed by the serial product.

It will be recalled that the number of members of the sequence $\{y\}$ exceeds the number of members of $\{x\}$ if $\{x\}$ has a finite number of members. To avoid the awkwardness that arises if $\mathbf{X}$ and $\mathbf{Y}$ are vectors in spaces with different numbers of dimensions, the vector interpretation is usually applied to infinite sequences $\{x\}$ and $\{y\}$. If the sequences arising in a given problem are in actual fact not infinite, it is often permissible to convert them to infinite sequences by including an infinite number of extra members all of which are zero.

As in the case of representation of sequences by column matrices, the commutative property $\{a\} * \{x\} = \{x\} * \{a\}$ is abandoned. One of the sequences entering into the serial product is interpreted as a vector, and one in an entirely different way. To reverse the roles of $\{x\}$ and $\{a\}$—for example, to talk of the vector $\mathbf{A}$—would be unthinkable in physical fields where the vector interpretation is used. This apparent rigidification of the notation is compensated by greater generality. To illustrate this we may note first that, to the extent that matrix methods are used for dealing with vectors, there is no distinction between the representations of sequences as column matrices and as vectors. Now, examination of the square matrix in the preceding section that was built up from members of the sequence $\{f\}$ reveals that in each row the sequence $\{f\}$ was written backward, the sequence shifting one step from each row to the next. This reflects the mode of calculation of the serial product, where the sequence $\{f\}$ would be written on a movable strip of paper alongside a sequence $\{g\}$ and shifted one step after each cycle of forming products and adding. In the procedure for taking serial products there is no provision for changing the sequence $\{f\}$ from step to step. Indeed it is the essence of convolution, and of serial multiplication, that no such change occurs. But in the square-matrix formulation, it is just as convenient to express such changes as not. Thus the matrix notation allows for a general situation of which the serial product is a special case.

This more general situation is simply the general linear transformation as contrasted with the general shift-invariant linear transformation. If the indexing of a sequence is with respect to time, then serial multiplication is the most general time-invariant linear transformation, and would for example be applied in problems concerned with linear filters whose elements did not change as time went by. If, on the other hand, one were dealing with the passage of a signal through a filter containing time-varying linear elements such as motor-driven potentiometers, then the relationship between output and input could not be expressed in the form of a serial product. In such circumstances, where convolution is inapplicable, the property referred to loosely as "harmonic response to harmonic excitation" also breaks down.

The autocorrelation function

The self-convolution of a function $f(x)$ is given by

$$f * f = \int_{-\infty}^{\infty} f(u)f(x-u)\,du.$$

Suppose, however, that prior to multiplication and integration we do not reverse one of the two component factors; then we have the integral

$$\int_{-\infty}^{\infty} f(u)f(u-x)\,du,$$

which may be denoted by $f \star f$. A single value of $f \star f$ is represented by

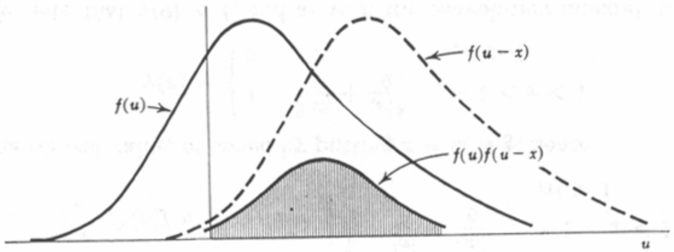


Fig. 3.11 The autocorrelation function represented by an area (shown shaded).

the shaded area in Fig. 3.11. A moment's thought will show that if the function f is to be displaced relative to itself by an amount x (without reversal), then the integral of the product will be the same whether x is positive or negative. In other words, if $f(x)$ is a real function, then $f \star f$ is an even function, a fact which is not true in general of the convolution integral. It follows that

$$f \star f = \int_{-\infty}^{\infty} f(u)f(u-x)\,du = \int_{-\infty}^{\infty} f(u)f(u+x)\,du,$$

which, of course, is deducible from the previous expression by substitution of $w = u - x$.

It is shown in the appendix to this chapter that the value of $f \star f$ at its origin is a maximum; that is, as soon as some shift is introduced, the integral of the product falls off.

When $f(x)$ is complex it is customary to use the expression

$$\int_{-\infty}^{\infty} f(u)f^*(u-x)\,du$$

or

$$\int_{-\infty}^{\infty} f^*(u)f(u+x)\,du.$$

(Note that it is possible to place the asterisk in the wrong place and thus obtain the conjugate of the standard version.)

It is often convenient to normalize by dividing by the central value. Then we may define a quantity $\gamma(x)$ given by

$$\gamma(x) = \frac{\int_{-\infty}^{\infty} f^*(u)f(u+x)\,du}{\int_{-\infty}^{\infty} f(u)f^*(u)\,du}$$

and it is clear that

$$\gamma(0) = 1.$$

We shall refer to $\gamma(x)$ as the autocorrelation function of $f(x)$. However, it often happens that the question of normalization is unimportant in a particular application, and the character of the autocorrelation is of more interest than the magnitude; then the nonnormalized form is referred to as the autocorrelation function.

As an example, take the function $f(x)$ defined by

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 - x & 0 < x < 1 \\ 0 & x > 1. \end{cases}$$

Then, for $0 < x < 1$,

$$\int_{-\infty}^{\infty} f^*(u)f(u+x)\,du = \int_0^{1-x} (1-u)[1-(u+x)]\,du = \frac{1}{3} - \frac{x}{2} + \frac{x^3}{6}.$$

The best way to determine the limits of integration is to make a graph such as the one in Fig. 3.11. Where $x > 1$, the integral is zero, and since the integral is known to be an even function of x, we have

$$\int_{-\infty}^{\infty} f^*(u)f(u+x)\,du = \begin{cases} \frac{1}{3} - \frac{|x|}{2} + \frac{|x|^3}{6} & -1 < x < 1 \\ 0 & |x| > 1. \end{cases}$$

The central value, obtained by putting $x = 0$, is $\frac{1}{3}$; hence

$$\gamma(x) = \begin{cases} 1 - \frac{3|x|}{2} + \frac{|x|^3}{2} & -1 < x < 1 \\ 0 & |x| > 1. \end{cases}$$

We note that $\gamma(0) = 1$, and as with the convolution integral, the area under the autocorrelation function may be checked (it is $\frac{1}{4}$) to verify that it is equal to the square of the area under $f(x)$.

A second example is furnished by

$$f(x) = \begin{cases} e^{-\alpha x} & x > 0 \\ 0 & x < 0. \end{cases}$$

Then

$$\int_{-\infty}^{\infty} f^*(u)f(u+x)\,du = \int_0^{\infty} e^{-\alpha u}\, e^{-\alpha(x+u)}\,du = \frac{e^{-\alpha|x|}}{2\alpha},$$

and

$$\gamma(x) = e^{-\alpha|x|}.$$

The autocorrelation function is often used in the study of functions representing observational data, especially observations exhibiting some degree of randomness, and ingenious computing machines have been devised to carry out the integration on data in various forms. In any case, the digital computation is straightforward in principle. In the theory of such phenomena, however, as distinct from their observation and analysis, one wishes to treat functions that run on indefinitely, which often means that the infinite integral does not exist. This may always be handled by considering a segment of the function of length X and replacing the values outside the range of this segment by zero; any difficulty associated with the fact that the original function ran on indefinitely

is thus removed. The autocorrelation $\gamma(x)$ of the new function $f_X(x)$, which is zero outside the finite segment, is given, according to the definition, by

$$\gamma(x) = \frac{\int_{-\infty}^{\infty} f_X^*(u)f_X(u+x)\,du}{\int_{-\infty}^{\infty} f_X(u)f_X^*(u)\,du} = \frac{\int_{-\frac{1}{2}X}^{\frac{1}{2}X} f^*(u)f(u+x)\,du}{\int_{-\frac{1}{2}X}^{\frac{1}{2}X} f_X(u)f_X^*(u)\,du}.$$

Thus the infinite integrals are reduced to finite ones, and, of course, this is precisely what happens in fact when a calculation is made on a finite quantity of observational data.

Now if we are given a function $f(x)$ to which the definition of $\gamma(x)$ does not apply because of its never dying out, and we have calculated $\gamma(x)$ for finite segments, we can then make the length X of the segment as long as we wish. It may happen that as X increases, the values of $\gamma(x)$ settle down to a limit; in fact, the circumstances under which this happens are of very wide interest. This limit, when it exists, will be denoted by $C(x)$, and it is given by

$$C(x) = \lim_{X\to\infty} \frac{\int_{-\frac{1}{2}X}^{\frac{1}{2}X} f(u)f(u+x)\,du}{\int_{-\frac{1}{2}X}^{\frac{1}{2}X} [f(u)]^2\,du}.$$

As this branch of the subject is often restricted to real functions, such as signal waveforms, allowance for complex f has been dropped.

Exercise Show that when this expression is evaluated for a real function for which $\gamma(x)$ exists, then $C(x) = \gamma(x)$.

Since the limiting autocorrelation C is identical with γ in cases where γ is applicable, it would be logical to take C as defining the autocorrelation function, and this is often done. In the discussion that follows, however, we shall understand the term "autocorrelation function" to include the operation of passing to a limit only where that is necessary.

As an example of a function that does not die out and for which the infinite integral does not exist, consider a time-varying signal which is a combination of three sinusoids of arbitrary amplitudes and phases, given by

$$V(t) = A\sin(\alpha t + \phi) + B\sin(\beta t + \chi) + C\sin(\gamma t + \psi).$$

Then

$$\int_{-\frac{1}{2}T}^{\frac{1}{2}T} V(t')V(t'+t)\,dt'$$
$$= \int_{-\frac{1}{2}T}^{\frac{1}{2}T} [A\sin(\alpha t'+\phi)A\sin(\alpha t'+\alpha t+\phi)$$
$$+ B\sin(\beta t'+\chi)B\sin(\beta t'+\beta t+\chi)$$
$$+ C\sin(\gamma t'+\psi)C\sin(\gamma t'+\gamma t+\psi)$$
$$+ \text{cross-product terms}]\,dt'$$
$$= \int_{-\frac{1}{2}T}^{\frac{1}{2}T} \{A^2[\cos\alpha t - \cos(2\alpha t'+\alpha t+2\phi)]$$
$$+ B^2[\cos\beta t - \cos(2\beta t'+\beta t+2\chi)]$$
$$+ C^2[\cos\gamma t - \cos(2\gamma t'+\gamma t+2\psi)]$$
$$+ \text{cross-product terms}\}\,dt'$$
$$= A^2T\cos\alpha t + B^2T\cos\beta t + C^2T\cos\gamma t + F(t,T) + G(t,T),$$

where F stands for oscillatory terms and G for terms arising from cross products. As T increases, F and G become negligible with respect to the three leading terms, which increase in proportion to T. Hence

$$\frac{\int_{-\frac{1}{2}T}^{\frac{1}{2}T} V(t')V(t'+t)\,dt'}{\int_{-\frac{1}{2}T}^{\frac{1}{2}T} [V(t')]^2\,dt'} = \frac{A^2T\cos\alpha t + B^2T\cos\beta t + C^2T\cos\gamma t + F(t,T) + G(t,T)}{A^2T + B^2T + C^2T + F(0,T) + G(0,T)}$$

and

$$C(t) = \lim_{T\to\infty} \frac{\int_{-\frac{1}{2}T}^{\frac{1}{2}T} V(t')V(t'+t)\,dt'}{\int_{-\frac{1}{2}T}^{\frac{1}{2}T} [V(t')]^2\,dt'}$$
$$= \frac{1}{A^2+B^2+C^2}(A^2\cos\alpha t + B^2\cos\beta t + C^2\cos\gamma t).$$

Note that the limiting autocorrelation function $C(t)$ is a superposition of three *cosine* functions at the same frequencies as contained in the signal $V(t)$, but with different relative amplitudes, and that $C(0) = 1$.

Since the principal terms in the numerator of $\gamma(x)$ increase in proportion to X, another way to obtain a nondivergent result is to divide by X before proceeding to the limit. The expression

$$\lim_{X\to\infty} \frac{1}{X}\int_{-\frac{1}{2}X}^{\frac{1}{2}X} f(u)f(u+x)\,du$$

is not equal to unity at its origin but is generally referred to simply as the "autocorrelation function" because it becomes the same as $C(x)$ if divided by its central value (provided the central value is not zero). It may therefore be regarded as a nonnormalized form of $C(x)$.

When time is the independent variable it is customary to refer to the time average of the product $V(t)V(t+\tau)$, and write

$$\langle V(t)V(t+\tau)\rangle \equiv \lim_{T\to\infty}\frac{1}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} V(t)V(t+\tau)\,dt,$$

where the operation of time averaging is denoted by angular brackets according to the definition

$$\langle\ .\ .\ .\ \rangle \equiv \lim_{T\to\infty}\frac{1}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} \ .\ .\ .\ dt.$$

Using this notation for the example worked, we have

$$\langle V(t)V(t+\tau)\rangle = A^2\cos\alpha\tau + B^2\cos\beta\tau + C^2\cos\gamma\tau,$$

which, of course, is a little simpler than the normalized expression. It should be noticed, however, that we do not get a useful result in cases for which γ exists.

A conspicuous feature of C in this example is the absence of any trace of the original phases. Hence, autocorrelation is not a reversible process; it is not possible to get back from the autocorrelation function to the original function from which it was derived. Autocorrelation thus involves a loss of information. In some branches of physics, such as radio interferometry and X-ray diffraction analysis of substances, it is easier to observe the autocorrelation of a desired function than to observe the function itself, and a lot of ingenuity is expended to fill in the lost information.

The character of the lost information can be seen by considering a cosinusoidal function of x. When this function is displaced relative to itself, multiplied with the unshifted function, and the result integrated, clearly the result will be the same as for a sinusoidal function of the same period and amplitude. Furthermore, the result will be the same for any harmonic function of x, with the same period and amplitude, and arbitrary origin of x. Thus the autocorrelation function does not reveal the phase of a harmonic function. Now, if a function is composed of several harmonic waves present simultaneously, then when it is displaced, multiplied, and integrated, the result can be calculated simply by considering the different periods one at a time. This is possible because the product of harmonic variations of different frequencies integrates to a negligible quantity. Consequently, each of the periodic waves may be slid along the x axis into any arbitrary phase, without affecting the autocorrelation. In particular, all the components may be shifted until they become cosine components, thus generating an even function which, among all possible functions with the same autocorrelation, possesses a certain uniqueness.

Pentagram notation for cross correlation

The cross correlation $f(x)$ of two real functions $g(x)$ and $h(x)$ is defined as

$$f(x) = g \star h = \int_{-\infty}^{\infty} g(u - x)h(u)\, du.$$

It is thus similar to the convolution integral except that the component $g(u)$ is simply displaced to $g(u - x)$, without reversal. To denote this operation of g on h we use a five-pointed star, or pentagram, instead of the asterisk that denotes convolution. While cross correlation is slightly simpler than convolution, its properties are less simple. Thus, whereas $g * h = h * g$, the cross-correlation operation of h on g is not the same as that of g on h; that is,

$$g \star h \neq h \star g.$$

By change of variables it can be seen that

$$f(x) = g \star h = \int_{-\infty}^{\infty} g(u - x)h(u)\, du = \int_{-\infty}^{\infty} g(u)h(u + x)\, du$$
$$f(-x) = h \star g = \int_{-\infty}^{\infty} h(u - x)g(u)\, du = \int_{-\infty}^{\infty} h(u)g(u + x)\, du;$$
$$g \star h = g(-) * h.$$

As in the case of the autocorrelation function, the cross correlation is often normalized so as to be equal to unity at the origin, and when appropriate the average $\langle g(u - x)h(u)\rangle$ is used instead of the infinite integral. When the functions are complex it is customary to define the (complex) cross-correlation function by

$$g^* \star h = \int_{-\infty}^{\infty} g^*(u - x)h(u)\, du = \int_{-\infty}^{\infty} g^*(u)h(u + x)\, du.$$

The energy spectrum

We shall refer to the squared modulus of a transform as the energy spectrum; that is, $|F(s)|^2$ is the energy spectrum of $f(x)$. The term is taken directly from the physical fields where it is used. It will be seen that there is not a one-to-one relationship between $f(x)$ and its energy spectrum, for although $f(x)$ determines $F(s)$ and hence also $|F(s)|^2$, it would be necessary to have[2] pha $F(s)$ as well as $|F(s)|$ in order to reconstitute $f(x)$.

[2] The phase angle of the complex quantity $F(s)$ is written pha $F(s)$ and is defined as follows: if there is a complex variable $z = r \exp i\theta$, then pha $z = \theta$.

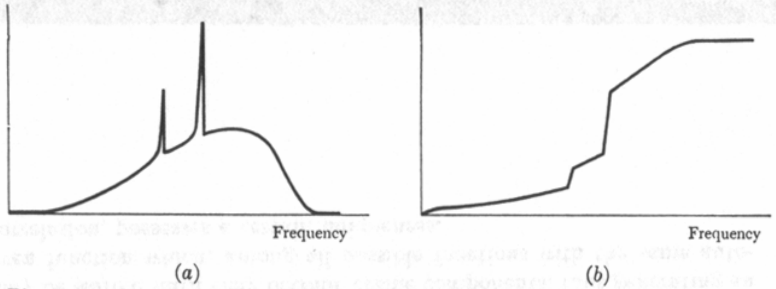


Fig. 3.12 Spectra of X radiation from molybdenum: (a) power spectrum; (b) cumulative spectrum.

Knowledge of the energy spectrum thus conveys a certain kind of information about $f(x)$ which says nothing about the phase of its Fourier components. It is the kind of information about an acoustical waveform which results from measuring the sound intensity as a function of frequency.

The information lost when only the energy spectrum can be given is of precisely the same character as that which is lost when the autocorrelation function has to do duty for the original function. The autocorrelation theorem, to be given later, expresses this equivalence.

When $f(x)$ is real, as it would be if it represented a physical waveform, the energy spectrum is an even function and is therefore fully determined by its values for $s \geqslant 0$. To stress this fact, the term "positive-frequency energy spectrum" may be used to mean the part of $|F(s)|^2$ for which $s \geqslant 0$.

Since $|F(s)|^2$ has the character of energy density measured per unit of s, it would have to become infinite if a nonzero amount of energy were associated with a discrete value of s. This is the situation with an infinitely narrow spectral line. Now we can consider a cumulative distribution function which gives the amount of energy in the range 0 to s:

$$\int_0^s |F(s)|^2\, ds.$$

Any spectral *lines* would then appear as finite discontinuities in the cumulative energy spectrum as suggested by Fig. 3.12, and some mathematical convenience would be gained by using the cumulative spectrum in conjunction with Stieltjes integral notation. The convenience is especially marked when it is a question of using the theory of distributions for some question of rigor. However, the matter is purely one of notation, and in cases where we have to represent concentrations of energy within bands much narrower than can be resolved in the given context, we shall use the delta-symbol notation described later.

Appendix

Prove that the autocorrelation of the real function $f(x)$ is a maximum at the origin, that is,

$$\int_{-\infty}^{\infty} f(u)f(u+x)\,du \leqslant \int_{-\infty}^{\infty} [f(u)]^2\,du.$$

Let ϵ be a real number. Then

$$\int_{-\infty}^{\infty} [f(u) + \epsilon f(u+x)]^2\,du > 0$$

and $$\int_{-\infty}^{\infty} [f(u)]^2\,du + 2\epsilon \int_{-\infty}^{\infty} f(u)f(u+x)\,du + \epsilon^2 \int_{-\infty}^{\infty} [f(u+x)]^2\,du > 0;$$

that is, $$a\epsilon^2 + b\epsilon + c > 0,$$

where $$a = c = \int_{-\infty}^{\infty} [f(u)]^2\,du$$

$$b = 2\int_{-\infty}^{\infty} f(u)f(u+x)\,du.$$

Now, if the quadratic expression in ϵ may not be zero, that is, if it has no real root, then

$$b^2 - 4ac \leqslant 0.$$

Hence in this case $b/2 \leqslant a$, or

$$\frac{\int_{-\infty}^{\infty} f(u)f(u+x)\,du}{\int_{-\infty}^{\infty} [f(u)]^2\,du} \leqslant 1.$$

The equality is achieved at $x = 0$; consequently the autocorrelation function can nowhere exceed its value at the origin. The argument is the one used to establish the Schwarz inequality and readily generalizes to give the similar result for the complex autocorrelation function.

Exercise Extend the argument with a view to showing that the equality cannot be achieved for any value of x save zero.

Problems

1 Calculate the following serial products, checking the results by summation. Draw graphs to illustrate.

a. {6 9 17 20 10 1} * {3 8 11}
b. {1 1 1 1 1} * {1 1 1 1}
c. {1 4 2 3 5 3 3 4 5 7 6 9} * {1 1}
d. {1 4 2 3 5 3 3 4 5 7 6 9} * {$\frac{1}{2}$ $\frac{1}{2}$}

e. {1 4 2 3 5 3 3 4 5 7 6 9} * {1 2 1}
f. {1 4 2 3 5 3 3 4 5 7 6 9} * {$\frac{1}{4}$ $\frac{1}{2}$ $\frac{1}{4}$}
g. {1 2 3 7 12 19 21 22 18 13 7 5 3 2 1} * {1 −1}
h. {1 2 3 7 12 19 21 22 21 18 13 7 5 3 2 1} * {1 1 1 1 . . . }
i. {1 1} * {1 1} * {1 1}
j. {1 1 0 0 1 0 1} * {1 1 0 0 1 0 1}
k. {1 1 0 0 1 0 1} * {1 0 1 0 0 1 1}
l. {*a b c d e*} * {*e d c b a*}
m. {1 3 1} * {1 0 0 0 0 0 ±1}
n. {1 3 1} * {1 2 2}
o. Multiply 131 by 122
p. Multiply 10,301 by 10,202
q. {1 8 1} * {1 2 2}
r. Comment on the smoothness of your results in *d* and *f* relative to the longer of the two given sequences.
s. Consider the result of *i* in conjunction with Pascal's pyramid of binomial coefficients.
t. Seek longer sequences with the same property you discovered in *k*.
u. Contemplate *j*, *k*, *l*, *m*, and *n* with a view to discerning what leads to serial products which are even.
v. Master the implication of *n*, *o*, *p*, and *q*, and design a mechanical desk computer to perform serial multiplication.

2 Derive the following results, where $H(x)$ is the Heaviside unit step function (Chapter 4):

$$\begin{aligned} x^2H(x) * e^xH(x) &= (2e^x - x^2 - 2x - 2)H(x) \\ [\sin x\, H(x)]^{*2} &= \tfrac{1}{2}(\sin x - x\cos x)H(x) \\ [(1-x)H(x)] * [e^xH(x)] &= xH(x) \\ H(x) * [e^xH(x)] &= (e^x - 1)H(x) \\ [e^xH(x)]^{*2} &= xe^xH(x) \\ [e^xH(x)]^{*3} &= \tfrac{1}{2}x^2e^xH(x) \end{aligned}$$

3 Prove the commutative property of convolution, that is, that $f * g \equiv g * f$.

4 Prove the associative rule $f * (g * h) \equiv (f * g) * h$.

5 Prove the distributive rule for addition $f * (g + h) \equiv f * g + f * h$.

6 The function f is the convolution of g and h. Show that the self-convolutions of f, g, and h are related in the same way as the original functions.

7 If $f = g \star h$, show that $f \star f = (g \star g) \star (h \star h)$.

8 Show that if a is a constant, $a(f * g) = (af) * g = f * (ag)$.

9 Establish a theorem involving $f(g * h)$.

10 Prove that the autocorrelation function is hermitian, that is, that $C(-u) = C^*(u)$, and hence that when the autocorrelation function is real it is even. Note that if the autocorrelation function is imaginary it is also odd; give some thought to devising a function with an odd autocorrelation function.

11 Prove that the sum and product of two autocorrelation functions are each hermitian.

12 Alter the origin of $f(x)$ until $f * f\big|_0$ is a maximum. Investigate the assertion that the new origin defines an axis of maximum symmetry, making any necessary modification. Investigate the merits of the parameter

$$\frac{f * f\big|_0}{f \star f\big|_0},$$

to be considered a measure of "degree of evenness."

13 Show that if $f(x)$ is real,

$$\int_{-\infty}^{\infty} f(x)f(-x)\,dx = \int_{-\infty}^{\infty} [E(x)]^2\,dx - \int_{-\infty}^{\infty} [O(x)]^2\,dx,$$

and note that the left-hand side is the central value of the self-convolution of $f(x)$; that is, $f * f\big|_0$.

14 Find reciprocal sequences for {1 3 3 1} and {1 4 6 4 1}.

15 Find reciprocal sequences for {1 1} and {1 1 1}.

16 Establish a general procedure for finding the reciprocal of finite or semi-infinite sequences and test it on the following cases:

$$\{64\ 32\ 16\ 8\ 4\ 2\ 1\ \ldots\}$$
$$\{64\ 64\ 48\ 32\ 20\ 12\ 7\ 4\ \ldots\}$$
$$\left\{1\quad e^{-\frac{1}{10}}\cos\left(\frac{\pi}{10}\right)\quad e^{-\frac{2}{10}}\cos\left(\frac{2\pi}{10}\right)\quad \ldots\quad e^{-n/10}\cos\left(\frac{n\pi}{10}\right)\quad \ldots\right\}$$

17 Find approximate numerical values for a function $f(x)$ such that

$$f(x) * E(x)$$

is zero when evaluated numerically by serial multiplication of values taken at intervals of 0.2 in x, except at the origin. Normalize $f(x)$ so that its integral is approximately unity.

18 The cross correlation $g \star h$ is to be normalized to unity at its maximum value. It is argued that

$$0 \leqslant \int [g(u) - h(u+x)]^2\,du = \int g^2\,du - 2\int g(u)h(u+x)\,du + \int h^2\,du,$$

and therefore that

$$\int g(u)h(u+x)\,du \leqslant \tfrac{1}{2}\int g^2\,du + \tfrac{1}{2}\int h^2\,du = M.$$

Consequently $(g \star h)/M$ is the desired quantity. Correct the fallacy in this argument.

Chapter 4 Notation for some useful functions

Many useful functions in Fourier analysis have to be defined piecewise because of abrupt changes. For example, we may consider the function $f(x)$ such that

$$f(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leqslant x \leqslant 1 \\ 1 & x > 1. \end{cases}$$

This function, though simple in itself, is awkwardly expressed in comparison with a function such as, for example, $1 + x^2$, whose algebraic expression compactly states, over the infinite range of x, the arithmetical operations by which it is formed. For many mathematical purposes a function which is piecewise analytic is not simple to deal with, but for physical purposes a "sloping step function," to give it a name, may be at least as simple as a smoother function.

Fourier himself was concerned with the representation of functions given graphically, and according to E. W. Hobson "was the first fully to grasp the idea that a single function may consist of detached portions given arbitrarily by a graph."

To regain compactness and clarity of notation, we introduce a number of simple functions embodying various kinds of abrupt behavior. Also included here is a section dealing with sinc x, the important interpolating function, which is the transform of a discontinuous function, and some reference material on notations for the Gaussian function.

Rectangle function of unit height and base, $\Pi(x)$

The rectangle function of unit height and base, which is illustrated in Fig. 4.1, is defined by

$$\Pi(x) = \begin{cases} 0 & |x| > \frac{1}{2} \\ (\frac{1}{2} & |x| = \frac{1}{2}) \\ 1 & |x| < \frac{1}{2}. \end{cases}$$

It provides simple notation for segments of functions which have simple expressions, for example, $f(x) = \Pi(x) \cos \pi x$ is compact notation for

$$f(x) = \begin{cases} 0 & x < -\frac{1}{2} \\ \cos \pi x & -\frac{1}{2} < x < \frac{1}{2} \\ 0 & \frac{1}{2} < x \end{cases}$$

(see Fig. 4.2). We may note that $h\Pi[(x - c)/b]$ is a displaced rectangle function of height h and base b, centered on $x = c$ (see Fig. 4.3). Hence, purely by multiplication by a suitably displaced rectangle function, we

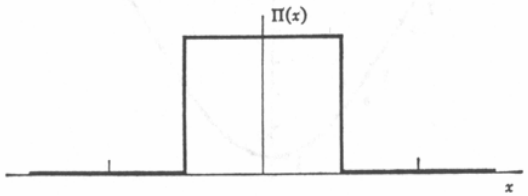


Fig. 4.1 The rectangle function of unit height and base, $\Pi(x)$.

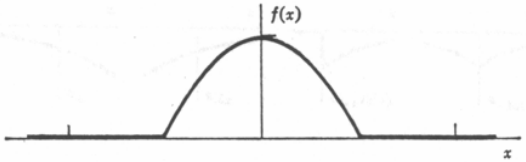


Fig. 4.2 A segmented function expressed by $\Pi(x) \cos \pi x$.

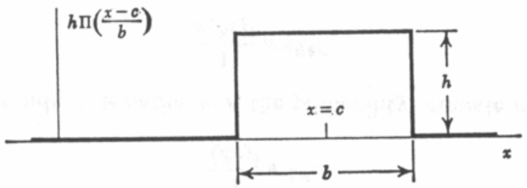


Fig. 4.3 A displaced rectangle function of arbitrary height and base expressed in terms of $\Pi(x)$.

can select, or gate, any segment of a given function, with any amplitude, and reduce the rest to zero.

The rectangle function also enters through convolution into expressions for running means, and, of course, in the frequency domain multiplication by the rectangle function is an expression of ideal low-pass filtering. It is important in the theory of convergence of Fourier series, where it is generally known as Dirichlet's discontinuous factor. The notation rect x, which has been suggested for this function, does not lend itself readily to notation such as $\overline{\Pi}$, $\Pi * \Pi$, and Πf.

For reasons explained below, it is almost never important to specify the values at $x = \pm\frac{1}{2}$, that is, at the points of discontinuity, and we shall normally omit mention of those values. Likewise, it is not necessary or desirable to emphasize the values $\Pi(\pm\frac{1}{2}) = \frac{1}{2}$ in graphs; it is preferable to show graphs which are reminiscent of high-quality oscillograms (which, of course, would never show extra brightening halfway up the discontinuity).

The triangle function of unit height and area, $\Lambda(x)$

By definition,

$$\Lambda(x) = \begin{cases} 0 & |x| > 1 \\ 1 - |x| & |x| < 1. \end{cases}$$

This function, which is illustrated in Fig. 4.4, gains its importance largely from being the self-convolution of $\Pi(x)$, but it has other uses—for example, in giving compact notation for polygonal functions (continuous functions consisting of linear segments).

Note that $h\Lambda(x/\frac{1}{2}b)$ is a triangle function of height h, base b, and area $\frac{1}{2}hb$.

Various exponentials and Gaussian and Rayleigh curves

Figure 4.5 shows, from left to right, a rising exponential, a falling exponential, a truncated falling exponential, and a double-sided falling exponential.

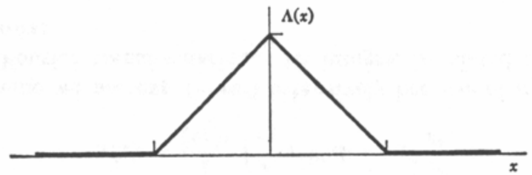


Fig. 4.4 The triangle function of unit height and area, $\Lambda(x)$.

Figure 4.6 shows the Gaussian function $\exp(-\pi x^2)$. Of the various ways of normalizing this function, we have chosen one in which both the central ordinate and the area under the curve are unity, and certain advantages follow from this choice. The Fourier transform of the Gaussian function is also Gaussian, and proves to be normalized in precisely the same way under our choice. In statistics the Gaussian distribution is referred to as the "normal (error) distribution with zero mean" and is normalized so that the area and the standard deviation are unity. We may use the term "probability ordinate" to distinguish the form

$$\frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}x^2}.$$

When the standard deviation is σ, the probability ordinate is

$$\frac{1}{\sigma(2\pi)^{\frac{1}{2}}} e^{-x^2/2\sigma^2},$$

and the area under the curve remains unity. The central ordinate is equal to $0.3989/\sigma$. Prior to the strict standardization now prevailing

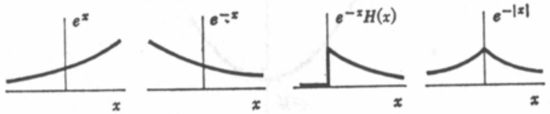


Fig. 4.5 Various exponential functions.

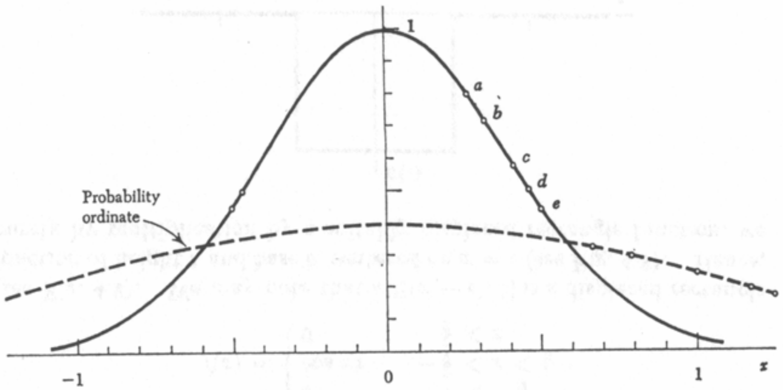


Fig. 4.6 The Gaussian function exp $(-\pi x^2)$ *and the probability ordinate* $(2\pi)^{-\frac{1}{2}}$ *exp* $(-\frac{1}{2}x^2)$.

in statistics, the error integral erf x was introduced as

$$\operatorname{erf} x = \frac{2}{\pi^{\frac{1}{2}}} \int_0^x e^{-t^2}\,dt.$$

The complementary error integral erfc x is defined by

$$\operatorname{erfc} x = 1 - \operatorname{erf} x.$$

The probability integral $\alpha(x)$, now widely tabulated, is

$$\alpha(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-x}^{x} e^{-\frac{1}{2}t^2}\,dt = \operatorname{erf} \frac{x}{2^{\frac{1}{2}}}.$$

In this volume we use $\exp(-\pi x^2)$ extensively because of its symmetry under the Fourier transformation. Its integral is related to erf x and $\alpha(x)$ as follows:

$$\int_0^x e^{-\pi t^2}\,dt = \tfrac{1}{2} \operatorname{erf} \pi^{\frac{1}{2}}x = \tfrac{1}{2}\alpha[(2\pi)^{\frac{1}{2}}x]$$
$$\int_{-\infty}^{x} e^{-\pi t^2}\,dt = \tfrac{1}{2} + \tfrac{1}{2} \operatorname{erf} \pi^{\frac{1}{2}}x = \tfrac{1}{2} + \tfrac{1}{2}\alpha[(2\pi)^{\frac{1}{2}}x].$$

The customary dispersion parameters of $\exp(-\pi x^2)$ are (see Fig. 4.6) as follows:

a. Probable error		$= 0.2691$	$= 0.6745\sigma$
b. Mean absolute error (mean of $\lvert x\rvert$)	$= \pi^{-1}$	$= 0.3183$	$= 0.7979\sigma$
c. Standard deviation (mean of x^2)	$= (2\pi)^{-\frac{1}{2}}$	$= 0.3989$	$= \sigma$
d. Width to half-peak		$= 0.9394$	$= 2.355\sigma$
e. Equivalent width		$= 1.0000$	$= 2.5066\sigma$

In two dimensions the Gaussian distribution generalizes to

$$e^{-\pi(x^2+y^2)},$$

again with symmetry under the Fourier transformation, with unit central ordinate, and with unit volume. The version used in statistics, for arbitrary standard deviations σ_x and σ_y, is

$$\frac{1}{2\pi\sigma_x\sigma_y} e^{-x^2/2\sigma_x^2 - y^2/2\sigma_y^2}.$$

Under conditions of circular symmetry, and putting $x^2 + y^2 = r^2$, the two-dimensional probability ordinate becomes

$$\frac{1}{2\pi\sigma^2} e^{-r^2/2\sigma^2}.$$

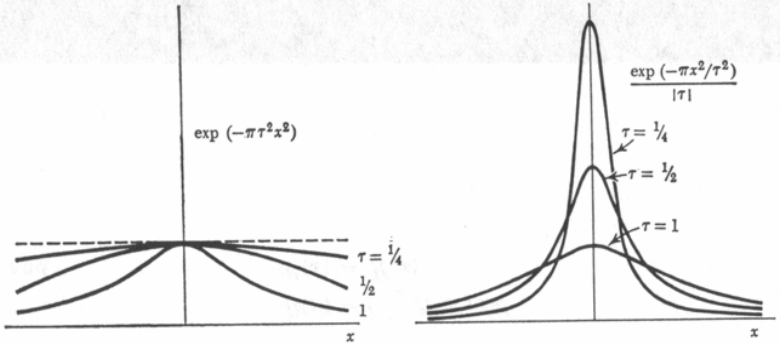


Fig. 4.7 Standard sequences of Gaussian functions.

The probability $R(r)\,dr$ of finding the radial distance in the range r to $r + dr$ is $2\pi r\,dr$ times the above expression; hence

$$R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}.$$

This is referred to as Rayleigh's distribution, since it occurred in the famous problem of the drunkard's walk discussed by Rayleigh. In a simple version of the problem the drunkard always falls down after taking one step, and the direction of each step bears no relation to the previous step. After a long time has elapsed, the probability of finding him at (x,y) is a two-dimensional Gaussian function (according to which he is more likely to be at the origin than elsewhere), and the probability of finding him at a distance r from the origin is given by a Rayleigh distribution. Since the Rayleigh distribution has a peak that is not at the origin, the above statements may appear contradictory. If they do, the reader will find it instructive to contemplate the matter further.

The following are some infinite integrals often needed for checking:

$$\int_{-\infty}^{\infty} e^{-\pi x^2}\,dx = 1 \qquad \int_{-\infty}^{\infty} e^{-x^2}\,dx = \pi^{\frac{1}{2}} \qquad \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} = (2\pi)^{\frac{1}{2}}$$

$$\int_{-\infty}^{\infty} e^{-Ax^2}\,dx = \left(\frac{\pi}{A}\right)^{\frac{1}{2}} \qquad \int_0^{\infty} xe^{-x^2}\,dx = \tfrac{1}{2} \qquad \int_0^{\infty} x^2e^{-x^2}\,dx = \frac{\pi^{\frac{1}{2}}}{4}.$$

Sequences of Gaussian functions play a special role in connection with transforms in the limit. The sequence $\exp(-\pi\tau^2x^2)$ as τ approaches zero is useful for multiplying with functions whose integrals do not converge. The limiting member of the sequence is unity. The sequence $|\tau|^{-1}\exp(-\pi x^2/\tau^2)$ is used for recovering ordinary functions, in cases of impulsive behavior, by convolution. The properties which make the Gaussian function useful in these contexts are that its derivatives are all continuous

and that it dies away more rapidly than any power of x; that is,

$$\lim_{x\to\infty} x^n e^{-x^2} = 0$$

for all n.

Figure 4.7 shows these two important sequences; later we emphasize that corresponding members are Fourier transform pairs.

Heaviside's unit step function, H(x)

An indispensable aid in the representation of simple discontinuities, the unit step function is defined by

$$H(x) = \begin{cases} 0 & x < 0 \\ (\frac{1}{2} & x = 0) \\ 1 & x > 0, \end{cases}$$

and is illustrated in Fig. 4.8. It represents voltages which are suddenly switched on or forces which begin to act at a definite time and are constant thereafter. Furthermore, any function with a jump can be decomposed into a continuous function plus a step function suitably displaced. As a simple example of additive use, consider the rectangle function $\Pi(x)$, which has two unit discontinuities, one positive and one negative. If these are removed, nothing remains. Hence $\Pi(x)$ is expressible entirely in terms of step functions as follows (see Fig. 4.9):

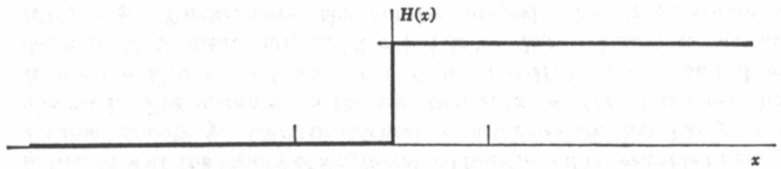


Fig. 4.8 The unit step function.

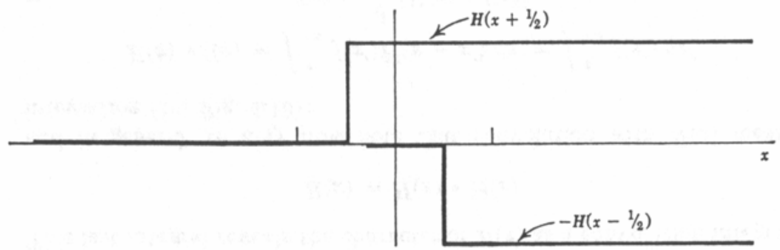


Fig. 4.9 Two functions whose sum is $\Pi(x)$.

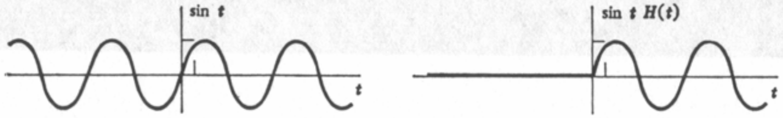


Fig. 4.10 Graphs of $\sin t$ *and* $\sin t\, H(t)$ *which exhibit a principal use of the unit step function.*

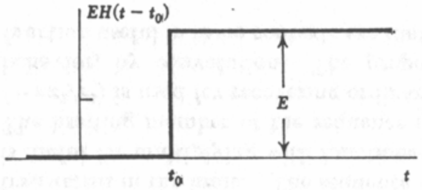


Fig. 4.11 A voltage E which appears at $t = t_0$ *represented in step-function notation by* $EH(t - t_0)$.

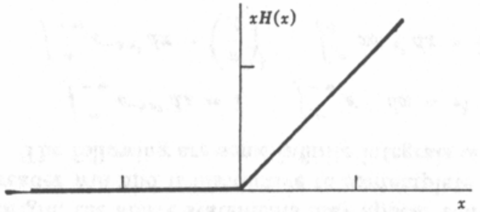


Fig. 4.12 The ramp function $xH(x)$.

$$\Pi(x) = H(x + \tfrac{1}{2}) - H(x - \tfrac{1}{2}).$$

Since multiplication of a given function by $H(x)$ reduces it to zero where x is negative but leaves it intact where x is positive, the unit step function provides a convenient way of representing the switching on of simply expressible quantities. For instance, we can represent a sinusoidal quantity which switches on at $t = 0$ by $\sin t\, H(t)$ (see Fig. 4.10). A voltage which has been zero until $t = t_0$ and then jumps to a steady value E is represented by $EH(t - t_0)$ (see Fig. 4.11).

The ramp function $R(x) = xH(x)$ furnishes a further example of notation involving $H(x)$ multiplicatively (see Fig. 4.12). $(F/m)R(t)$ represents the velocity of a mass m to which a steady force $FH(t)$ has been applied, or the current in a coil of inductance m across which the potential difference is $FH(t)$. It will be seen that $R(x)$ is also the integral of $H(x)$ and, conversely, that $H(x)$ is the derivative of $R(x)$. Thus

$$R(x) = \int_{-\infty}^{x} H(x')\, dx'$$

and

$$R'(x) = H(x).$$

Step-function notation plays a role in simplifying integrals with variable limits of integration by reducing the integrand to zero in the range beyond the original limits. Then constant limits such as $-\infty$ to ∞ or 0 to ∞ can be used. For example, the function $H(x - x')$ is zero where $x' > x$, and therefore $\int_{-\infty}^{x} f(x')\, dx'$ can always be written $\int_{-\infty}^{\infty} f(x')H(x - x')\, dx'$. Thus in the example appearing above we can write

$$R(x) = \int_{-\infty}^{x} H(x')\, dx' = \int_{-\infty}^{\infty} H(x')H(x - x')\, dx'.$$

This last integral reveals the character of $R(x)$ as a convolution integral,

$$R(x) = H(x) * H(x),$$

and in general we may now note that convolution with $H(x)$ means integration (see Fig. 4.13):

$$H(x) * f(x) = \int_{-\infty}^{\infty} f(x')H(x - x')\, dx' = \int_{-\infty}^{x} f(x')\, dx'$$

or

$$f(x) = \frac{d}{dx}[H(x) * f(x)].$$

Usually, it is not important to define $H(0)$, but for the sake of compatibility with the theory of single-valued functions it is desirable to assign a value, usually $\frac{1}{2}$. Certain internal consistencies are then likely to be observed. For instance, in the equation $R'(x) = H(x)$ it is clear that $R'(0+) = H(0+) = 1$ and that $R'(0-) = H(0-) = 0$, and if we deem $R'(0)$ to mean $\lim_{\Delta x \to 0} [R(x + \frac{1}{2}\Delta x) - R(x - \frac{1}{2}\Delta x)]/\Delta x$, we find $R'(0) = \frac{1}{2}$. Furthermore, the Fourier integral, when it converges at a point of discontinuity, gives the midvalue.

However, there is no obligation to take $H(0) = \frac{1}{2}$; it is not uncommonly taken as zero. This is a natural consequence of a point of view according to which (see Fig. 4.14)

$$\hat{H}(x) = \lim_{\tau \to 0} [(1 - e^{-x/\tau})H(x)].$$

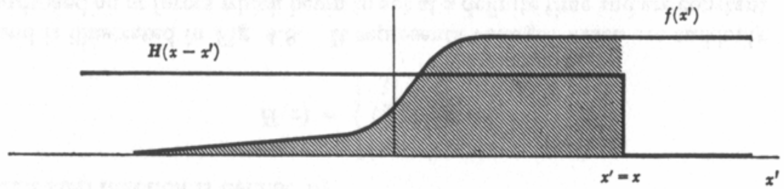


Fig. 4.13 The shaded area is $\int_{-\infty}^{x} f(x')\, dx'$, *or a value of the convolution of* $f(x)$ *with* $H(x)$.

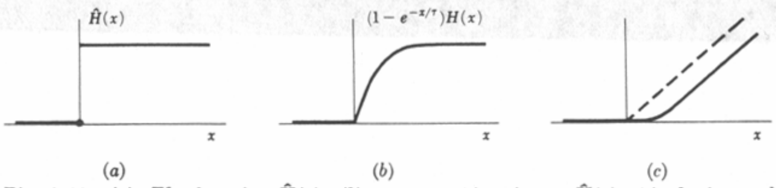


Fig. 4.14 (a) *The function* $\hat{H}(x)$; (b) *an approximation to* $\hat{H}(x)$; (c) *the integral of* (b).

Then $\hat{H}(0) = 0$. The circumflex indicates that there is a slight difference from the definition already decided on.

Thus
$$\hat{H}(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leqslant 0 \end{cases}$$
and
$$H(x) = \hat{H}(x) + \tfrac{1}{2}\delta^0(x)$$
where
$$\delta^0(x) = \begin{cases} 0 & x \neq 0 \\ 1 & x = 0. \end{cases}$$

The function $\delta^0(x)$ is one of a class of *null functions* referred to below. No discrepancy need arise from the relation $R'(x) = \hat{H}(x)$, since the integral of $\hat{H}(x)$, regarded as

$$\lim_{\tau\to 0}\int_{-\infty}^{x}(1-e^{-u/\tau})H(u)\,du,$$

is certainly $R(x)$; and $R'(0)$, regarded as the limiting slope at $x = 0$ of approximations such as those shown in Fig. 4.14*c*, is certainly equal to $\hat{H}(0)$, namely, 0.

It is sometimes useful to have continuous approximations to $H(x)$. The following examples all approach $H(x)$ as a limit for all x as $\tau \to 0$.

$$\tfrac{1}{2} + \frac{1}{\pi}\arctan\frac{x}{\tau}$$
$$\tfrac{1}{2}\operatorname{erfc}\left(-\frac{x}{\tau}\right) = \frac{1}{\pi^{\frac{1}{2}}}\int_{-x}^{\infty}\tau^{-1}e^{-u^2/\tau^2}\,du$$
$$\tfrac{1}{2} + \frac{1}{\pi}\operatorname{Si}\frac{\pi x}{\tau} = \int_{-\infty}^{x}\tau^{-1}\operatorname{sinc}\frac{u}{\tau}\,du$$
$$\int_{-\infty}^{x}\tau^{-1}\Pi\left(\frac{u}{\tau}\right)du$$
$$\tfrac{1}{2} + \begin{cases} \tfrac{1}{2}(1-e^{-x/\tau}) & x > 0. \\ -\tfrac{1}{2}(1-e^{x/\tau}) & x < 0. \end{cases}$$

An example which approaches $H(x)$ as $\tau \to 0$ for all x except $x = 0$ is

$$f(x,\tau) = \begin{cases} 0 & x < -\tau \\ 1 - e^{-(x+\tau)/\tau} & x > -\tau. \end{cases}$$

In this case
$$\lim_{\tau\to 0} f(x,\tau) = H(x) + (\tfrac{1}{2} - e^{-1})\,\delta^0(x),$$
since $f(0,\tau) = 1 - e^{-1}$ for all τ. A further example which approaches $H(x)$ as $\tau \to 0$ is

$$\int_{-\infty}^{x}\tau^{-1}\Lambda\left(\frac{u-\frac{1}{2}\tau}{\tau}\right)du.$$

The difference between $H(x)$ and any version such that $H(0) \neq \frac{1}{2}$ is a null function whose integral is always zero. If it were necessary to make physical observations of a quantity varying as $H(x)$ or $\hat{H}(x)$, with the finite resolving power to which physical observations are limited, it would not be possible to distinguish between the mathematically distinct alternatives, since the weighted means over nonzero intervals, which are the only quantities measurable, would be unaffected by the presence or absence of null functions. For physical applications of $H(x)$ it is therefore perhaps more graceful not to mention $H(0)$.

The sign function, sgn x

The function sgn x (pronounced signum x) is equal to $+1$ or -1, according to the sign of x (see Fig. 4.15). Thus

$$\operatorname{sgn} x = \begin{cases} -1 & x < 0 \\ 1 & x > 0. \end{cases}$$

Clearly it differs little from the step function $H(x)$ and has most of its properties. It has a positive discontinuity of 2. The relation to $H(x)$ is

$$\operatorname{sgn} x = 2H(x) - 1.$$

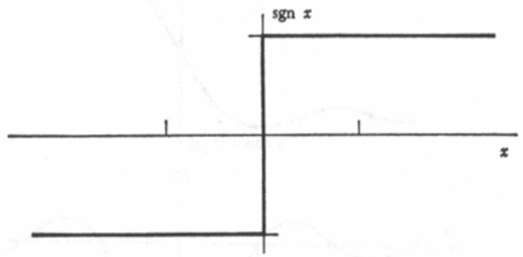


Fig. 4.15 *The odd function sgn x.*

However, $\lim_{A\to\infty}\int_{-A}^{A}\operatorname{sgn} x\,dx = 0$, whereas $\int_{-\infty}^{\infty} H(x)$ does not exist. Furthermore, we may note that sgn x, unlike $H(x)$, is an odd function, and this property sometimes makes it more useful in symbolic expressions.

The filtering or interpolating function, sinc x

We define

$$\operatorname{sinc} x = \frac{\sin \pi x}{\pi x},$$

a function with the properties that

$$\operatorname{sinc} 0 = 1$$

$$\operatorname{sinc} n = 0 \qquad n = \text{nonzero integer}$$

$$\int_{-\infty}^{\infty} \operatorname{sinc} x\,dx = 1.$$

This function will appear so frequently, in many different connections, that it is convenient to have a special symbol for it, especially in some agreed normalized form. In the form chosen here, the central ordinate is unity and the total area under the curve is unity (see Fig. 4.16*a*); the word "sinc" appears in Woodward's book[1] and has achieved some currency. A table of the sinc function appears on page 368.

The unique properties of sinc x go back to its spectral character: it contains components of all frequencies up to a certain limit and none beyond. Furthermore, the spectrum is flat up to the cutoff frequency. By our choice of notation, sinc x and $\Pi(s)$ are a Fourier transform pair; the cutoff frequency of sinc x is thus 0.5 (cycles per unit of x).

When sinc x enters into convolution it performs ideal low-pass filtering; that is, it removes all components above its cutoff and leaves all below unaltered, and under certain special circumstances discussed later under the sampling theorem, it performs an important kind of interpolation.

In terms of the widely tabulated sine integral Si x (shown in Fig. 4.16*c*), where

$$\operatorname{Si} x = \int_0^x \frac{\sin u}{u}\,du,$$

we have the relations

$$\int_0^x \operatorname{sinc} u\,du = \frac{\operatorname{Si}(\pi x)}{\pi}$$

$$\operatorname{sinc} x = \frac{d}{dx}\frac{\operatorname{Si}(\pi x)}{\pi},$$

[1] P. M. Woodward, "Probability and Information Theory with Applications to Radar," McGraw-Hill Book Company, New York, 1953.

and for the integral of sinc x (see Fig. 4.16*b*) we have

$$H(x) * \operatorname{sinc} x = \int_{-\infty}^{x} \operatorname{sinc} u\,du = \frac{1}{2} + \frac{\operatorname{Si}(\pi x)}{\pi}.$$

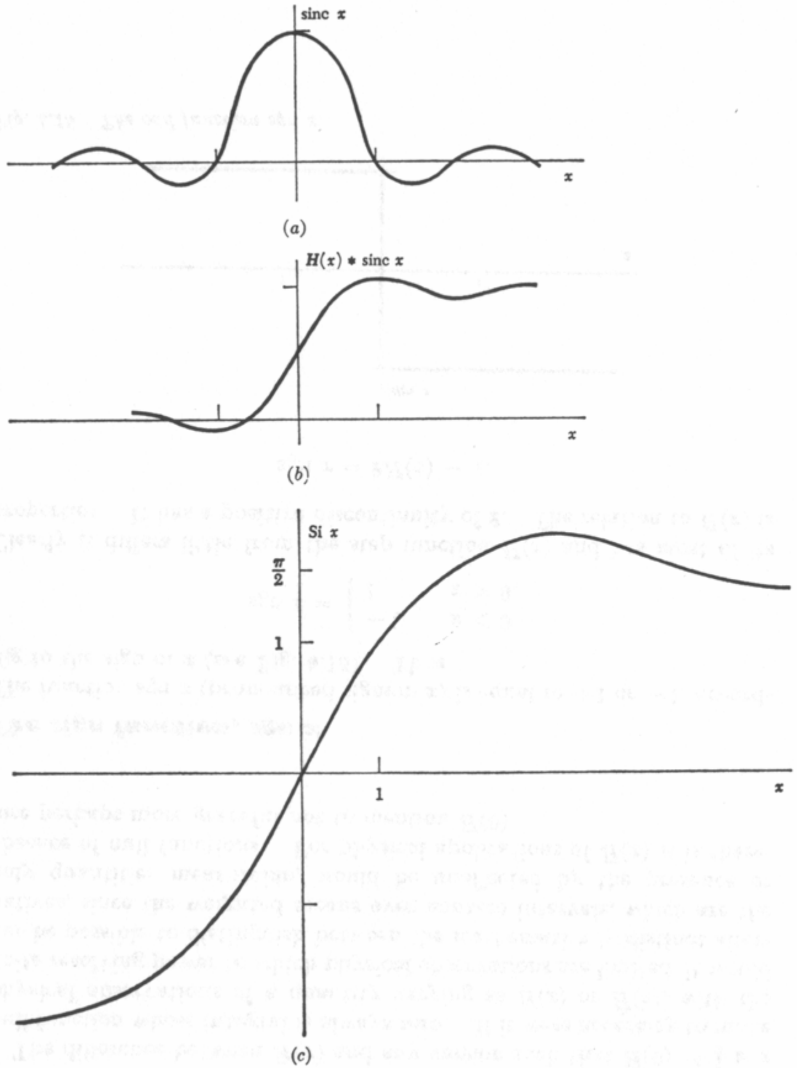


Fig. 4.16 (a) The filtering or interpolating function sinc x; (b) its integral; (c) the sine integral Si x.

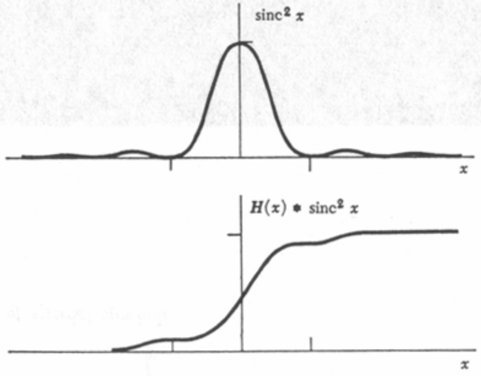


Fig. 4.17 The square of sinc x and its integral.

Another frequently needed function, tabulated on page 368, is the square of sinc x (see Fig. 4.17):

$$\operatorname{sinc}^2 x = \left(\frac{\sin \pi x}{\pi x}\right)^2.$$

This function represents the power radiation pattern of a uniformly excited antenna, or the intensity of light in the Fraunhofer diffraction pattern of a slit. Naturally, it shares with sinc x the property of having a cutoff spectrum, since squaring cannot generate frequencies higher than the sum-frequency of any pair of sinusoidal constituents. A little quantitative thought along the lines of this appeal to physical principles would soon reveal that the Fourier transform of $\operatorname{sinc}^2 x$ is $\Lambda(s)$. The cutoff frequency is one cycle per unit of x.

Among the properties of $\operatorname{sinc}^2 x$ are the following:

$$\operatorname{sinc}^2 0 = 1$$

$$\operatorname{sinc}^2 n = 0 \qquad n = \text{nonzero integer}$$

$$\int_{-\infty}^{\infty} \operatorname{sinc}^2 x = 1.$$

In two dimensions a function analogous to sinc x is

$$\frac{J_1(\pi r)}{2r},$$

which has unit volume, a central value of $\pi/4$, and a two-dimensional Fourier transform $\Pi(s)$. Another generalization to two dimensions, which has analogous filtering and interpolating properties, is

$$\operatorname{sinc} x \operatorname{sinc} y.$$

The two-dimensional Fourier transform of this function is $\Pi(u)\Pi(v)$.

Pictorial representation

Certain conventions in graphical representations will be adopted for the purpose of clarity. For example, theorems relating to Fourier transforms will be illustrated, where possible, by examples which are real. In these diagrams the points where the abscissa and ordinate are equal to unity are marked if it is appropriate to do so.

In representing purely imaginary quantities a dashed line is always

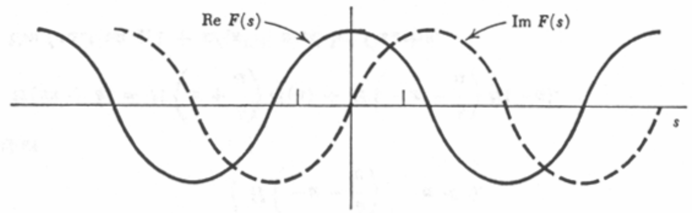


Fig. 4.18 Representation of a complex function F(s) by its real and imaginary parts.

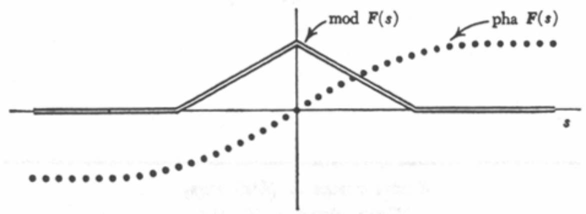


Fig. 4.19 Complex function shown in modulus and phase.

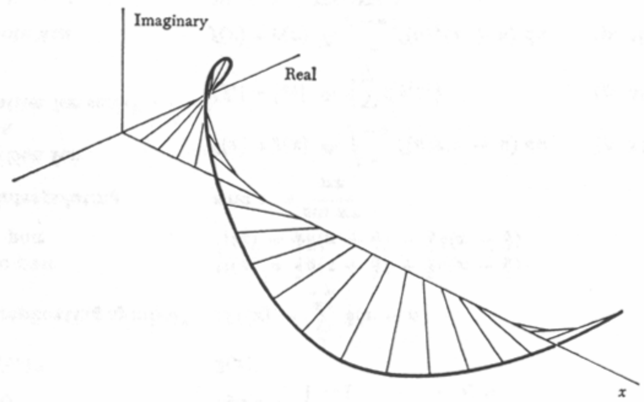


Fig. 4.20 Complex function in three dimensions.

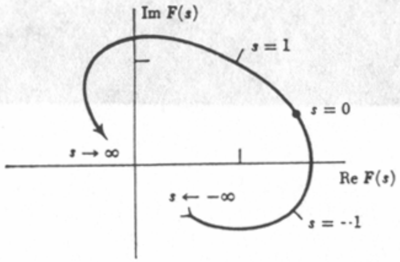


Fig. 4.21 Values of F(s) on the complex plane of F(s).

used; hence complex quantities may be shown unambiguously and clearly by their real and imaginary parts (see Fig. 4.18).

Sometimes a complex function may be shown to advantage by plotting its modulus and phase, as in Fig. 4.19. Another method is to show a three-dimensional diagram (see Fig. 4.20), and a third method is to show a locus on the complex plane (see Fig. 4.21).

Summary of special symbols

A small number of special symbols which are used extensively throughout this work are summarized in Table 4.1 and in Fig. 4.22 for reference.

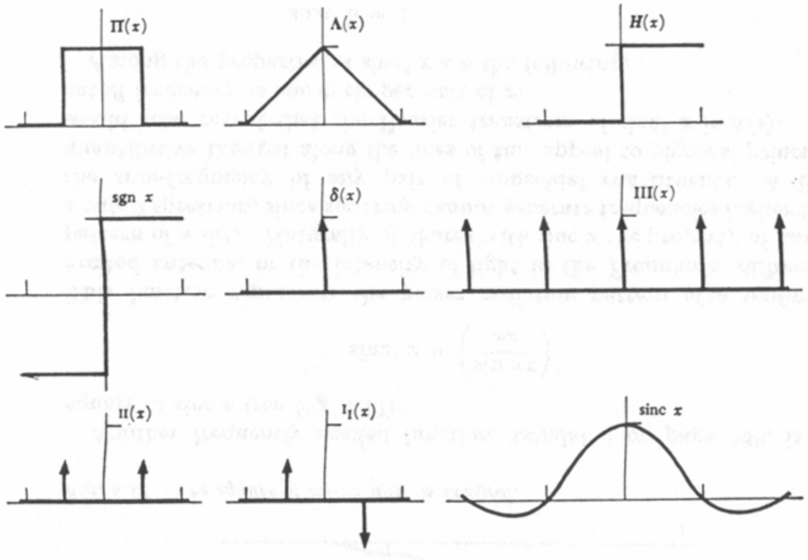


Fig. 4.22 Summary of special symbols.

Table 4.1 Special symbols

Function	*Notation*	
Rectangle	$\Pi(x) = \begin{cases} 1 & \lvert x\rvert < \frac{1}{2} \\ 0 & \lvert x\rvert > \frac{1}{2} \end{cases}$	
Triangle	$\Lambda(x) = \begin{cases} 1 - \lvert x\rvert & \lvert x\rvert < 1 \\ 0 & \lvert x\rvert > 1 \end{cases}$	
Heaviside unit step	$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$	
Sign (signum)	$\operatorname{sgn} x = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$	
Impulse symbol*	$\delta(x)$	
Sampling or replicating symbol*	$\mathrm{III}(x) = \sum_{-\infty}^{\infty} \delta(x - n)$	
Even impulse pair	$\mathrm{II}(x) = \frac{1}{2}\delta(x + \frac{1}{2}) + \frac{1}{2}\delta(x - \frac{1}{2})$	
Odd impulse pair	${}^{\mathrm{I}}{}_{\mathrm{I}}(x) = \frac{1}{2}\delta(x + \frac{1}{2}) - \frac{1}{2}\delta(x - \frac{1}{2})$	
Filtering or interpolating	$\operatorname{sinc} x = \dfrac{\sin \pi x}{\pi x}$	
Asterisk notation for convolution	$f(x) * g(x) \triangleq \int_{-\infty}^{\infty} f(u)g(x - u)\, du$	(p. 25)
Asterisk notation for serial products	$\{f_i\} * \{g_i\} \triangleq \{\sum_j f_j g_{i-j}\}$	(p. 32)
Pentagram notation	$f(x) \star g(x) \triangleq \int_{-\infty}^{\infty} f(u)g(x + u)\, du$	(p. 40)
Various two-dimensional functions	${}^2\Pi(x,y) = \Pi(x)\Pi(y)$ ${}^2\delta(x,y) = \delta(x)\,\delta(y)$ ${}^2\mathrm{III}(x,y) = \mathrm{III}(x)\mathrm{III}(y)$ ${}^2\operatorname{sinc}(x,y) = \operatorname{sinc} x \operatorname{sinc} y$	

* See Chapter 5.

Problems

1 Show that

$$H(ax + b) = \begin{cases} H\left(x + \dfrac{b}{a}\right) & a > 0 \\ H\left(-x - \dfrac{b}{a}\right) & a < 0, \end{cases}$$

and hence that

$$H(ax + b) = H\left(x + \frac{b}{a}\right)H(a) + H\left(-x - \frac{b}{a}\right)H(-a).$$

2 Discuss the function $\frac{1}{2}[1 + x/(x^2)^{\frac{1}{2}}]$ used by Cauchy.

3 Show that the operation $H(x)\,*$ is an integrating operation in the sense that

$$H(x) * [f(x)H(x)] = \int_0^x f(x)\,dx.$$

4 Calculate $(d/dx)\,[\Pi(x) * H(x)]$ and prove that $(d/dx)[f(x) * H(x)] = f(x)$.

5 By evaluating the integral, prove that $\operatorname{sinc} x * \operatorname{sinc} x = \operatorname{sinc} x$.

6 Prove that $\operatorname{sinc} x * J_0(\pi x) = J_0(\pi x)$.

7 Prove that $4 \operatorname{sinc} 4x * \sin x = \sin x$.

8 Show that

$$\begin{aligned}\Pi(x) &= H(x + \tfrac{1}{2}) - H(x - \tfrac{1}{2})\\ &= H(\tfrac{1}{2} + x) + H(\tfrac{1}{2} - x) - 1\\ &= H(\tfrac{1}{4} - x^2)\\ &= \tfrac{1}{2}[\operatorname{sgn}(x + \tfrac{1}{2}) - \operatorname{sgn}(x - \tfrac{1}{2})]\end{aligned}$$

and that $\Pi(x^2) = \Pi(x/2^{\frac{1}{2}})$.

9 Show that

$$\begin{aligned}\Lambda(x) &= \Pi(x) * \Pi(x)\\ &= \Pi(x) * H(x + \tfrac{1}{2}) - \Pi(x) * H(x - \tfrac{1}{2}).\end{aligned}$$

10 Experiment with the equation $f[f(x)] = f(x)$ and note that $f(x) = \operatorname{sgn} x$ is a solution. Find other solutions and attempt to write down the general solution compactly with the aid of step-function notation.

11 Show that $\operatorname{erf} x = 2\Phi(2^{\frac{1}{2}}\sigma x) - 1$, where

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} e^{-u^2/2\sigma^2}\,du.$$

12 Show that the first derivative of $\Lambda(x)$ is given by

$$\Lambda'(x) = -\Pi\left(\frac{x}{2}\right) \operatorname{sgn} x$$

and calculate the second derivative.

13 In abbreviated notation the relation of $\Lambda(x)$ to $\Pi(x)$ could be written $\Lambda = \Pi * \Pi$ or $\Lambda = \Pi^{*2}$. Show that

$$\Pi^{*3} = \Pi * \Lambda = \tfrac{1}{2}(x + 1\tfrac{1}{2})^2\Pi(x + 1) + (\tfrac{3}{4} - x^2)\Pi(x) + \tfrac{1}{2}(x - 1\tfrac{1}{2})^2\Pi(x - 1).$$

Show also that

$$\begin{aligned}\Pi^{*3} = \tfrac{1}{2}(x + 1\tfrac{1}{2})^2H(x + 1\tfrac{1}{2}) - \tfrac{3}{2}(x + \tfrac{1}{2})^2H(x + \tfrac{1}{2}) + \tfrac{3}{2}(x - \tfrac{1}{2})^2H(x - \tfrac{1}{2})\\ - \tfrac{1}{2}(x - 1\tfrac{1}{2})^2H(x - 1\tfrac{1}{2}).\end{aligned}$$

14 Examine the derivatives of Π^{*3} at $x = \frac{1}{2}\pm$ and $x = 1\frac{1}{2}\pm$ and reach some conclusion about the continuity of slope and curvature.

15 Show that $(d/dx)|x| = \operatorname{sgn} x$ and that $(d/dx) \operatorname{sgn} x = 2\delta(x)$. Comment on the fact that

$$\frac{d^2|x|}{dx^2} = \frac{d^2}{dx^2}[2xH(x)] = 2\delta(x).$$

Chapter 5 The impulse symbol

It is convenient to have notation for intense unit-area pulses so brief that measuring equipment of a given resolving power is unable to distinguish between them and even briefer pulses. This concept is covered in mechanics by the term "impulse." The important attribute of an impulse is its integral; the precise details of its form are unimportant. The idea has been current for a century[1] or more in mathematical circles and was extensively employed by Heaviside, for example, who used the symbol $p1$.[2] The notation $\delta(x)$, which was subsequently introduced into quantum mechanics by Dirac,[3] is now in general use. The underlying concept permeates physics. Point masses, point charges, point sources, concentrated forces, line sources, surface charges, and the like are familiar and accepted entities in physics. Of course, these things do not exist. Their conceptual value stems from the fact that the impulse response—the effect associated with the impulse (point mass, point charge, and the like)—may be indistinguishable, given measuring equipment of specified resolving power, from the response due to a physically realizable pulse. It is then a convenience to have a name for pulses which are so brief and intense that making them any briefer and more intense does not matter.

We have in mind an infinitely brief or concentrated, infinitely strong

[1] For historical examples from the writings of Hermite, Cauchy, Poisson, Kirchhoff, Helmholtz, Kelvin, and Heaviside, see B. van der Pol and H. Bremmer, "Operational Calculus Based on the Two-sided Laplace Integral," Cambridge University Press, Cambridge, England, 1955.

[2] This symbol means the derivative of the unit step function.

[3] P. A. M. Dirac, "The Principles of Quantum Mechanics," 3d ed., Oxford University Press, Oxford, 1947.

unit-area impulse and therefore wish to write

$$\delta(x) = 0 \qquad x \neq 0$$

and

$$\int_{-\infty}^{\infty} \delta(x)\,dx = 1.$$

However, the impulse symbol[4] $\delta(x)$ does not represent a function in the sense in which the word is used in analysis (to stress this fact Dirac coined the term "improper function"), and the above integral is not a meaningful quantity until some *convention* for interpreting it is declared. Here we use it to mean

$$\lim_{\tau \to 0} \int_{-\infty}^{\infty} \tau^{-1}\Pi\left(\frac{x}{\tau}\right) dx.$$

The function $\tau^{-1}\Pi(x/\tau)$ is a rectangle function of height τ^{-1} and base τ and has unit area; as τ tends to zero a sequence of unit-area pulses of ever-increasing height is generated. The limit of the integral is, of course, equal to unity. In other words, to interpret expressions containing the impulse symbol, we fall back on certain sequences of finite unit-area pulses of brief but nonzero duration, and of some particular shape. We perform the operations indicated, such as integration, differentiation, multiplication, and then discuss limits as the duration approaches zero. There is some convenience mathematically in taking the pulse shape always Gaussian; obviously, one has to prepare for possible awkwardness in retaining $\Pi(x)$ as a choice, where differentiation is involved. However, the essence of the physics is that the pulse shape should not matter, and we therefore proceed under the expectation that the choice of pulse shape will remain at our disposal.

We adapt our approach to each case as it arises. Later we give a systematic presentation of the theory of generalized functions, a recently developed exposition of the sequence idea in tidied-up mathematical form.

The need to broaden Fourier transform theory was mentioned earlier in connection with functions, such as periodic functions, which do not possess Fourier transforms. The term "transform pair in the limit" was introduced to describe cases where one or both members of the transform pair are generalized functions. All these cases can be expressed with the aid of the impulse symbol and its derivatives, $\delta(x)$, $\delta'(x)$, and so on, which thus furnish the notation for the broadened theory.

The convenience of the impulse symbol lies in its reserve over detail. As a specific example of the relevance of this feature to physical systems, consider an electrical network, say a low-pass filter. An applied pulse of voltage produces a certain transient response, and it is readily observable, as the applied pulses are made briefer and briefer, that the response settles down to a definite form. It is also observable that the form of the response is then independent of the input pulse shape, be it rectangular, triangular, or even a pair of pulses. This happens because the high-frequency components, which distinguish the different applied pulses, produce negligible response. The network is thus characterized by a certain definite, readily observable form of response, which can be elicited by a multiplicity of applied waveforms, the details of which are irrelevant; it is necessary only that they be brief enough. Since the response may be scrutinized with an oscilloscope of the highest precision and time resolution, we must, of course, be prepared to keep the applied pulse duration shorter than the minimum set by the quality of the measuring instrument. The impulse symbol enables us to make abbreviated statements about arbitrarily shaped indefinitely brief pulses.

[4] We use the word "symbol" systematically to signpost the entities that are not functions; we may also use the term "generalized function," introduced in 1953 by Temple.

An intimate relationship between the impulse symbol and the unit step function follows from the property that $\int_{-\infty}^{x} \delta(x')\,dx'$ is unity if x is positive but zero if x is negative. Hence

$$\int_{-\infty}^{x} \delta(x')\,dx' = H(x).$$

This equation furnishes an opportunity to illustrate the interpretation of an expression containing the impulse symbol. First we replace $\delta(x)$ by the pulse sequence $\tau^{-1}\Pi(x/\tau)$ and contemplate the sequence of integrals

$$\int_{-\infty}^{x} \tau^{-1}\Pi\left(\frac{x'}{\tau}\right) dx'.$$

As long as τ is not zero or infinite, each such integral is a function of x that may be described as a ramp-step function, as illustrated in Fig. 5.1. Now fix x, and consider the limit of the sequence of values of the integral generated as τ approaches zero. We see that if we have fixed on a positive (negative) x, then the limit of the integral will be unity (zero). Therefore, in accordance with the definition of $H(x)$ we can write

$$\lim_{\tau \to 0} \int_{-\infty}^{x} \tau^{-1}\Pi\left(\frac{x'}{\tau}\right) dx' = H(x).$$

The equation

$$\int_{-\infty}^{x} \delta(x')\,dx' = H(x)$$

is shorthand for this.

Since under ordinary circumstances

$$\int_{-\infty}^{x} f(x')\,dx' = g(x)$$

implies that

$$f(x) = g'(x),$$

one writes by analogy

$$\delta(x) = \frac{d}{dx} H(x)$$

and states that "the derivative of the unit step function is the impulse symbol." Since the unit step function does not in fact possess a derivative at the origin, this statement must be interpreted as shorthand for "the derivatives of a sequence of differentiable functions that approach $H(x)$ as a limit constitute a suitable defining sequence for $\delta(x)$." The ramp-step functions of Fig. 5.1 are differentiable and approach $H(x)$. Since the amount of the step is in all cases unity, the area under each derivative function is unity, thus qualifying the sequence of derivative functions as a suitable sequence to define a unit impulse.

It is not *necessary* to deal in terms of sequences of rectangle functions to discuss impulses. Representations of $\delta(x)$ in terms of various pulse shapes include the following sequences, generated as τ approaches zero (through positive values).

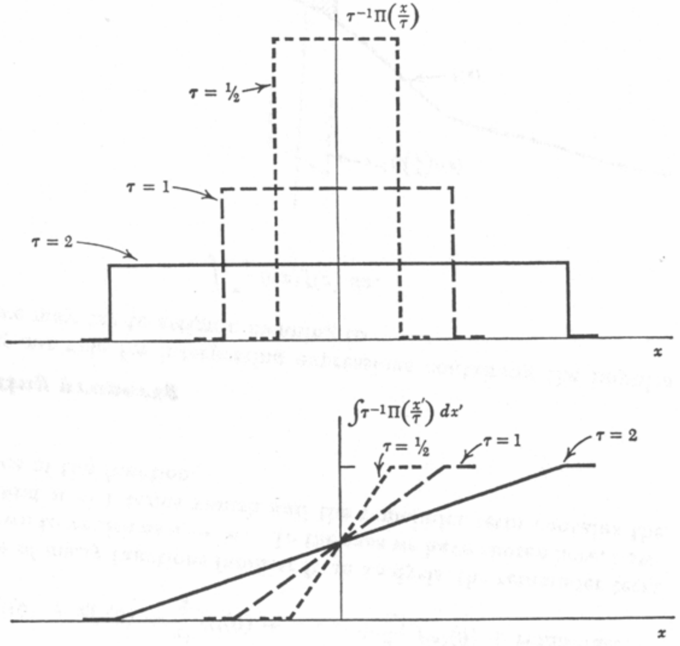


Fig. 5.1 A rectangular pulse sequence and the sequence of ramp-step functions obtained by integration.

The sequence

$$\tau^{-1}\Pi\left(\frac{x - \frac{1}{2}\tau}{\tau}\right)$$

is composed of rectangle functions, all having their left-hand edge at the origin; it is often considered in the analysis of circuits where transients cannot exist prior to $t = 0$, the instant at which some switch is thrown. If one uses this sequence instead of the centered sequence, the results will not necessarily be the same. This is discussed below in connection with the sifting property. The sequence

$$\tau^{-1}\, e^{-\pi x^2/\tau^2}$$

of Gaussian profiles, which has been mentioned previously, has the convenient property that derivatives of all orders exist. On the other hand, these profiles lack the convenience of being nonzero over only a finite range of x. The sequence

$$\tau^{-1}\Lambda\left(\frac{x}{\tau}\right)$$

of triangle functions is useful for discussing situations where first derivatives are needed, since the profiles are continuous. In addition, they have an advantage in being zero outside the interval in which $|x| < \tau$. The sequence

$$\tau^{-1} \operatorname{sinc} \frac{x}{\tau}$$

has the curious property of not dying out to zero where $x \neq 0$; at any value of x not equal to zero the value oscillates without diminishing as $\tau \to 0$. The sequence serves perfectly well to define $\delta(x)$ for a reason that is given below in connection with the sifting property. The resonance profiles

$$\frac{\tau}{\pi(x^2 + \tau^2)}$$

decay rather slowly with increasing x. A product with an arbitrary function may well have infinite area. To eliminate this possibility completely, one is led to contemplate sequences that are zero outside a finite range, thus obtaining freedom to accept products with functions having any kind of asymptotic behavior. In addition, one would like to have derivatives of all orders exist.

$$\tau^{-1} \exp\left[\frac{-1}{1 - (x/\tau)^2}\right] \Pi\left(\frac{x}{2\tau}\right)$$

is a specific example of a sequence of profiles, each of which is zero outside the interval in which $|x| < \tau$, and each of which possesses derivatives of all orders. To see that it is possible for a function to descend to zero with zero slope, zero curvature, and all higher derivatives zero, differentiate the function

$$e^{-1/x}H(x),$$

and evaluate the derivatives at $x = 0$. There may seem to be a clash with the Maclaurin formula, according to which

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2}f''(0) + \cdots + \frac{x^n}{n!}f^{(n)}(0) + \text{remainder}.$$

In the case of many functions familiar from analysis, the remainder term can be shown to vanish as $n \to \infty$. In the case we have chosen here, however, the first $n + 1$ terms vanish and the remainder term contains the whole value of the function.

The sifting property

Following our rule for interpreting expressions containing the impulse symbol, we may try to assign a meaning to

$$\int_{-\infty}^{\infty} \delta(x)f(x)\,dx.$$

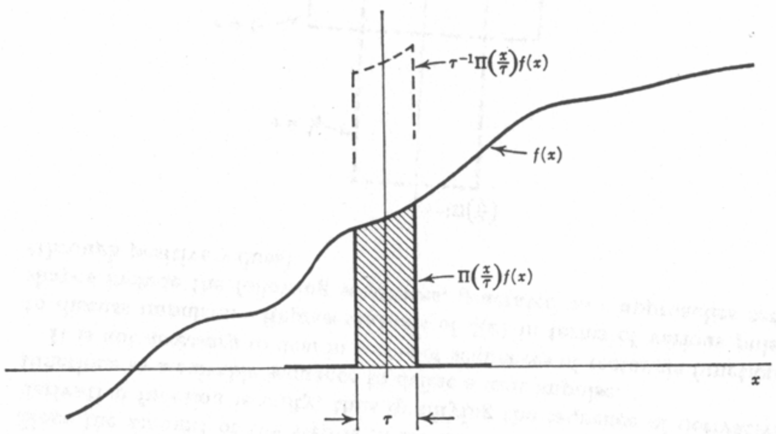


Fig. 5.2 Explaining the sifting property. The shaded area is approximately $\tau f(0)$.

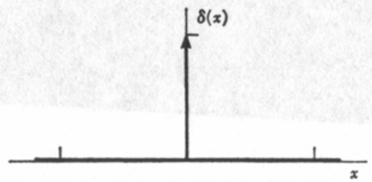


Fig. 5.3 Graphical representation of the impulse symbol $\delta(x)$ as a spike of unit height.

Thus we substitute the sequence $\tau^{-1}\Pi(x/\tau)$ for $\delta(x)$, perform the multiplication and integration, and finally take the limit of the integral as $\tau \to 0$:

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} \tau^{-1}\Pi\left(\frac{x}{\tau}\right)f(x)\,dx.$$

In Fig. 5.2 the integrand is indicated in broken outline. Its area is τ^{-1} times the shaded area. The shaded area, whose width is τ and whose average height is approximately $f(0)$, amounts to approximately $\tau f(0)$. Hence the area under the integrand approaches $f(0)$ as τ approaches zero. Thus we write

$$\int_{-\infty}^{\infty} \delta(x)f(x)\,dx = f(0)$$

and refer to this statement as the sifting property of the impulse symbol, since the operation on $f(x)$ indicated on the left-hand side sifts out a single value of $f(x)$. It will be seen that it is immaterial what sort of pulse is incorporated in the integrand, and this fact is the essence of the utility of $\delta(x)$. It just stands for a unit pulse whose duration is much smaller than any interval of interest, and consequently whose pulse *shape* means nothing to us; only its integral counts. It will be represented on graphs (see Fig. 5.3) as a spike of unit height, and impulses in general will be shown as spikes of height equal to their integral.

It is clear that we can also write

$$\int_{-\infty}^{\infty} \delta(x-a)f(x)\,dx = f(a)$$

and

$$\int_{-\infty}^{\infty} \delta(x)f(x-a)\,dx = f(-a).$$

The resemblance to the convolution integral can be emphasized by writing

$$\int_{-\infty}^{\infty} \delta(x')f(x-x')\,dx' = \int_{-\infty}^{\infty} \delta(x-x')f(x')\,dx' = f(x)$$

or, in asterisk notation,

$$\delta(x) * f(x) = f(x) * \delta(x) = f(x).$$

If $f(x)$ has a jump at $x = 0$, a little thought devoted to a diagram such as the one Fig. 5.2 will show that the sifting integral will have a limiting value of $\frac{1}{2}[f(0+) + f(0-)]$. Consequently, it is more general to write

$$\delta(x) * f(x) = \frac{f(x+) + f(x-)}{2}.$$

The expression on the right-hand side differs from $f(x)$ only by a null function, and hence the refinement is ordinarily not important. This does not alter the fact that $\frac{1}{2}[f(x+) + f(x-)]$ can be different in value from $f(x)$.

The asymmetrical sequence $\tau^{-1}\Pi[(x - \frac{1}{2}\tau)/\tau]$ mentioned above will be seen to have the property of sifting out $f(x+)$. At points of discontinuity of $f(x)$, the use of this asymmetrical sequence therefore gives a different result. In transient analysis, where discontinuities at the switching instant $t = 0$ are particularly common, the choice of sequence can thus appear to give different answers. The difference, however, can only be instantaneous.

The impulse symbol has many fascinating properties, most of which can be proved easily. An important one which must be watched carefully in algebraic manipulation is

$$\delta(ax) = \frac{1}{|a|}\,\delta(x);$$

that is, if the scale of x is compressed by a factor a, thus reducing the area of the pulses which previously had unit area, then the strength of the impulse is reduced by the factor $|a|$. The modulus sign allows for the property

$$\delta(-x) = \delta(x).$$

From this it would seem that the impulse symbol has the property of evenness; however, we gave an equation earlier involving a sequence of displaced rectangle functions which were not themselves even (Prob. 19).

It can easily be shown by considering sequences of pulses that we may write, if $f(x)$is continuous at $x = 0$,

$$f(x)\,\delta(x) = f(0)\,\delta(x).$$

From the sifting property, putting $f(x) = x$, we have

$$\int_{-\infty}^{\infty} x\,\delta(x)\,dx = 0.$$

One generally writes

$$x\,\delta(x) \equiv 0,$$

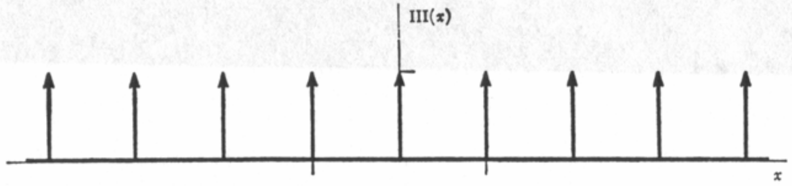


Fig. 5.4 The shah symbol III(x).

although if the prelimit graphs are contemplated, it will be seen that this equation conceals a nonvanishing component reminiscent of the Gibbs phenomenon in Fourier series. Thus it is true that

$$\lim_{\tau\to 0}\left[x\tau^{-1}\Pi\left(\frac{x}{\tau}\right)\right] = 0 \qquad \text{for all } x;$$

nevertheless,

$$\lim_{\tau\to 0}\left[x\tau^{-1}\Pi\left(\frac{x}{\tau}\right)\right]_{\max} = \tfrac{1}{2},$$

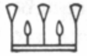
and, moreover, the limit of the minimum value is $-\frac{1}{2}$. Consequently, among those functions which are identically zero, $x\delta(x)$ is rather curious, and one has the feeling that if it could be applied to the deflecting electrodes of an oscilloscope, one would see spikes.

The sampling or replicating symbol III(x)

Consider an infinite sequence of unit impulses spaced at unit interval as shown in Fig. 5.4. Any reservations that apply to the impulse symbol $\delta(x)$ apply equally in this case; indeed, even more may be needed because we have to deal with an infinite number of infinite discontinuities and a nonconvergent infinite integral. For example, *all* the conditions for existence of a Fourier transform are violated. The conception of an infinite sequence of impulses proves, however, to be extremely useful—and easy to manipulate algebraically.

To describe this conception we introduce the *shah*[5] symbol III(x) and write

$$\text{III}(x) = \sum_{n=-\infty}^{\infty} \delta(x - n).$$

[5] The symbol III is pronounced *shah* after the Cyrillic character Ш, which is said to have been modeled on the Hebrew letter  (*shin*), which in turn may derive from the Egyptian , a hieroglyph depicting papyrus plants along the Nile.

Various obvious properties may be pointed out:

$$\begin{aligned} \mathrm{III}(ax) &= \frac{1}{|a|}\sum \delta\left(x - \frac{n}{a}\right) \\ \mathrm{III}(-x) &= \mathrm{III}(x) \\ \mathrm{III}(x+n) &= \mathrm{III}(x) \qquad n \text{ integral} \\ \mathrm{III}(x-\tfrac{1}{2}) &= \mathrm{III}(x+\tfrac{1}{2}) \\ \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} \mathrm{III}(x)\,dx &= 1 \\ \mathrm{III}(x) &= 0 \qquad x \neq n. \end{aligned}$$

Evidently, III(x) is periodic with unit period.

A periodic *sampling* property follows as a generalization of the sifting integral already discussed in connection with the impulse symbol. Thus multiplication of a function $f(x)$ by III(x) effectively samples it at unit intervals:

$$\mathrm{III}(x)f(x) = \sum_{n=-\infty}^{\infty} f(n)\,\delta(x-n).$$

The information about $f(x)$ in the intervals between integers where III(x) = 0 is not contained in the product; however, the values of $f(x)$ at integral values of x are preserved (see Fig. 5.5).

The sampling property makes III(x) a valued tool in the study of a wide variety of subjects (for example, the radiation patterns of antenna arrays, the diffraction patterns of gratings, raster scanning in television and radar, pulse modulation, data sampling, Fourier series, and computing at discrete tabular intervals).

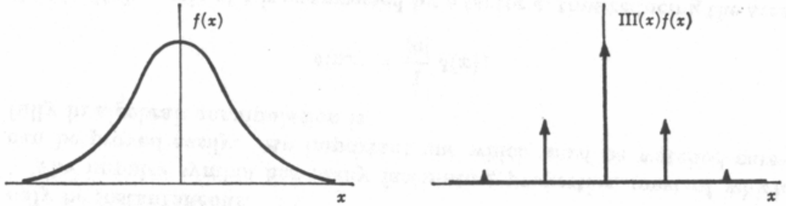


Fig. 5.5 The sampling property of III(x).

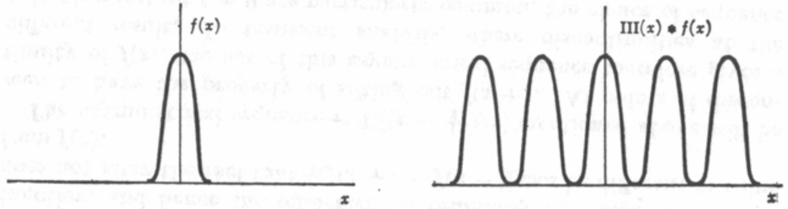


Fig. 5.6 The replicating property of III(x).

Just as important as the sampling property under multiplication is a *replicating* property exhibited when III(x) enters into convolution with a function $f(x)$. Thus

$$\mathrm{III}(x) * f(x) = \sum_{n=-\infty}^{\infty} f(x-n);$$

as shown in Fig. 5.6, the function $f(x)$ appears in replica at unit intervals of x *ad infinitum* in both directions. Of course, if $f(x)$ spreads over a base more than one unit wide, there is overlapping.

The III symbol is thus also applicable wherever there are periodic structures. This twofold character is not accidental, but is connected with the fact that III is its own Fourier transform (in the limit), which of course makes it twice as useful as it otherwise would have been.

The self-reciprocal property under the Fourier transformation is derived later.

The even and odd impulse pairs II(x) *and* $^{\mathrm{I}}_{\mathrm{I}}$($x$)

Figure 5.7 shows the often-needed impulse-pair symbols defined by

$$\begin{aligned} \mathrm{II}(x) &= \tfrac{1}{2}\delta(x+\tfrac{1}{2}) + \tfrac{1}{2}\delta(x-\tfrac{1}{2}), \\ {}^{\mathrm{I}}_{\mathrm{I}}(x) &= \tfrac{1}{2}\delta(x+\tfrac{1}{2}) - \tfrac{1}{2}\delta(x-\tfrac{1}{2}). \end{aligned}$$

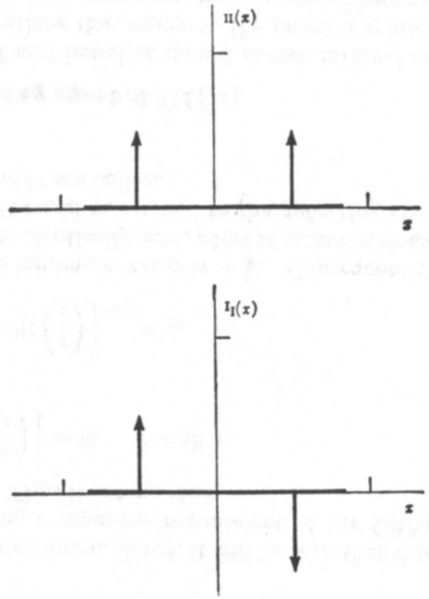


Fig. 5.7 The even and odd impulse pairs, II(x) *and* $^{\mathrm{I}}_{\mathrm{I}}$($x$).

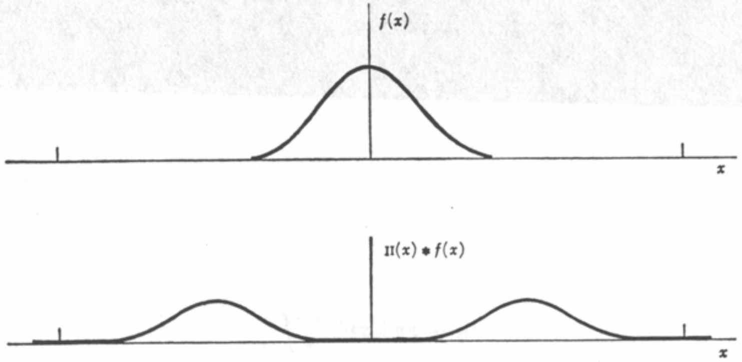


Fig. 5.8 The convolution of $\text{II}(x)$ *with* $f(x)$.

The impulse pairs derive importance from their transform relationship to the cosine and sine functions. Thus

$$\text{II}(x) \supset \cos \pi s \qquad \cos \pi x \supset \text{II}(s)$$
$$\text{I}_\text{I}(x) \supset i \sin \pi s \qquad \sin \pi x \supset i\text{I}_\text{I}(s).$$

When convolved with a function $f(x)$, the even impulse pair $\text{II}(x)$ has a duplicating property. Thus, as illustrated in Fig. 5.8,

$$\text{II}(x) * f(x) = \tfrac{1}{2}f(x + \tfrac{1}{2}) + \tfrac{1}{2}f(x - \tfrac{1}{2}).$$

There are occasions when $\text{II}(x)$ might better consist of two unit impulses, but as defined it is normalized to unit area; that is,

$$\int_{-\infty}^{\infty} \text{II}(x)\, dx = 1,$$

which has advantages.

If the finite difference of $f(x)$ is defined by

$$\Delta f(x) = f(x + \tfrac{1}{2}) - f(x - \tfrac{1}{2}),$$

then

$$\Delta f(x) = 2\,\text{I}_\text{I}(x) * f(x).$$

Thus the finite difference operator can be expressed as

$$\Delta \equiv 2\,\text{I}_\text{I} * .$$

Derivatives of the impulse symbol

The first derivative of the impulse symbol is defined symbolically by

$$\delta'(x) = \frac{d}{dx}\,\delta(x).$$

The mental picture that accompanies this conception is the same as that involved in the conception of an infinitesimal dipole in electrostatics, and it is well known that the idea of an infinitesimal dipole is convenient and easy to think with physically. The $\delta'(x)$ notation carries this facility over into mathematical form, but, of course, there are difficulties because we cannot ask a function to go positively infinite just to the left of the origin and negatively infinite just to the right, and to be zero where $|x| > 0$. To cap this we would wish to write $\delta'(0) = 0$.

For rigorous interpretation of statements involving $\delta'(x)$ we may fall back on sequences of pulses such as were invoked in connection with $\delta(x)$, and consider their derivatives (two examples are given in Fig. 5.9). Then statements such as

$$\int_{-\infty}^{\infty} \delta'(x)\, dx = 0$$

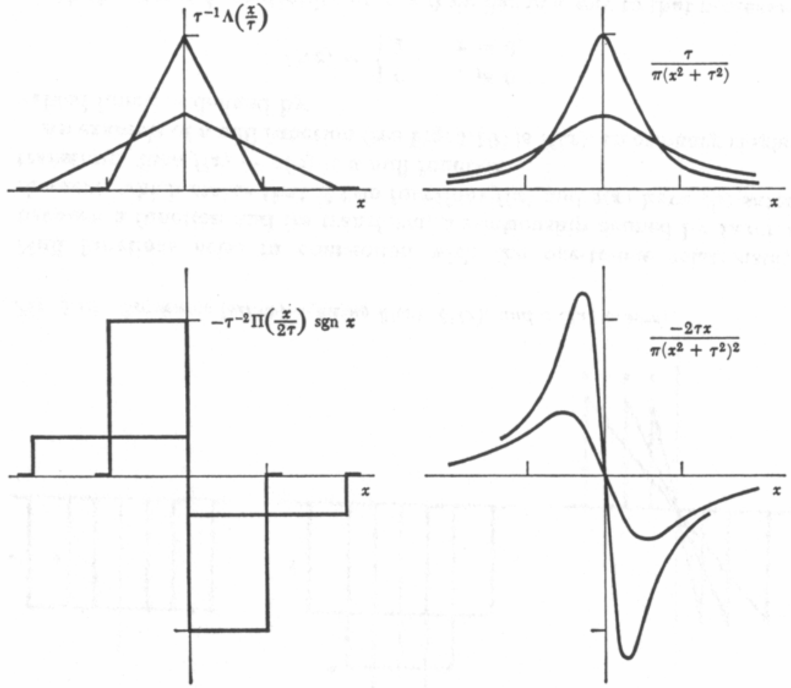


Fig. 5.9 Pulse sequences (above) and their derivatives (below) which, as $\tau \to 0$, *are used for contemplating the meaning of* $\delta(x)$ *and* $\delta'(x)$.

are deemed to be shorthand for statements such as

$$\lim_{\tau \to 0} \int_{-\infty}^{\infty} \left[\frac{-2\tau x}{\pi(x^2 + \tau^2)^2} \right] dx = 0,$$

where the quantity in brackets is the derivative of one of the pulse shapes considered previously in discussing $\delta(x)$. The precise form of pulse adopted is unimportant—even a rectangular one will do—but in later work the possibility that a differentiable pulse shape may offer an advantage should be considered.

A derivative-sifting property

$$\delta' * f \equiv \int_{-\infty}^{\infty} \delta'(x - x') f(x')\, dx' = f'(x)$$

may be established in this way. Further properties are

$$\int_{-\infty}^{\infty} x\, \delta'(x)\, dx = -1$$

$$\int_{-\infty}^{\infty} |\delta'(x)|\, dx = \infty$$

$$x^2\, \delta'(x) = 0$$

$$\delta'(-x) = -\delta'(x) \qquad x\, \delta'(x) = -\delta(x)$$

$$f(x)\, \delta'(x) = f(0)\, \delta'(x) - f'(0)\, \delta(x).$$

The following relations apply to derivatives of higher order.

$$\int_{-\infty}^{\infty} \delta''(x)\, dx = 0$$

$$\delta''(x) * f(x) = f''(x)$$

$$\int_{-\infty}^{\infty} x^2\, \delta''(x)\, dx = 2$$

$$\delta^{(n)}(x) = (-1)^n n!\, x^{-n}\, \delta(x)$$

$$\delta^{(n)}(x) * f(x) = f^{(n)}(x)$$

$$\int_{-\infty}^{\infty} \delta^{(n)}(x) f(x)\, dx = (-1)^n f^{(n)}(0)$$

Null functions

Null functions are known chiefly for having Fourier transforms which are zero, while not themselves being identically zero. By definition, $f(x)$ is a null function if

$$\int_a^b f(x)\, dx = 0$$

for all a and b. An alternative statement is

$$\int_{-\infty}^{\infty} |f(x)|\, dx = 0.$$

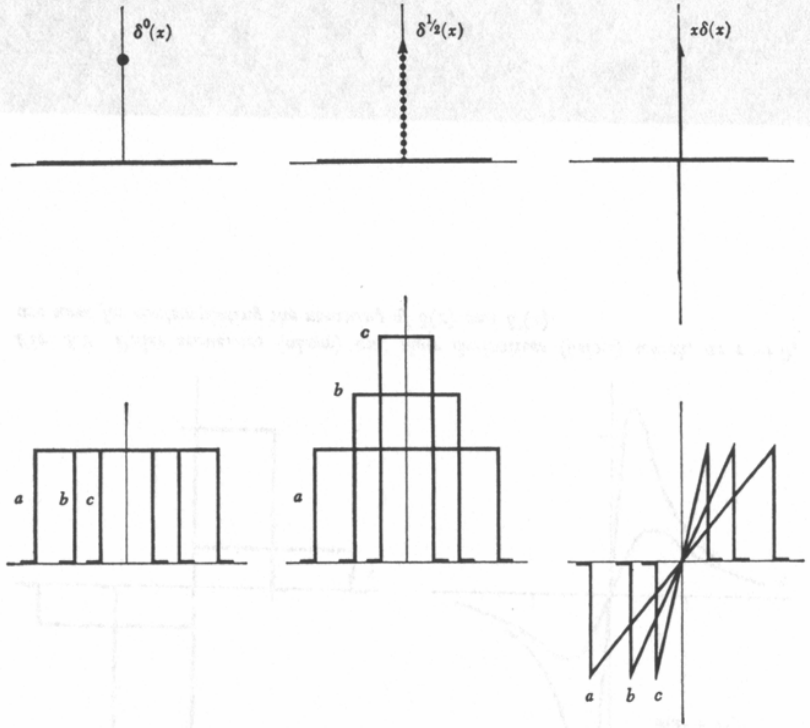


Fig. 5.10 Sequences (below) defining $\delta^0(x)$, $\delta^{\frac{1}{2}}(x)$, and $x\,\delta(x)$ (above).

Null functions arise in connection with the one-to-one relationship between a function and its transform, a relationship defined by Lerch's theorem, which states that if two functions $f(x)$ and $g(x)$ have the same transform, then $f(x) - g(x)$ is a null function.

An example of a null function (see Fig. 5.10) is $\delta^0(x)$, an ordinary single-valued function defined by

$$\delta^0(x) = \begin{cases} 0 & x \neq 0 \\ 1 & x = 0, \end{cases}$$

which thus has a discontinuity at $x = 0$ similar in a way to that possessed by $H(x)$ [defined so that $H(0) = \frac{1}{2}$]. Under the ordinary rules of integration, the integral of $\delta^0(x)$ is certainly zero. However, it aptly describes the current taken by a series combination of a resistance and a capacitance from a battery, in the limit as the capacitance approaches zero.

We can now more succinctly state the relation between $H(x)$ and the step function $\hat{H}(x)$ of Fig. 4.14; thus

$$H(x) = \hat{H}(x) + \tfrac{1}{2}\delta^0(x),$$

the difference between the two being a null function.

The symbol $\delta^{\frac{1}{2}}(x)$ has to be considered in terms of sequences of pulses in the same way as $\delta(x)$. Consider the sequence

$$\tau^{-\frac{1}{2}}\Pi\left(\frac{x}{\tau}\right)$$

as $\tau \to 0$. Then we can attach meaning to the statements

$$\delta^{\frac{1}{2}}(x) = \begin{cases} 0 & x \neq 0 \\ \infty & x = 0, \end{cases}$$

$$\int_{-\infty}^{\infty} \delta^{\frac{1}{2}}(x)\,dx = 0,$$

and

$$\int_{-\infty}^{\infty} [\delta^{\frac{1}{2}}(x)]^2\,dx = 1.$$

We could describe $\delta^{\frac{1}{2}}(x)$ as a null symbol.

Exercise Would we wish to call $\delta'(x)$ a null symbol?

Some functions in two and more dimensions

One encounters the two- and three-dimensional impulse symbols $^2\delta(x,y)$, $^3\delta(x,y,z)$, as natural generalizations of $\delta(x)$. For example, $^2\delta(x,y)$ describes the pressure distribution over the xy plane when a concentrated unit force is applied at the origin; $^3\delta(x,y,z)$ describes the charge density in a volume containing a unit charge at the point (0,0,0). In establishing properties of $^2\delta(x,y)$ one considers a sequence, as $\tau \to 0$, of functions such as $\tau^{-2}\Pi(x/\tau)\Pi(y/\tau)$ or $(4/\pi)\tau^{-2}\Pi[(x^2+y^2)^{\frac{1}{2}}/\tau]$, which have unit volume

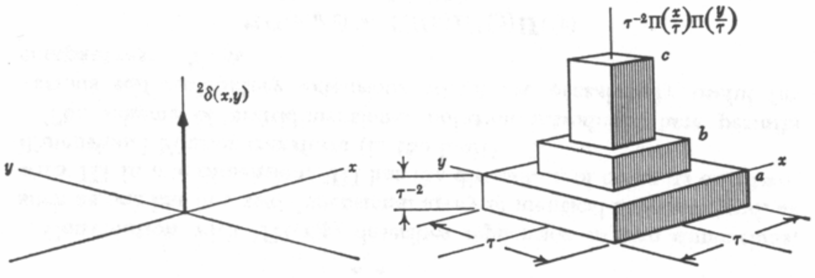


Fig. 5.11 The two-dimensional impulse symbol $^2\delta(x,y)$ and a defining sequence of functions a,b,c.

(Fig. 5.11). Then we have

$$^2\delta(x,y) = \begin{cases} 0 & x^2+y^2 \neq 0 \\ \infty & x^2+y^2 = 0, \end{cases}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} {}^2\delta(x,y)\,dx\,dy = 1,$$

$$^2\delta(ax,by) = \frac{1}{|ab|}\,{}^2\delta(x,y),$$

and the very interesting relation

$$^2\delta(x,y) = \delta(x)\,\delta(y).$$

Introducing the radial coordinate r such that $r^2 = x^2 + y^2$, we can express $^2\delta(x,y)$ in terms of $\delta(r)$:

$$^2\delta(x,y) = \frac{\delta(r)}{\pi|r|}.$$

In three dimensions,

$$^3\delta(x,y,z) = \begin{cases} 0 & x^2+y^2+z^2 \neq 0 \\ \infty & x^2+y^2+z^2 = 0, \end{cases}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} {}^3\delta(x,y,z)\,dx\,dy\,dz = 1,$$

and

$$^3\delta(x,y,z) = \delta(x)\,\delta(y)\,\delta(z) = {}^2\delta(x,y)\,\delta(z).$$

In cylindrical coordinates $r^2 = x^2 + y^2$,

$$^3\delta(x,y,z) = \frac{\delta(r)\,\delta(z)}{\pi|r|},$$

and with $\rho^2 = x^2 + y^2 + z^2$,

$$^3\delta(x,y,z) = \frac{\delta(\rho)}{2\pi\rho^2}.$$

For describing arrays in two dimensions we have the bed-of-nails symbol $^2\mathrm{III}(x,y)$, illustrated in Fig. 5.12 and defined by

$$^2\mathrm{III}(x,y) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} {}^2\delta(x-m,\,y-n).$$

Figure 5.13 shows an approach to the discussion of its properties. It has the property

$$^2\mathrm{III}(x,y) = \mathrm{III}(x)\mathrm{III}(y)$$

and various extensions of the integral properties of $^2\delta(x,y)$ and $\mathrm{III}(x)$, for example,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,{}^2\mathrm{III}(x,y)\,dx\,dy = \sum_m\sum_n f(m,n).$$

It is doubly periodic,

$$^2\mathrm{III}(x+m,\,y+n) = {}^2\mathrm{III}(x,y) \qquad m,n \text{ integral,}$$

and

$$\frac{1}{|XY|}\,{}^2\mathrm{III}\left(\frac{x}{X},\frac{y}{Y}\right)$$

represents a doubly periodic array of two-dimensional unit impulses with period X in the x direction and Y in the y direction.

Tabulation at discrete intervals of two independent variables (two-dimensionally sampled data), and the coefficients of double Fourier series are handled through the relation

$$f(x,y)\,{}^2\mathrm{III}(x,y) = \sum_m \sum_n f(m,n)\,{}^2\delta(x-m,\,y-n).$$

Convolution with $^2\mathrm{III}(x,y)$ describes replication in two dimensions, such as one has in a two-dimensional array of identical antennas, and, as with III in one dimension, ^{2}III has the distinction of being its own two-dimensional Fourier transform (in the limit).

The scheme of multidimensional notation introduced here permits various self-explanatory extensions which are occasionally useful for compactness. Thus

$$\begin{aligned} {}^3\mathrm{III}(x,y,z) &= \mathrm{III}(x)\mathrm{III}(y)\mathrm{III}(z)\\ {}^2\Pi(x,y) &= \Pi(x)\Pi(y)\\ {}^2\mathrm{sinc}\,(x,y) &= \frac{\sin \pi x \sin \pi y}{\pi^2 xy}\\ {}^2\mathrm{II}(x,y) &= \mathrm{II}(x)\mathrm{II}(y). \end{aligned}$$

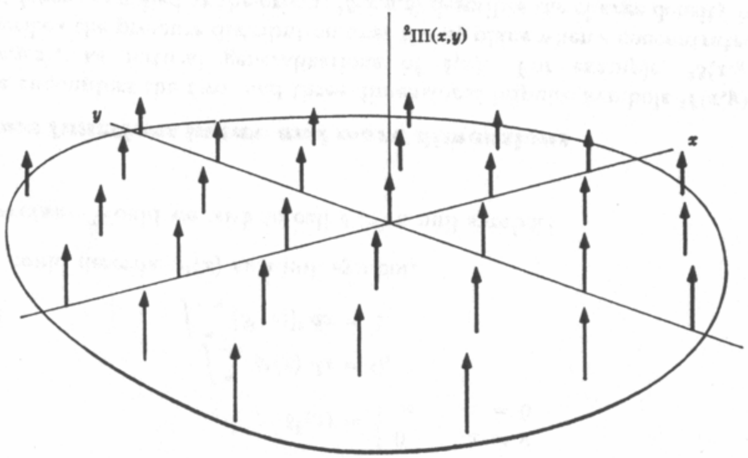


Fig. 5.12 The bed-of-nails symbol, $^2\mathrm{III}(x,y)$.

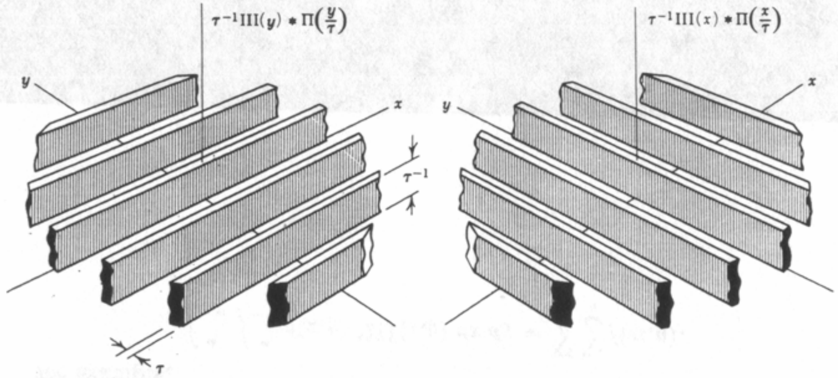


Fig. 5.13 Two functions whose product is suitable for discussing $^2\mathrm{III}(x,y)$.

Two other important two-dimensional distributions do not need new symbols. The row of spikes (see Fig. 5.14) is adequately expressed by $\mathrm{III}(x)\,\delta(y)$ and the grating by $\mathrm{III}(x)$. These two distributions form a two-dimensional Fourier transform pair and are suitable for discussing phenomena such as the diffraction of light by a row of pinholes or by a diffraction grating.

The concept of generalized function

As has been seen, a good deal of convenience attends the use of the impulse symbol $\delta(x)$ and other combinations of impulses such as $\mathrm{III}(x)$ and $\mathrm{II}(x)$. The word "symbol" has been used to call attention to the fact that these entities are not functions, but despite their apparent lack of status they

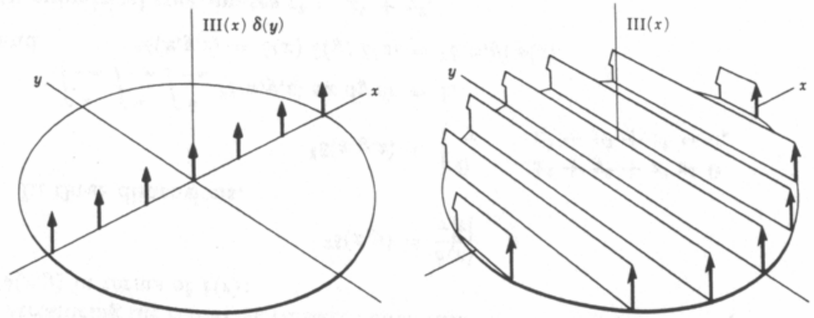


Fig. 5.14 The row of spikes (left) and grating (right).

have many uses, one of which is to provide derivatives for functions with simple discontinuities. Ordinarily we would say in such cases that the derivative does not exist, but the impulse symbol permits simple discontinuities to be accommodated.

The importance of dealing with discontinuous and impulsive behavior, even though it is nonphysical, was explained earlier in connection with the indispensable but nonphysical pure alternating and pure direct current.

Pure alternating current means an eternal harmonic variation, which cannot be generated. However, the response to a variation that is simple harmonic over a certain interval and zero outside that range can be made independent of the time of switching on, to a given precision, by waiting more than a certain length of time before observing the response. Since the details of time and manner of switching on are irrelevant, they might as well be relegated to the infinitely remote past, thus making it convenient to refer to the observable response as the response to pure alternating current.

Impulses are likewise impossible to generate physically, but the responses to different sufficiently brief but finite pulses can be made indistinguishable to an instrument of given finite temporal resolving power.

Since precision of measurement is known to be limited by the temporal and spectral resolution of the measuring instrument, the physical limitations referred to in the preceding paragraphs can be characterized as "finite resolution." The feature of finite resolving power invoked in the explanation contains the key to the mathematical interpretation of impulse-symbol notation: integrals containing the impulse symbol are to be interpreted as limits of a sequence of integrals in which the impulse is replaced by a sequence of unit-area rectangular pulses $\tau^{-1}\Pi(x/\tau)$. The limit of the sequence of integrals may exist, even though the rectangular pulses grow *without* limit.

Thus the statement

$$\int_{-\infty}^{\infty} \delta(x)f(x)\,dx = f(0)$$

is deemed to mean

$$\lim_{\tau\to 0} \int_{-\infty}^{\infty} \tau^{-1}\Pi\left(\frac{x}{\tau}\right)f(x)\,dx = f(0).$$

In this interpretation of the statement the integrals can exist, and the limit of the integrals as $\tau \to 0$ can exist. The physical situation to which this corresponds is a sequence of ever more compact stimuli producing responses which become indistinguishable under observation to a given precision, no matter how high that precision is.

A satisfactory mathematical formulation of the theory of impulses has

been evolved along these lines and is expounded in the books of Lighthill[1] and Friedman.[2] Lighthill credits Temple with simplifying the mathematical presentation; Temple[3] in turn credits the Polish mathematician Mikusiński[4] with introducing the presentation in terms of sequences in 1948. Schwartz's two volumes[5] on the theory of distributions unify "in one systematic theory a number of partial and special techniques proposed for the analytical interpretation of 'improper' or 'ideal' functions and symbolic methods."[6]

The idea of sequences was current in physical circles before 1948, however.[7]

The introduction of rectangular pulse sequences was not meant to imply that other pulse shapes are not equally valid. In fact the essence of the approach is that the detailed pulse shape is unimportant. The advantage of rectangular pulses is the purely practical one of facilitating integration. However, rectangular pulses do not lend themselves to discussing the derivative of an impulse. For that we need something smoother which does not itself have an impulsive derivative. Now for a general theory in which we wish to discuss derivatives of any order it is advantageous to have a pulse sequence such that derivatives of all orders exist. Schwartz and Temple introduce pulse shapes which have all derivatives and furthermore are zero outside a finite range; an example mentioned earlier in this chapter is[8]

$$\begin{cases} \tau^{-1}e^{-\tau^2/(\tau^2-x^2)} & |x| < \tau \\ 0 & |x| \geqslant \tau. \end{cases}$$

In actual fact one never inserts such a function explicitly into an integral; when it becomes necessary to integrate, a pulse shape with a *sufficient* number of derivatives is chosen. Often a rectangular pulse suffices.

Particularly well-behaved functions The term "generalized function" may be defined as follows. First we consider the class S of functions which possess derivatives of all orders at all points and which,

[1] M. J. Lighthill, "An Introduction to Fourier Analysis and Generalised Functions," Cambridge University Press, Cambridge, England, 1958.

[2] B. Friedman, "Principles and Techniques of Applied Mathematics," John Wiley & Sons, New York, 1956.

[3] G. Temple, Theories and Applications of Generalised Functions, *J. Lond. Math. Soc.*, vol. 28, p. 181, 1953.

[4] J. G.-Mikusiński, Sur la méthode de généralisation de Laurent Schwartz et sur la convergence faible, *Fundamenta Mathematicae*, vol. 35, p. 235, 1948.

[5] L. Schwartz, "Théorie des distributions," vols. 1 and 2, Herman & Cie, Paris, 1950 and 1951.

[6] Temple, *op. cit.*, p. 175.

[7] B. van der Pol, Discontinuous Phenomena in Radio Communication, *J. Inst. Elec. Engrs.*, vol. 81, p. 381, 1937.

[8] Schwartz, *op. cit.*, p. 22.

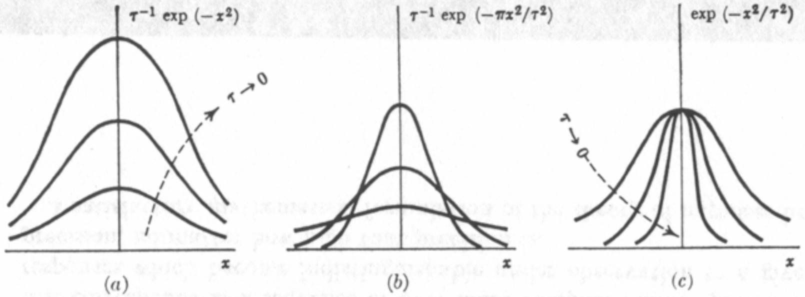


Fig. 5.15 Sequences of particularly well-behaved functions: (a) not regular; (b) regular; (c) exercise.

together with all the derivatives, die off at least as rapidly as $|x|^{-N}$ as $|x| \to \infty$, no matter how large N may be. We shall refer to members of the class S as particularly well-behaved functions. We note that the derivative and the Fourier transform of a particularly well-behaved function are also particularly well behaved. To prove the second of these statements let

$$\bar{F}(s) = \int_{-\infty}^{\infty} F(x)e^{-i2\pi sx}\,dx.$$

The conditions met by the particularly well-behaved function more than suffice to ensure that the Fourier integral exists. Differentiating p times, we have

$$\bar{F}^{(p)}(s) = \int_{-\infty}^{\infty} [(-i2\pi x)^p F(x)]e^{-i2\pi sx}\,dx,$$

and integrating by parts N times, we have

$$\begin{aligned} |\bar{F}^{(p)}(s)| &= \left| \frac{1}{(i2\pi s)^N} \int_{-\infty}^{\infty} \frac{d^N}{dx^N}[(-i2\pi x)^p F(x)]e^{-i2\pi sx}\,dx \right| \\ &\leqslant \frac{(2\pi)^{p-N}}{|s|^N} \int_{-\infty}^{\infty} \left| \frac{d^N}{dx^N}[x^p F(x)] \right| dx \\ &= O(|s|^{-N}); \end{aligned}$$

hence $\bar{F}(s)$ belongs to S.

Regular sequences Among sequences of particularly well-behaved functions we distinguish sequences $p_\tau(x)$ which lead to limits when multiplied by any other particularly well-behaved function $F(x)$ and integrated. Thus if

$$\lim_{\tau\to 0} \int_{-\infty}^{\infty} p_\tau(x)F(x)\,dx$$

exists, then we call $p_\tau(x)$ a *regular* sequence of particularly well-behaved functions.

A sequence of particularly well-behaved functions which is not regular (see Fig. 5.15) is

$$\tau^{-1}e^{-x^2}.$$

An example of a regular sequence is

$$\tau^{-1}e^{-\pi x^2/\tau^2}.$$

In this example each member of the sequence has unit area, but that is not essential; for instance, consider the regular sequence $(1 + \tau^{-1})\exp(-x^2/\tau^2)$.

Generalized functions A generalized function $p(x)$ is then taken to be defined by a regular sequence $p_\tau(x)$ of particularly well-behaved functions. In fact the generalized function *is* the regular sequence, and since the limit to which a regular sequence leads can be the same for more than one sequence, a generalized function is finally defined as the class of all regular sequences of particularly well-behaved functions equivalent to a given regular sequence. The symbol $p(x)$ thus represents an entity rather different from an ordinary function. It stands for a class of functions, and is itself not a function. Therefore, when we write it in a context where ordinary functions are customary, the meaning to be assigned must be stated. For example, we shall deem that

$$\int_{-\infty}^{\infty} p(x)F(x)\,dx,$$

where $p(x)$ is a generalized function and $F(x)$ is any particularly well-behaved function, shall mean

$$\lim_{\tau\to 0} \int_{-\infty}^{\infty} p_\tau(x)F(x)\,dx,$$

where $p_\tau(x)$ is any regular sequence of particularly well-behaved functions defining $p(x)$.

Two points may be noticed. First, the sequences $\tau^{-1}\Pi(x/\tau)$ and $\tau^{-1}\Lambda(x/\tau)$ do not define a generalized function in the present sense, for the members of these sequences do not possess derivatives of all orders at all points. Second, the function $F(x)$ on which the limiting process is tested has to be particularly well-behaved.

These highly restrictive conditions enable one to make a logical development at the cost of appearing to exclude the simple sequences and simple functions we ordinarily handle. We must bear in mind, however, that where differentiability is not in question we may fall back on rectangular pulses, and that where only first or second derivatives are required, Π^{*2} and Π^{*3} suffice. The requirement on asymptotic behavior is met by the rectangular pulse. On the other hand, the sequence $\tau^{-1}\operatorname{sinc}(x/\tau)$, which satisfies the requirement on differentiability, does not die away quickly enough to be a regular sequence in the strict sense; nevertheless, it is usable when it enters into a product with a function which is zero outside a finite interval.

The advantage of the Gaussian pulse as a special but simple case of a particularly well-behaved function has been exploited systematically by Lighthill,[9] whose line of development is followed.

The sequence $\exp(-\tau^2 x^2)$ defines a generalized function $I(x)$, for

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} e^{-\tau^2x^2}F(x)\,dx = \int_{-\infty}^{\infty} F(x)\,dx,$$

which integral exists. Hence we can make the following statement about $I(x)$:

$$\int_{-\infty}^{\infty} I(x)F(x)\,dx = \int_{-\infty}^{\infty} F(x)\,dx,$$

where $F(x)$ is any particularly well-behaved function.

The sequence $\tau^{-1}\exp(-\pi x^2/\tau^2)$ defines a generalized function, for

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} \tau^{-1}e^{-\pi x^2/\tau^2}F(x)\,dx$$

exists and is equal to $F(0)$, where $F(x)$ is any particularly well-behaved function. To prove this note that

$$\left|\int_{-\infty}^{\infty} \tau^{-1}e^{-\pi x^2/\tau^2}F(x)\,dx - F(0)\right| = \left|\int_{-\infty}^{\infty} \tau^{-1}e^{-\pi x^2/\tau^2}[F(x) - F(0)]\,dx\right|$$
$$\leqslant \max|F'(x)|\int_{-\infty}^{\infty}\tau^{-1}e^{-\pi x^2/\tau^2}|x|\,dx$$
$$= \frac{\tau}{\pi}\max|F'(x)|,$$

which approaches zero as $\tau \to 0$. The generalized function defined by this and equivalent sequences we call $\delta(x)$, and we can state immediately that

$$\int_{-\infty}^{\infty}\delta(x)F(x)\,dx = F(0),$$

where $F(x)$ is any particularly well-behaved function.

Algebra of generalized functions We have introduced one rule for handling the symbol standing for a generalized function, namely, that giving the meaning of

$$\int_{-\infty}^{\infty} p(x)F(x)\,dx.$$

Further rules are needed for handling the symbols for generalized functions where they appear in other algebraic situations.

Let $p(x)$ and $q(x)$ be two generalized functions, defined by the regular

[9] Lighthill, *op. cit.*

sequences $p_\tau(x)$ and $q_\tau(x)$, respectively. Now consider the sequence $p_\tau(x) + q_\tau(x)$. First, we note that it is a sequence of particularly well-behaved functions. Next, we see whether the sequence is a regular one, that is, whether

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty}[p_\tau(x) + q_\tau(x)]F(x)\,dx$$

exists, where $F(x)$ belongs to S. The integral splits into two terms, each of which has a limit, since $p_\tau(x)$ and $p_\tau(x)$ are by definition regular sequences. The sum of the two limits is the limit whose existence thus establishes that $p_\tau(x) + q_\tau(x)$ is a regular sequence that consequently defines a generalized function. This generalized function we would wish to assign as the denotation of

$$p(x) + q(x);$$

it remains only to verify that the result is the same irrespective of the choice of the defining sequences $p_\tau(x)$ and $q_\tau(x)$, and indeed we see that the defining sequences $p_\tau(x) + q_\tau(x)$ are equivalent, since the sum of the two limits is independent of the choice of $p_\tau(x)$ and $q_\tau(x)$.

We now have a meaning for the addition of generalized functions.

Let $p(x)$ be a generalized function defined by a sequence $p_\tau(x)$. From the formula for integration by parts,

$$\int_{-\infty}^{\infty} p'_\tau(x)F(x)\,dx = -\int_{-\infty}^{\infty} p_\tau(x)F'(x)\,dx,$$

where $F(x)$ is any particularly well-behaved function and so therefore is $F'(x)$. Since $F'(x)$ is a particularly well-behaved function, and since $p_\tau(x)$ is by definition a regular sequence, it follows that

$$-\lim_{\tau\to 0}\int_{-\infty}^{\infty} p_\tau(x)F'(x)\,dx$$

exists, hence

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} p'_\tau(x)F(x)\,dx$$

exists. Thus $p'_\tau(x)$ is a regular sequence of particularly well-behaved functions, and all such sequences are equivalent. To the generalized function so defined we assign the notation

$$p'(x).$$

This gives us a meaning for the derivative of a generalized function. Here is an example of a statement which can be made about the derivative $p'(x)$ of a generalized function $p(x)$:

$$\int_{-\infty}^{\infty} p'(x)F(x)\,dx = -\int_{-\infty}^{\infty} p(x)F'(x)\,dx.$$

Similarly, $\int_{-\infty}^{\infty} p^{(n)}(x)F(x)\,dx = (-1)^n \int_{-\infty}^{\infty} p(x)F^{(n)}(x)\,dx.$

Since by definition $F^{(n)}(x)$ exists, however large n may be, it follows that we have an interpretation for the nth derivative of a generalized function, for any n.

Differentiation of ordinary functions Generalized functions possess derivatives of all orders, and if an ordinary function could be regarded as a generalized function, then there would be a satisfactory basis for formulas such as

$$\frac{d}{dx}[H(x)] = \delta(x).$$

If $f(x)$ is an ordinary function and we form a sequence $f_\tau(x)$ such that

$$\lim_{\tau\to 0} \int_{-\infty}^{\infty} f_\tau(x)F(x)\,dx = \int_{-\infty}^{\infty} f(x)F(x)\,dx,$$

where $F(x)$ is any particularly well-behaved function, then the sequence defines a *generalized* function, which we may denote by the same symbol $f(x)$. The symbol $f(x)$ then has two meanings. We shall limit attention to functions $f(x)$ which as $|x| \to \infty$ behave as $|x|^{-N}$ for some value of N.

A suitable sequence $f_\tau(x)$ is given by

$$[\tau^{-1}e^{-\pi x^2/\tau^2}] * [f(x)e^{-\tau^2 x^2}].$$

With this enlargement of the notion of generalized functions we can embrace the unit step function $H(x)$ as a generalized function and assign meaning to its derivative $H'(x)$. Thus

$$\begin{aligned}\int_{-\infty}^{\infty} H'(x)F(x)\,dx &= -\int_{-\infty}^{\infty} H(x)F'(x)\,dx\\ &= -\int_{0}^{\infty} F'(x)\,dx\\ &= \int_{\infty}^{0} F'(x)\,dx\\ &= F(0),\end{aligned}$$

but $$\int_{-\infty}^{\infty} \delta(x)F(x)\,dx = F(0),$$

hence $$H'(x) = \delta(x).$$

The generalized function $\delta(x)$ is thus the derivative of the generalized function $H(x)$, and this is how we interpret formulas such as $H'(x) = \delta(x)$; we take the symbol for an ordinary function such as $H(x)$ to stand for the corresponding generalized function.

Problems

1 What is the even part of

$$\delta(x+3) + \delta(x+2) - \delta(x+1) + \tfrac{1}{2}\delta(x) + \delta(x-1) - \delta(x-2) - \delta(x-3)?$$

2 Attempting to clarify the meaning of $\delta(xy)$, a student gave the following explanation. "Where u is zero, $\delta(u)$ is infinite. Now xy is zero where $x = 0$ and where $y = 0$, therefore $\delta(xy)$ is infinite along the x and y axes. Hence $\delta(xy) = \delta(x) + \delta(y)$." Explain the fallacy in this argument, and show that

$$\delta(xy) = \frac{\delta(x) + \delta(y)}{(x^2 + y^2)^{\frac{1}{2}}}.$$

3 Show that

$$\text{II}(x) = \delta(2x^2 - \tfrac{1}{2})$$

and that $$\delta(x^2 - a^2) = \tfrac{1}{2}|a|^{-1}\{\delta(x-a) + \delta(x+a)\}.$$

4 Show that

$$\int_{-\infty}^{\infty} e^{-i2\pi xs}\,ds = \delta(x)$$

and that

$$\int_{-\infty}^{\infty} \delta(x)e^{i2\pi sx}\,dx = 1.$$

5 Show that

$$\delta(ax+b) = \frac{1}{|a|}\,\delta\left(x + \frac{a}{b}\right), \qquad a \neq 0.$$

6 If $f(x) = 0$ has roots x_n, show that

$$\delta[f(x)] = \sum_n \frac{\delta(x - x_n)}{|f'(x_n)|}$$

wherever $f'(x_n)$ exists and is not zero. Consider the ideas suggested by $\delta(x^3)$ and $\delta(\operatorname{sgn} x)$.

7 Show that

$$\pi\delta(\sin \pi x) = \text{III}(x)$$

and $$\delta(\sin x) = \pi^{-1}\,\text{III}\left(\frac{x}{\pi}\right).$$

8 Show that

$$\text{III}(x) + \text{III}(x - \tfrac{1}{2}) = 2\,\text{III}(2x) = \text{III}(x) * 4\text{II}(2x - \tfrac{1}{2}).$$

9 Show that

$$\text{III}(x)\Pi\left(\frac{x}{8}\right) = \text{III}(x)\Pi\left(\frac{x}{7}\right) + \frac{\text{II}(x/8)}{8}$$

and also that $$\text{III}(x)\Pi\left(\frac{x}{6\frac{1}{2}}\right) = \text{III}(x)\Pi\left(\frac{x}{7\frac{1}{2}}\right).$$

10 Can the following equation be correct?

$$x\,\delta(x-y) = y\,\delta(x-y).$$

11 Show that $\Lambda(x) * \sum_{-\infty}^{\infty} a_n\,\delta(x-n)$ is the polygon through the points (n,a_n).

12 Prove that

$$\delta'(-x) = -\delta'(x)$$
$$x\,\delta'(x) = -\delta(x).$$

Show also that

$$f(x)\,\delta'(x) = f(0)\,\delta'(x) - f'(0)\,\delta(x),$$

for example, by differentiating $f(x)\,\delta(x)$.

13 In attempting to show that $\delta'(x) = -\delta(x)/x$ a student presented the following argument. "A suitable sequence, as τ approaches zero, for defining $\delta(x)$ is $\tau/\pi(x^2+\tau^2)$. Therefore a suitable sequence for $\delta'(x)$ is the derivative

$$\frac{d}{dx}\frac{\tau}{\pi(x^2+\tau^2)} = \frac{-2\tau x}{\pi(x^2+\tau^2)^2} = \frac{-2x}{x^2+\tau^2}\,\frac{\tau}{\pi(x^2+\tau^2)}.$$

The second factor is the sequence for $\delta(x)$, and the first factor goes to $-2/x$ in the limit as τ approaches zero. Therefore $\delta'(x) = -2\delta(x)/x$." Explain the fallacy in this argument.

14 Show that

$$x^n\delta^{(n)}(x) = (-1)^n n!\,\delta(x)$$

and hence that $\quad x^2\,\delta''(x) = 2\delta(x)$

and $\quad x^3\,\delta''(x) = 0.$

15 The function $[x]$ is here defined as the mean of the greatest integer less than x and the greatest integer less than or equal to x. Show that

$$[x]' = \text{III}(x)$$

and also that $\quad \frac{d}{dx}\{[x]H(x)\} = \text{III}(x)H(x) - \tfrac{1}{2}\delta(x).$

(The common definition of $[x]$ as the greatest integer less than x is not fully suitable for the needs of this exercise; the two definitions differ by the null function which is equal to $\frac{1}{2}$ for integral values of x and is zero elsewhere.)

16 The sawtooth function $Sa(x)$ is defined by $Sa(x) = [x] - x + \frac{1}{2}$. Show that

$$Sa'(x) = \text{III}(x) - 1$$

and that $\quad \frac{d}{dx}[Sa(x)H(x)] = [\text{III}(x) - 1]H(x).$

17 Show that $\text{sgn}^2 x = 1 - \delta^0(x)$.

18 The Kronecker delta is defined by

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

Show that it may be expressed as a null function of $i - j$ as follows:

$$\delta_{ij} = \delta^0(i-j).$$

19 We wish to consider the suitability of a sequence of asymmetrical profiles, such as $\tau^{-1}\{\Lambda(x/\tau) + \frac{1}{2}\Lambda[(x-\tau)/\tau]\}$, for representing the impulse symbol. Discuss the sifting property that leads to a result of the form

$$\delta_a * f = \mu\,\delta_+ * f + \nu\,\delta_- * f,$$

where δ_a is a symbol based on the asymmetrical sequence, δ_+ is based on the sequence $\tau^{-1}\Pi[(x-\frac{1}{2}\tau)/\tau]$, and δ_- is based on the sequence $\tau^{-1}\Pi[(x+\frac{1}{2}\tau)/\tau]$ (τ positive).

20 Prove the relation ${}^2\delta(x,y) = \delta(r)/\pi|r|$.

21 Illustrate on an isometric projection the meaning you would assign to $\text{III}[(x^2+y^2)^{\frac{1}{2}}]$. How would you express something which on this diagram would have the appearance of equally spaced concentric rings of equal height?

22 The function $f_r(x)$ is formed from $f(x)$ by reversing it; that is, $f_r(x) = f(-x)$. Show that the operation of forming f_r from f can be expressed with the aid of the impulse symbol by

$$f \star \delta$$

and hence that $\quad (f\star\delta)\star\delta = f.$

23 Under what conditions could we say that $(f\star\delta)\star\delta = f\star(\delta\star\delta)$?

24 All the sequences $f(x,\tau)$ given on page 78 have the property that $f(0,\tau)$ increases without limit as $\tau \to 0$. Show that $\frac{1}{2}\tau^{-1}\Lambda[(x/\tau)-1] + \frac{1}{2}\tau^{-1}\Lambda[(x/\tau)+1]$ is an equivalent sequence which, however, possesses a limit of zero, as $\tau \to 0$, for all x. Show that $f(0,\tau)$, far from needing to approach ∞ as $\tau \to 0$, may indeed approach $-\infty$.

25 Show that

$$f(x)\,\delta''(x) = f(0)\,\delta''(x) - f'(0)\,\delta'(x) + f''(0)\,\delta(x)$$

and that in general

$$f(x)\,\delta^{(n)}(x) = f(0)\,\delta^{(n)}(x) - f'(0)\,\delta^{(n-1)}(x) + \cdots - f^{(n-1)}(0)\,\delta'(x) + f^{(n)}(0)\,\delta(x).$$

26 What can be said about the associativity of convolution in the case of $H(x) * \delta'(x) * H(-x)$?

Chapter 6 The basic theorems

A small number of theorems play a basic role in thinking with Fourier transforms. Most of them are familiar in one form or another, but here we collect them as simple mathematical properties of the Fourier transformation. Most of their derivations are quite simple, and their applicability to impulsive functions can readily be verified by consideration of sequences of rectangular or other suitable pulses. As a matter of interest, proofs based on the algebra of generalized functions as given in Chapter 5 are gathered for illustration at the end of this chapter.

The emphasis in this chapter, however, is on illustrating the *meaning* of the theorems and gaining familiarity with them. For this purpose a stock-in-trade of particular transform pairs is first provided so that the meaning of each theorem may be shown as it is encountered.

A few transforms for illustration

Six transform pairs for reference are listed below. They are all well known, and the integrals are evaluated in Chapter 7; we content ourselves at this point with asserting that the following integrals may be verified.

$$\int_{-\infty}^{\infty} e^{-\pi x^2} e^{-i2\pi xs}\,dx = e^{-\pi s^2} \quad \text{and} \quad \int_{-\infty}^{\infty} e^{-\pi s^2} e^{+i2\pi sx}\,ds = e^{-\pi x^2}$$

$$\int_{-\infty}^{\infty} \operatorname{sinc} x\, e^{-i2\pi xs}\,dx = \Pi(s) \quad \text{and} \quad \int_{-\infty}^{\infty} \Pi(s) e^{+i2\pi sx}\,ds = \operatorname{sinc} x$$

$$\int_{-\infty}^{\infty} \operatorname{sinc}^2 x\, e^{-i2\pi xs}\,dx = \Lambda(s) \quad \text{and} \quad \int_{-\infty}^{\infty} \Lambda(s) e^{+i2\pi sx}\,ds = \operatorname{sinc}^2 x$$

Thus the transform of the Gaussian function is the same Gaussian function, the transform of the sinc function is the unit rectangle function, and

the transform of the sinc² function is the triangle function of unit height and area.

These formulas are illustrated as the first three transform pairs in Fig. 6.1. Note that item (2) of the figure, which says that $\Pi(s)$ is the transform of sinc x, could be supplemented by a second figure, with left and right graphs interchanged, which would say that sinc s is the transform of $\Pi(x)$. A consequence of the reciprocal property of the Fourier transformation, this extra figure would appear redundant. However, the statement

$$\Pi(s) = \int_{-\infty}^{\infty} \operatorname{sinc} x\, e^{-i2\pi xs}\,dx$$

has quite a different character from

$$\operatorname{sinc} s = \int_{-\infty}^{\infty} \Pi(x) e^{-i2\pi xs}\,dx.$$

The first statement tells us that the integral of the product of certain rather ordinary functions is equal to unity for absolute values of the constant s less than $\frac{1}{2}$. Whether s is equal to say 0.3 or 0.35, the value of the integral is unchanged. However, if $|s|$ exceeds $\frac{1}{2}$, the situation changes abruptly, because the integral now comes to nothing and continues to do so, regardless of the precise value of s. Thus

$$\int_{-\infty}^{\infty} \frac{\sin \pi x}{\pi x} e^{-i2\pi xs}\,dx = \begin{cases} 1 & |s| < \frac{1}{2} \\ 0 & |s| > \frac{1}{2}. \end{cases}$$

This rather curious behavior is typical of many situations where the Fourier integral connects ordinary, continuous, and differentiable functions, on the one hand, with awkward, abrupt functions requiring piecewise definition, on the other. The second statement may be rewritten

$$\frac{\sin \pi s}{\pi s} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-i2\pi xs}\,dx.$$

Here an elementary definite integral of the exponential function is equal to an ordinary function of the parameter s. Thus the direct and inverse transforms express different things. Two separate and distinct physical meanings will later be seen to be associated with each transform pair.

Three further transforms required for illustrating the basic theorems are transform pairs in the limiting sense discussed earlier.

Taking the result for the Gaussian function, and making a simple substitution of variables, we have[1]

$$\int_{-\infty}^{\infty} e^{-\pi(ax)^2} e^{-i2\pi sx}\,dx = |a|^{-1} e^{-\pi(s/a)^2}.$$

[1] In this formula the absolute value of a is used in order to counteract the sign reversal associated with the interchange of the limits of integration when a is negative.

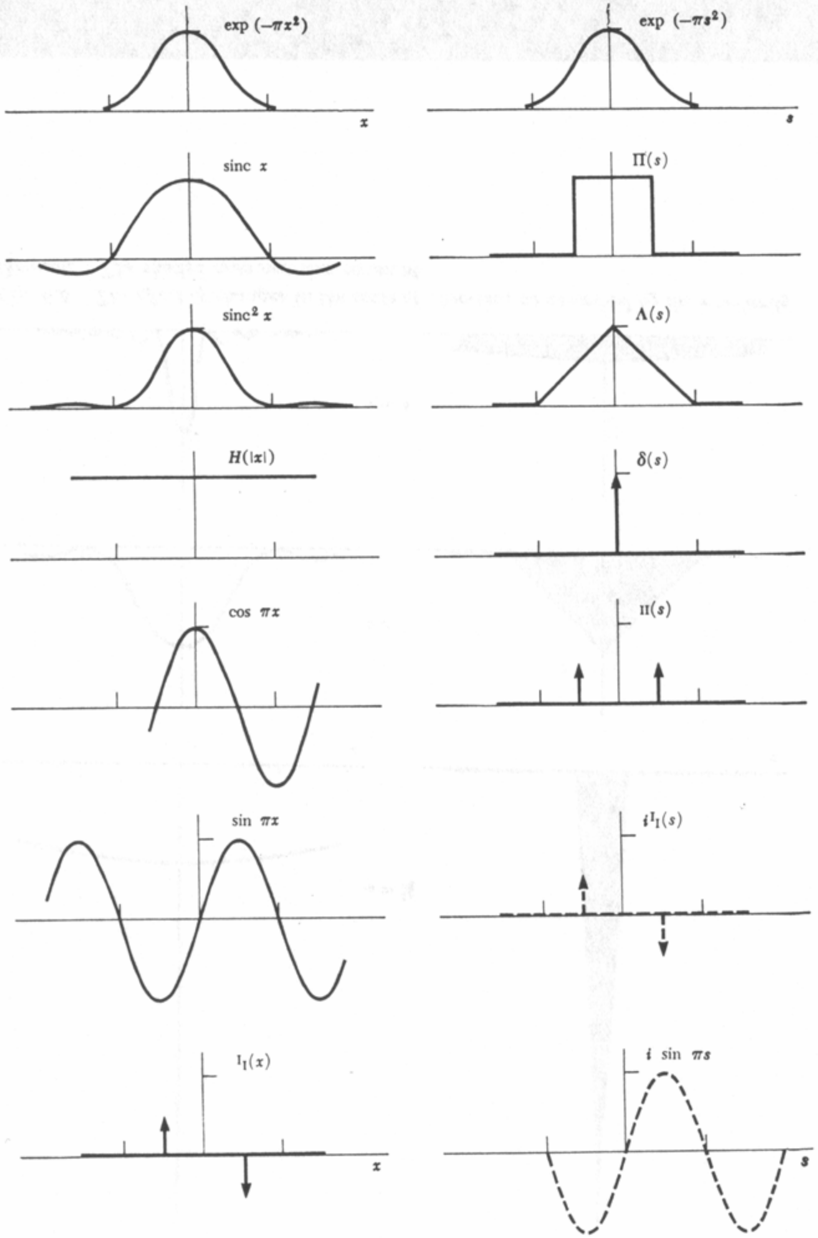


Fig. 6.1 *Some Fourier transform pairs for reference.*

As $a \to 0$, the right-hand side represents a defining sequence for $\delta(s)$; the left-hand side is the Fourier transform of what in the limit is unity. It follows that

1 is the Fourier transform in the limit of $\delta(s)$.

The remaining two examples come from the verifiable relation

$$\int_{-\infty}^{\infty} e^{i\pi x} e^{-(ax)^2} e^{-i2\pi xs}\,dx = |a|^{-1} e^{-[(s-\frac{1}{2})/a]^2},$$

whence

$e^{i\pi x}$ is the Fourier transform in the limit of $\delta(s - \frac{1}{2})$;

or, splitting the left-hand side into real and imaginary parts and the right-hand side into even and odd parts, $\cos \pi x$ is the Fourier transform in the limit of

$$\tfrac{1}{2}\delta(s + \tfrac{1}{2}) + \tfrac{1}{2}\delta(s - \tfrac{1}{2}) = \mathrm{II}(s)$$

and $i \sin \pi x$ is the minus-i Fourier transform in the limit of

$$-\tfrac{1}{2}\delta(s + \tfrac{1}{2}) + \tfrac{1}{2}\delta(s - \tfrac{1}{2}) = -{}^{\mathrm{I}}\mathrm{I}_{\mathrm{I}}(s).$$

Summarizing the examples,

$$\begin{aligned}
e^{-\pi x^2} &\supset e^{-\pi s^2}\\
\operatorname{sinc} x &\supset \Pi(s)\\
\operatorname{sinc}^2 x &\supset \Lambda(s)\\
1 &\supset \delta(s)\\
\cos \pi x &\supset \mathrm{II}(s) \equiv \tfrac{1}{2}\delta(s+\tfrac{1}{2}) + \tfrac{1}{2}\delta(s-\tfrac{1}{2})\\
\sin \pi x &\supset i\,{}^{\mathrm{I}}\mathrm{I}_{\mathrm{I}}(s) \equiv \tfrac{1}{2}i\delta(s+\tfrac{1}{2}) - \tfrac{1}{2}i\delta(s-\tfrac{1}{2})\\
{}^{\mathrm{I}}\mathrm{I}_{\mathrm{I}}(x) &\supset i \sin \pi s.
\end{aligned}$$

All the transform pairs chosen for illustration have physical interpretations, which will be brought out later. Many properties appear among the transform pairs chosen for reference, including discontinuity, impulsiveness, limited extent, nonnegativeness, and oddness. The only examples exhibiting complex or nonsymmetrical properties are

$$e^{i\pi x} \supset \delta(s - \tfrac{1}{2})$$

and

$$\delta(x - \tfrac{1}{2}) \supset e^{-i\pi s}.$$

Similarity theorem

If $f(x)$ has the Fourier transform $F(s)$, then $f(ax)$ has the Fourier transform $|a|^{-1}F(s/a)$.

Derivation:

$$\int_{-\infty}^{\infty} f(ax)e^{-i2\pi xs}\,dx = \frac{1}{|a|}\int_{-\infty}^{\infty} f(ax)e^{-i2\pi(ax)(s/a)}\,d(ax)$$
$$= \frac{1}{|a|}F\left(\frac{s}{a}\right).$$

This theorem is well known in its application to waveforms and spectra, where compression of the time scale corresponds to expansion of the frequency scale. However, as one member of the transform pair expands

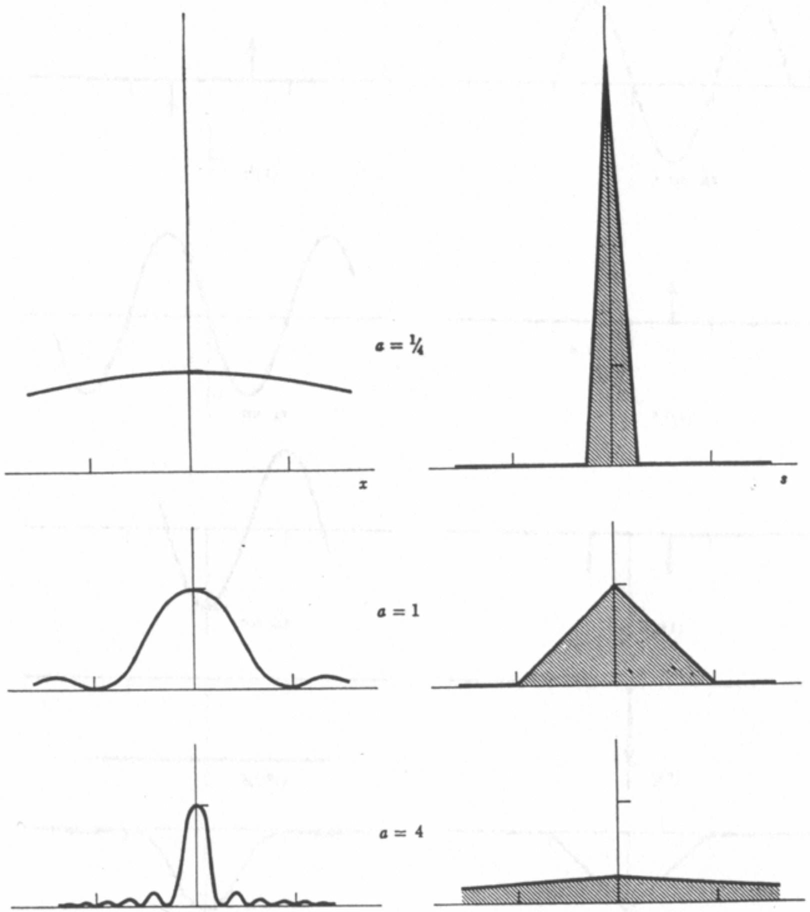


Fig. 6.2 The effect of changes in the scale of abscissas as described by the similarity theorem. The shaded area remains constant.

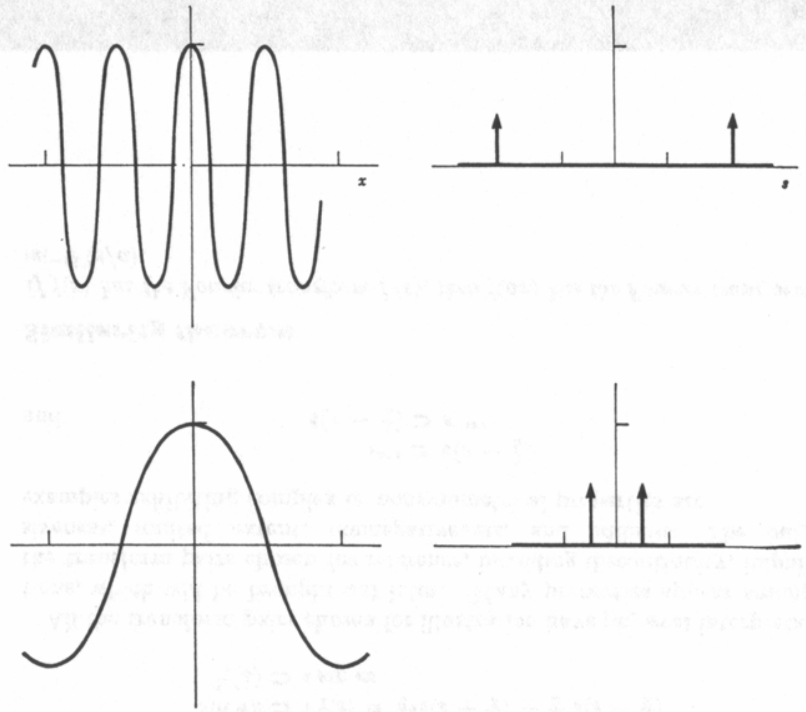


Fig. 6.3 Expansion of a cosinusoid and corresponding shifts in its spectrum.

horizontally, the other not only contracts horizontally but also grows vertically in such a way as to keep constant the area beneath it, as shown in Fig. 6.2.

A special case of interest arises with periodic functions and impulses. As Fig. 6.3 shows, expansion of a cosinusoid leads simply to shifts of the impulses constituting the transform. This is not simply a compression of the scale of s, for that would entail a reduction in strength of the impulses.

In a more symmetrical version of this theorem,

If $f(x)$ has the Fourier transform $F(s)$ then $|a|^{\frac{1}{2}}f(ax)$ has the Fourier transform $|b|^{\frac{1}{2}}F(bs)$, where $b = a^{-1}$.

Then, as each function expands or contracts it also shrinks or grows vertically (see Fig. 6.4) to compensate (in such a way that the integral of its square is maintained constant, as will be seen later from the power theorem).

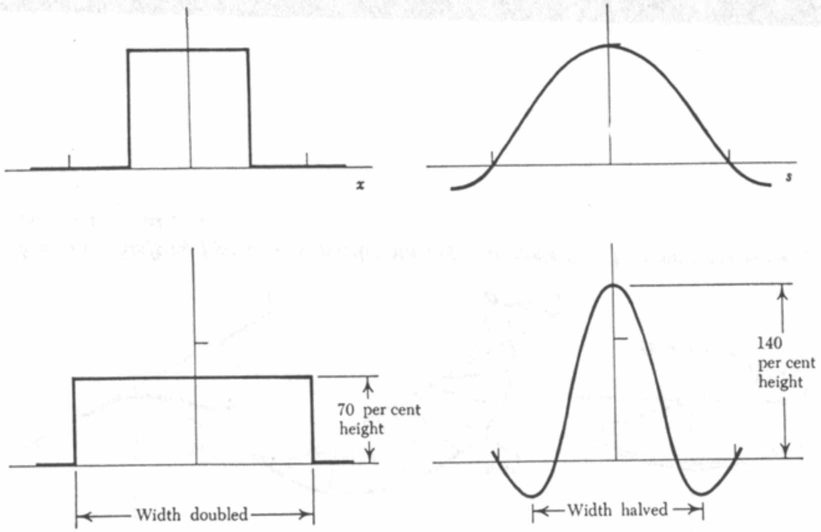


Fig. 6.4 A symmetrical version of the similarity theorem.

Addition theorem

If $f(x)$ and $g(x)$ have the Fourier transforms $F(s)$ and $G(s)$, respectively, then $f(x) + g(x)$ has the Fourier transform $F(s) + G(s)$.

Derivation:

$$\int_{-\infty}^{\infty} [f(x) + g(x)]e^{-i2\pi xs}\,dx = \int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx + \int_{-\infty}^{\infty} g(x)e^{-i2\pi xs}\,dx$$
$$= F(s) + G(s).$$

This theorem, which is illustrated by an example in Fig. 6.5, reflects the suitability of the Fourier transform for dealing with linear problems. A corollary is that $af(x)$ has the transform $aF(s)$, where a is a constant.

Shift theorem

If $f(x)$ has the Fourier transform $F(s)$, then $f(x - a)$ has the Fourier transform $e^{-2\pi ias}F(s)$.

Derivation:

$$\int_{-\infty}^{\infty} f(x - a)e^{-i2\pi xs}\,dx = \int_{-\infty}^{\infty} f(x - a)e^{-i2\pi(x-a)s}e^{-i2\pi as}\,d(x - a)$$
$$= e^{-i2\pi as}F(s).$$

If a given function is shifted in the positive direction by an amount a, no Fourier component changes in amplitude; it is therefore to be expected that the changes in its Fourier transform will be confined to phase changes. According to the theorem, each component is delayed in phase by an amount proportional to s; that is, the higher the frequency, the greater the change in phase angle. This occurs because the absolute shift a occupies a greater fraction of the period s^{-1} of a harmonic component in proportion to its frequency. Hence the phase delay is a/s^{-1} cycles or $2\pi as$ radians. The constant of proportionality describing the linear change of phase with s is $2\pi a$, the rate of change of phase with frequency being greater as the shift a is greater.

The shift theorem is one of those which are self-evident in a chosen physical embodiment. Consider parallel light falling normally on an aperture. To shift the diffracted beam through a small angle, one changes the angle of incidence by that amount. But this is simply a way of causing the phase of the illumination to change linearly across the aper-

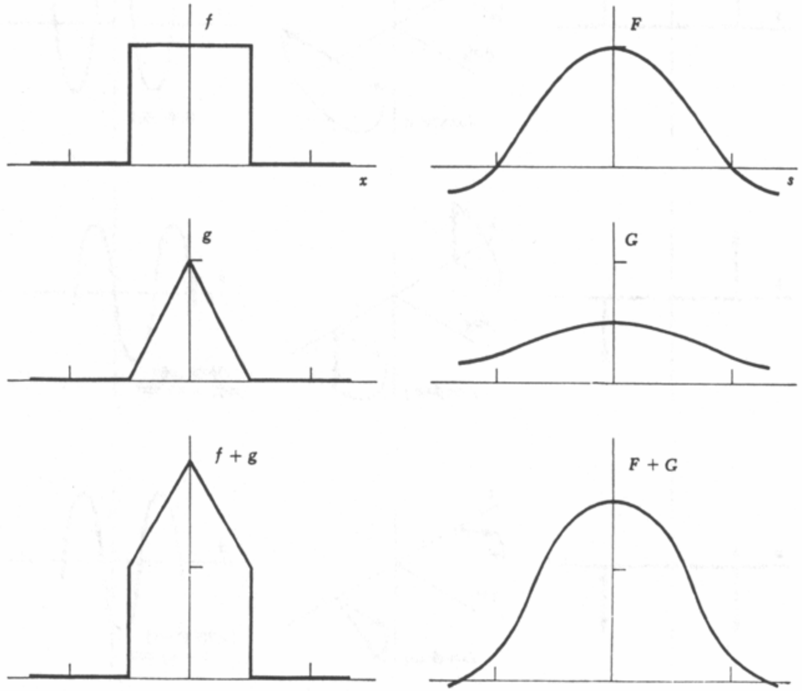


Fig. 6.5 The addition theorem $f + g \supset F + G$.

ture; another way is to insert a prism. These well-understood procedures for shifting the direction of a light beam are shown in Chapter 13 to exemplify the shift theorem.

In the example of Fig. 6.6, a function $f(x)$ is shown whose transform $F(s)$ is real. A shifted function $f(x - \frac{1}{4})$ has a transform which is derivable by subjecting $F(s)$ to a uniform twist of $\pi/2$ per unit of s. The figure attempts to show that the plane containing $F(s)$ has been deformed into a helicoid. The practical difficulties of representing a complex function of s in a three-dimensional plot are overcome by showing the modulus and phase of $F(s)$ separately; however, the three-dimensional diagram often gives a better insight.

The second example (see Fig. 6.7) shows familiar results for the cosine and sine functions and for the intermediate cases which arise as the cosine slides along the axis of x. In this case the helicoidal surface is not shown. An alternative representation in terms of real and imaginary parts is given, incorporating the convention introduced earlier of showing the imaginary part by a broken line. A small shift evidently leaves the real

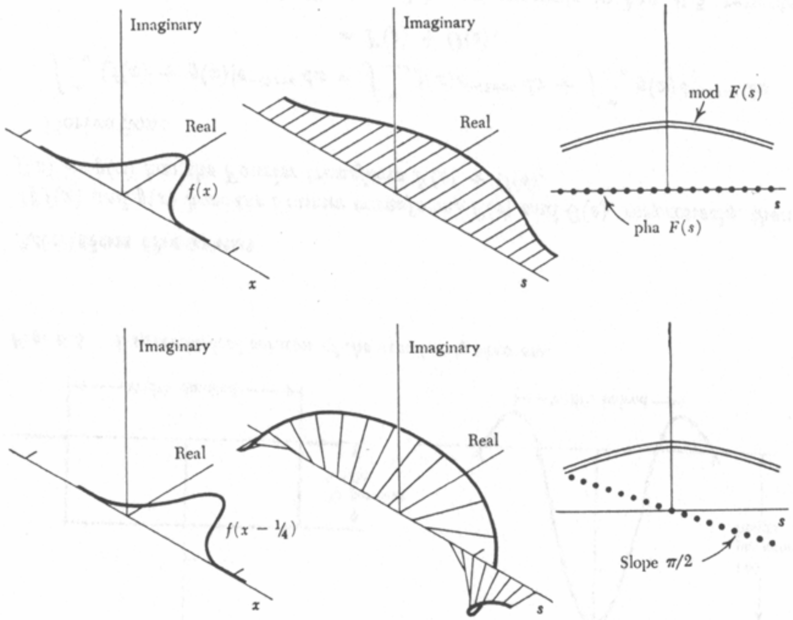


Fig. 6.6 Shifting $f(x)$ by one quarter unit of x subjects $F(s)$ to a uniform twist of 90 deg per unit of s.

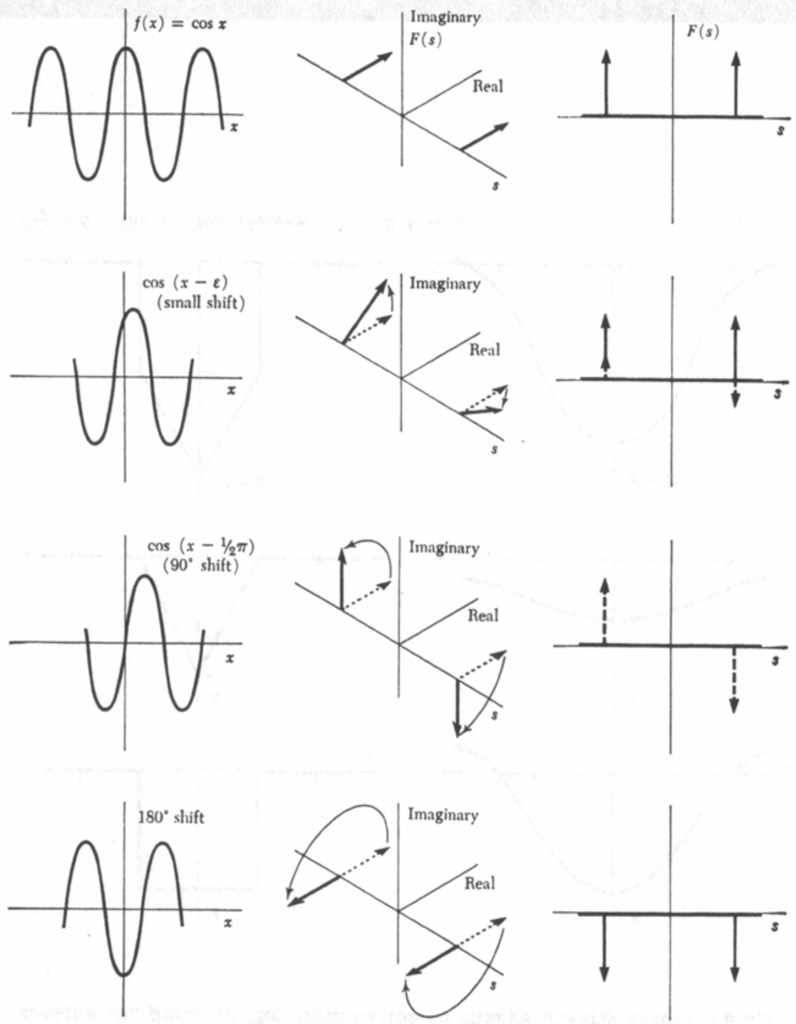


Fig. 6.7 The effect of shifting a cosinusoid.

part of the transform almost intact but introduces an odd imaginary part. With further shift the imaginary part increases until at a shift of $\pi/2$ there is no real part left. Then the real part reappears with opposite sign until at a shift of π both components have undergone a full reversal of phase.

Modulation theorem

If $f(x)$ has the Fourier transform $F(s)$, then $f(x)\cos\omega x$ has the Fourier transform $\frac{1}{2}F(s-\omega/2\pi)+\frac{1}{2}F(s+\omega/2\pi)$.

Derivation:

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\cos\omega x\, e^{-i2\pi xs}\,dx &= \tfrac{1}{2}\int_{-\infty}^{\infty} f(x)e^{i\omega x}e^{-i2\pi xs}\,dx\\ &\quad+\tfrac{1}{2}\int_{-\infty}^{\infty} f(x)e^{-i\omega x}e^{-i2\pi xs}\,dx\\ &= \tfrac{1}{2}\int_{-\infty}^{\infty} f(x)e^{-i2\pi(s-\omega/2\pi)x}\,dx\\ &\quad+\tfrac{1}{2}\int_{-\infty}^{\infty} f(x)e^{-i2\pi(s+\omega/2\pi)x}\,dx\\ &= \tfrac{1}{2}F(s-\omega/2\pi)+\tfrac{1}{2}F(s+\omega/2\pi).\end{aligned}$$

The new transform will be recognized as the convolution of $F(s)$ with $\frac{1}{2}\delta(s+\omega/2\pi)+\frac{1}{2}\delta(s-\omega/2\pi)=(\pi/\omega)\text{II}(\pi s/\omega)$. This is a special case of the convolution theorem, but it is important enough to merit special mention. It is well known in radio and television, where a harmonic carrier wave is modulated by an envelope. The spectrum of the envelope is separated into two parts, each of half the original strength. These two replicas of the original are then shifted along the s axis by amounts $\pm\omega/2\pi$, as shown in Fig. 6.8.

Convolution theorem

As stated earlier, the convolution of two functions f and g is another function h defined by the integral

$$h(x)=\int_{-\infty}^{\infty} f(u)g(x-u)\,du.$$

A great deal is implied by this expression. For instance, $h(x)$ is a linear functional of $f(x)$; that is, $h(x_1)$ is a linear sum of values of $f(x)$, duly weighted as described by $g(x)$. However, it is not the most general linear functional; it is the particular kind for which any other value $h(x_2)$ is given by a linear combination of values of $f(x)$ weighted in the *same* way. Another way of conveying this special property of convolution is to say that a shift of $f(x)$ along the x axis results simply in an equal shift of $h(x)$; that is, if $h(x)=f(x)*g(x)$, then

$$f(x-a)*g(x)=h(x-a).$$

Suppose that a train is slowly crossing a bridge. The load at the point x is $f(x)$, and the deflection at x is $h(x)$. Since the structural members are

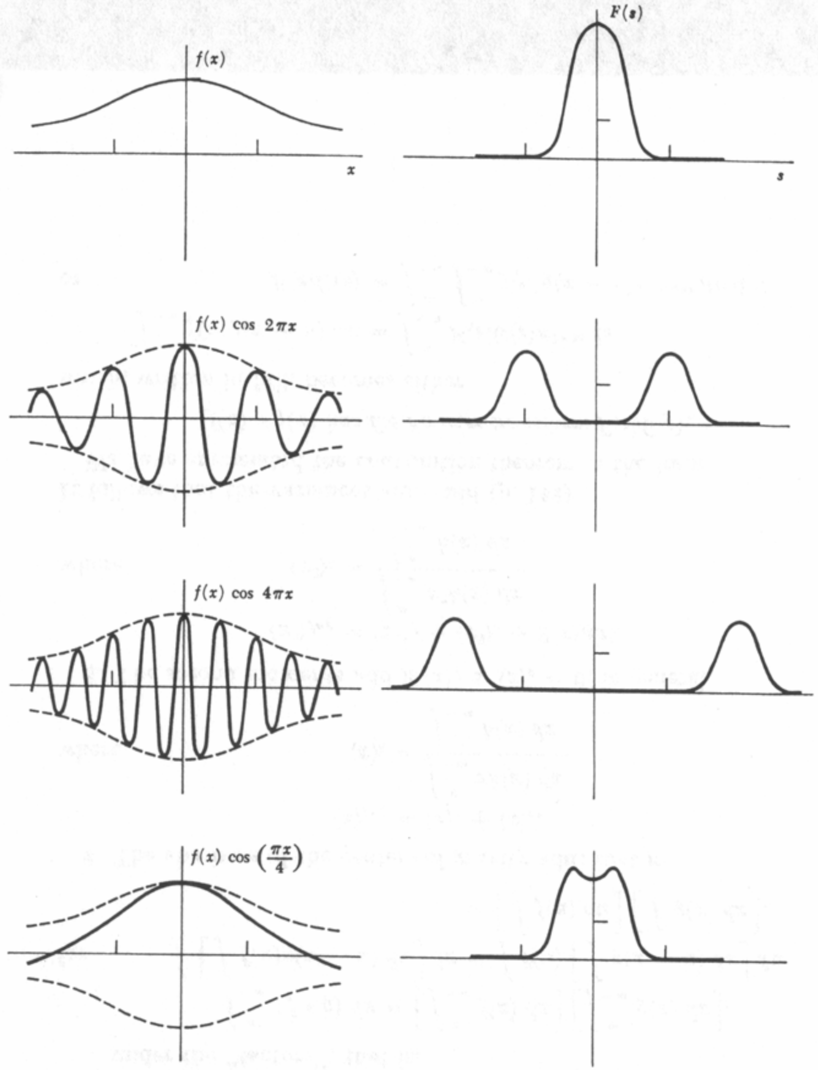


Fig. 6.8 An envelope function $f(x)$ multiplied by cosinusoids of various frequencies, with the corresponding spectra.

not being pushed beyond the regime where stress is proportional to strain, it follows that the deflection at x_1 is a duly weighted linear combination of values of the load distribution $f(x)$. But as the train moves on, the deflection pattern does not move on with it unchanged; it is not expressible as a convolution integral. All that can be said in this case is that $h(x)$ is a linear functional of $f(x)$; that is,

$$h(x) = \int_{-\infty}^{\infty} f(u)g(x,u)\,du.$$

It is the property of *linearity* combined with *x-shift invariance* which makes Fourier analysis so useful; as shown in Chapter 9, this is the condition that simple harmonic inputs produce simple harmonic outputs with frequency unaltered.

If the well-known and widespread advantages of Fourier analysis are concomitant with the incidence of convolution, one may expect in the transform domain a simple counterpart of convolution in the function domain. This counterpart is expressed in the following theorem.

If $f(x)$ has the Fourier transform $F(s)$ and $g(x)$ has the Fourier transform $G(s)$, then $f(x) * g(x)$ has the Fourier transform $F(s)G(s)$; that is, convolution of two functions means multiplication of their transforms.

Derivation:

$$\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x')g(x-x')\,dx'\right]e^{-i2\pi xs}\,dx$$
$$= \int_{-\infty}^{\infty} f(x')\left[\int_{-\infty}^{\infty} g(x-x')e^{-i2\pi xs}\,dx\right]dx'$$
$$= \int_{-\infty}^{\infty} f(x')e^{-i2\pi x's}G(s)\,dx'$$
$$= F(s)G(s).$$

Using bars to denote Fourier transforms, we can give compact statements of the theorem and its converse. Thus

$$\overline{f*g} = \bar f\bar g,$$
$$\overline{fg} = \bar f * \bar g.$$

Equivalent statements are

$$\overline{\bar f\bar g} = f*g$$
$$\overline{\bar f * \bar g} = fg.$$

We have stated earlier that

$$f*g = g*f \qquad \text{(commutative)}$$
$$f*(g*h) = (f*g)*h \qquad \text{(associative)}$$
$$f*(g+h) = f*g+f*h \qquad \text{(distributive).}$$

Further formulas are

$$\overline{f*g*h} = \bar f\bar g\bar h,$$
$$\overline{f*(gh)} = \bar f(\bar g * \bar h).$$

This powerful theorem and its converse play an important role in transforming a function which can be recognized as the convolution of two others or as the product of two others.

The following are statements in words of some of the above equations.

1. The transform of a convolution is the product of the transforms.
2. The transform of a product is the convolution of the transforms.
3. The convolution of two functions is the transform of the product of their transforms.
4. The product of two functions is the transform of the convolution of their transforms.

Three valuable properties often used for checking are the following.

1. The area under a convolution is equal to the product of the areas under the "factors"; that is,

$$\int_{-\infty}^{\infty}(f*g)\,dx = \left[\int_{-\infty}^{\infty} f(x)\,dx\right]\left[\int_{-\infty}^{\infty} g(x)\,dx\right],$$

for

$$\int\left[\int f(u)g(x-u)\,du\right]dx = \int f(u)\left[\int g(x-u)\,dx\right]du$$
$$= \left[\int f(u)\,du\right]\left[\int g(x)\,dx\right].$$

2. The abscissas of the centers of gravity add; that is,

$$\langle x\rangle_{f*g} = \langle x\rangle_f + \langle x\rangle_g,$$

where

$$\langle x\rangle_h = \frac{\int_{-\infty}^{\infty} xh(x)\,dx}{\int_{-\infty}^{\infty} h(x)\,dx}.$$

3. The second moments add if $\langle x\rangle_f$ or $\langle x\rangle_g = 0$; in general,

$$\langle x^2\rangle_{f*g} = \langle x^2\rangle_f + \langle x^2\rangle_g + 2\langle x\rangle_f\langle x\rangle_g,$$

where

$$\langle x^2\rangle_h = \frac{\int_{-\infty}^{\infty} x^2h(x)\,dx}{\int_{-\infty}^{\infty} h(x)\,dx}.$$

It follows that the variances must add (p. 142).

We have enunciated the convolution theorem in the form

$f(x) * g(x)$ has the Fourier transform $F(s)G(s)$,

which, written in full, becomes either

$$\int_{-\infty}^{\infty} f(u)g(x-u)\,du = \int_{-\infty}^{\infty} F(s)G(s)e^{i2\pi xs}\,ds$$

or

$$F(s)G(s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u)g(x-u)e^{-i2\pi xs}\,du\,dx.$$

There are no fewer than 20 versions in which the convolution theorem is constantly needed, when we allow for complex conjugates and sign reversals of the variables. The 10 abbreviated forms are

$$\begin{aligned}
f * g &\supset FG \\
f * g(-) &\supset FG(-) \\
f(-) * g(-) &\supset F(-)G(-) \\
f * g^*(-) &\supset FG^* \\
f * g^* &\supset FG^*(-) \\
f(-) * g^*(-) &\supset F(-)G^* \\
f(-) * g^* &\supset F(-)G^*(-) \\
f^*(-) * g^*(-) &\supset F^*G^* \\
f^*(-) * g^* &\supset F^*G^*(-) \\
f^* * g^* &\supset F^*(-)G^*(-)
\end{aligned}$$

The self-convolution formulas are

$$\begin{aligned}
f * f &\supset F^2 \\
f(-) * f(-) &\supset [F(-)]^2 \\
f^*(-) * f^*(-) &\supset [F^*]^2 \\
f^* * f^* &\supset [F^*(-)]^2
\end{aligned}$$

and for autocorrelation we have

$$\begin{aligned}
f \star f &\supset FF(-) \\
f(-) \star f(-) &\supset FF(-) \\
f^*(-) \star f^*(-) &\supset F^*F^*(-) \\
f^* \star f^* &\supset F^*F^*(-)
\end{aligned}$$

Rayleigh's theorem

The integral of the squared modulus of a function is equal to the integral of the squared modulus of its spectrum; that is,

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx = \int_{-\infty}^{\infty} |F(s)|^2\, ds.$$

Derivation:

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)f^*(x)\, dx &= \int_{-\infty}^{\infty} f(x)f^*(x)e^{-i2\pi xs'}\, dx && s' = 0 \\
&= F(s') * F^*(-s') && s' = 0 \\
&= \int_{-\infty}^{\infty} F(s)F^*(s - s')\, ds && s' = 0 \\
&= \int_{-\infty}^{\infty} F(s)F^*(s)\, ds.
\end{aligned}$$

This theorem, which corresponds to Parseval's theorem for Fourier series, was first used by Rayleigh[1] in his study of black-body radiation

[1] Lord Rayleigh, On the Character of the Complete Radiation at a Given Temperature, *Phil. Mag.*, series 5, vol. 27, 1889; "Scientific Papers," Cambridge University Press, Cambridge, England, 1902, and Dover Publications, New York, 1964, vol. 3, p. 273.

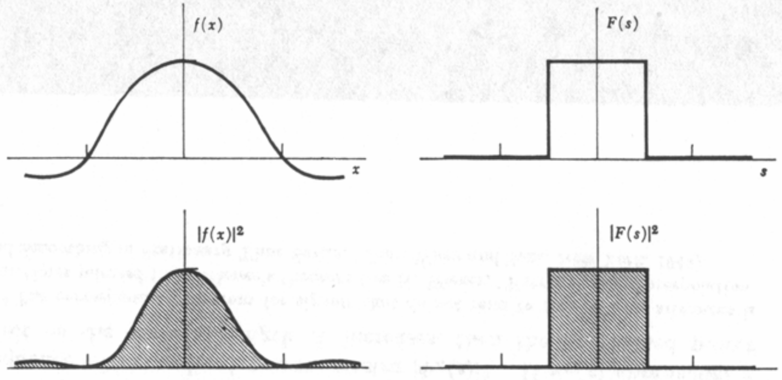


Fig. 6.9 Rayleigh's theorem: the shaded areas are equal.

In this, as in many other connections, each integral represents the amount of energy in a system, one integral being taken over all values of a coordinate, the other over all spectral components (see Fig. 6.9).

The theorem is sometimes referred to in mathematical circles as Plancherel's theorem,[2] after M. Plancherel, who in 1910 established conditions under which the theorem is true. The theorem is true if both the integrals exist. More recently, it has been shown by Carleman (see Bibliography, Chapter 2) that the theorem is true if one of the integrals exists. Rayleigh simply assumed in his derivation that the integrals existed.

Power theorem

$$\int_{-\infty}^{\infty} f(x)g^*(x)\, dx = \int_{-\infty}^{\infty} F(s)G^*(s)\, ds.$$

Derivation: The proof is as for Rayleigh's theorem when f^* is replaced by g^* and F^* by G^*. The following version illustrates a compact notation which is useful in its place.

$$\int fg^*\, dx = \overline{fg^*}\Big|_0 = \bar{f} * \overline{g^*}\Big|_0 = F * G^*(-)\Big|_0 = \int FG^*\, ds.$$

In many physical interpretations, each side of this equation represents energy or power (see Fig. 6.10), two different approaches being used to evaluate the energy or power. In one approach the instantaneous or local power or energy is evaluated as the product of a pair of canonically conjugate variables (electric and magnetic fields, voltage and current, force and velocity) integrated over time or space. In the second approach

[2] E. C. Titchmarsh, A Contribution to the Theory of Fourier Transforms, *Proc. Lond. Math. Soc.*, vol. 23, p. 279, 1924.

the temporal or spatial spectral components are multiplied and integrated over the whole spectrum.

It very often happens that both f and g are real quantities, as in the three examples cited. Then F and G may be complex, and

$$\begin{aligned} FG^* &= (\text{Re } F + i \text{ Im } F)(\text{Re } G - i \text{ Im } G) \\ &= (\text{Re } F)(\text{Re } G) + (\text{Im } F)(\text{Im } G) + \text{odd terms.} \end{aligned}$$

Inspection shows that the final terms are odd, for F and G are hermitian; that is, their real parts are even and imaginary parts odd. The odd terms do not contribute to the infinite integral. Hence for real f and g

$$\int_{-\infty}^{\infty} fg \, dx = \int_{-\infty}^{\infty} FG^* \, ds = \int_{-\infty}^{\infty} [(\text{Re } F)(\text{Re } G) + (\text{Im } F)(\text{Im } G)] \, ds.$$

This situation is illustrated in Fig. 6.11.

Exercise Show that, provided f and g are real, an alternative version of the power theorem is

$$\int_{-\infty}^{\infty} f(x)g(-x) \, dx = \int_{-\infty}^{\infty} F(s)G(s) \, ds.$$

By putting $g(x) = f(x)$ we obtain Rayleigh's theorem, which is thus appropriate to physical systems where, f/g and F/G (often interpretable as impedance in its general sense) being constant, energy or power may be

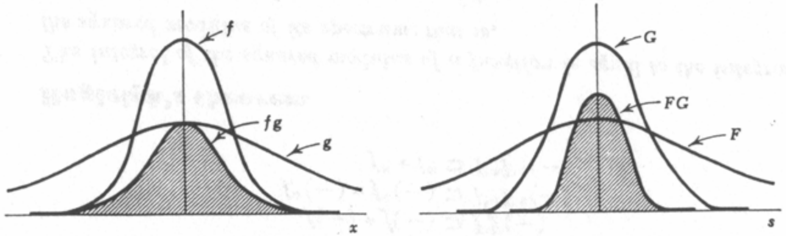


Fig. 6.10 The power theorem: the shaded areas are equal. In this example f and g are real and even.

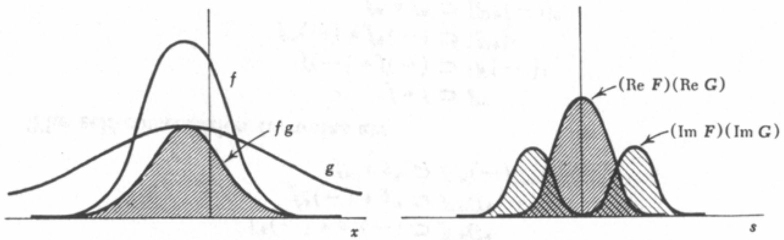


Fig. 6.11 The power theorem for f and g real, F and G complex. The shaded area on the left equals the sum of the shaded areas on the right.

expressed as the square of one variable alone. The theorem does not have a distinctive name of its own; some authors refer to it as Parseval's theorem, which is the well-established name of a theorem in the theory of Fourier series (Chapter 10).

Autocorrelation theorem

If $f(x)$ has the Fourier transform $F(s)$, then its autocorrelation function $\int_{-\infty}^{\infty} f^*(u)f(u+x)\,du$ has the Fourier transform $|F(s)|^2$.

Derivation:

$$\begin{aligned} \int_{-\infty}^{\infty} |F(s)|^2 e^{i2\pi xs} \, ds &= \int_{-\infty}^{\infty} F(s)F^*(s)e^{i2\pi xs} \, ds \\ &= f(x) * f^*(-x) \\ &= \int_{-\infty}^{\infty} f(u)f^*(u-x) \, du \\ &= \int_{-\infty}^{\infty} f^*(u)f(u+x) \, du. \end{aligned}$$

A special case of the convolution theorem, the autocorrelation theorem is familiar in communications in the form that the autocorrelation function of a signal is the Fourier transform of its power spectrum.[3] It is illustrated in Fig. 6.12. The unique feature of this theorem, as contrasted with a theorem that could be stated for the self-convolution, is that information about the phase of $F(s)$ is entirely missing from $|F(s)|^2$. The autocorrelation function correspondingly contains no information about the phase of the Fourier components of $f(x)$, being unchanged if phases are allowed to alter, as was shown on p. 45.

Exercise Show that the normalized autocorrelation function $\gamma(x)$, for which $\gamma(0) = 1$ (see p. 41), has as its Fourier transform the normalized power spectrum $|\Phi(s)|^2$ whose infinite integral is unity, and which is defined by

$$|\Phi(s)|^2 = \frac{|F(s)|^2}{\int_{-\infty}^{\infty} |F(s)|^2 \, ds}.$$

A statement may also be added about the function $C(x)$, which was defined in Chapter 3 by the sequence of autocorrelation functions $\gamma_X(x)$ generated from the functions $f(x)\Pi(x/X)$ as $X \to \infty$. If $\gamma_X(x)$ approached a limit, then the limit was called $C(x)$. Corresponding to the sequence of normalized autocorrelation functions, $\gamma_X(x)$ is the sequence of normalized power spectra $|\Phi_X(s)|^2$. If $\gamma_X(x)$ approaches a limit as the segment length X increases, then the normalized power

[3] The corresponding theorem for signals that do not tend to zero as time advances is sometimes referred to as Wiener's theorem (see N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," John Wiley and Sons, New York, 1949).

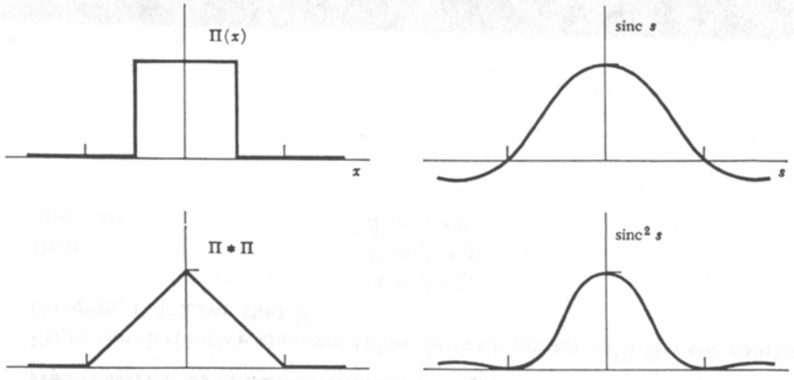


Fig. 6.12 The autocorrelation theorem: autocorrelating a function corresponds to squaring (the modulus of) its transform.

spectrum settles down to a limiting form $|\Phi_\infty(s)|^2$. In these circumstances the autocorrelation theorem takes the form

$$C(x) \supset |\Phi_\infty(s)|^2$$

and one generally says, as before, that the autocorrelation is the Fourier transform of the power spectrum, suiting the definitions to the needs of the case.

Clearly it may happen that the sequence of transforms of $\gamma_X(x)$ does not approach limits for all s but is of a character describable with impulse symbols $\delta(s)$. Therefore situations may be entertained where the transform of $C(x)$ is a generalized function. For example, an ideal line spectrum such as is possessed by a signal carrying finite power at a single frequency is such a case. We know that if $f(x) = \cos \alpha x$, then $C(x) = \cos \alpha x$. The Fourier transform of $C(x)$ is thus a generalized function $\frac{1}{2}\delta(s + \alpha/2\pi) + \frac{1}{2}\delta(s - \alpha/2\pi)$, and if necessary we could work out the sequence of transforms of $\gamma_X(x)$ that define it. The interesting point here, however, is that the power spectrum as a generalized function is not deducible from the autocorrelation theorem, for no interpretation has been given for products such as $[\delta(x)]^2$.

Exercise Give an interpretation for $[\delta(x)]^2$ by attempting to apply the autocorrelation theorem to $f(x) = \cos \alpha x$ and test it on some other simple example such as $f(x) = 1$.

Exercise Show that the situation cannot arise where the sequence $\gamma_X(x)$ calls for the use of $\delta(x)$ in representing $C(x)$.

Exercise We wish to discuss the ideal situation of a power spectrum which is flat and extends to infinite frequency. Determine $C(x)$ and

its transform. Show that the nonnormalized autocorrelation function lends itself to this requirement, and that a good version of the autocorrelation theorem can be devised in which the power spectrum is normalized so as to be equal to unity at its origin. What does this form of the theorem say when $f(x) = \cos \alpha x$?

Derivative theorem

If $f(x)$ has the Fourier transform $F(s)$ then $f'(x)$ has the Fourier transform $i2\pi sF(s)$.

Derivation:

$$\int_{-\infty}^{\infty} f'(x)e^{-i2\pi xs}\,dx = \int_{-\infty}^{\infty} \lim \frac{f(x+\Delta x) - f(x)}{\Delta x}\, e^{-i2\pi xs}\,dx$$

$$= \lim \int_{-\infty}^{\infty} \frac{f(x+\Delta x)}{\Delta x}\, e^{-i2\pi xs}\,dx - \lim \int_{-\infty}^{\infty} \frac{f(x)}{\Delta x}\, e^{-i2\pi xs}\,dx$$

$$= \lim \frac{e^{i2\pi\,\Delta x s}F(s) - F(s)}{\Delta x}$$

$$= i2\pi sF(s).$$

Since taking the derivative of a function multiplies its transform by $i2\pi s$, we can say that differentiation enhances the higher frequencies, attenuates the lower frequencies, and suppresses any zero-frequency component. Examples are given in Figs. 6.13 and 6.14.

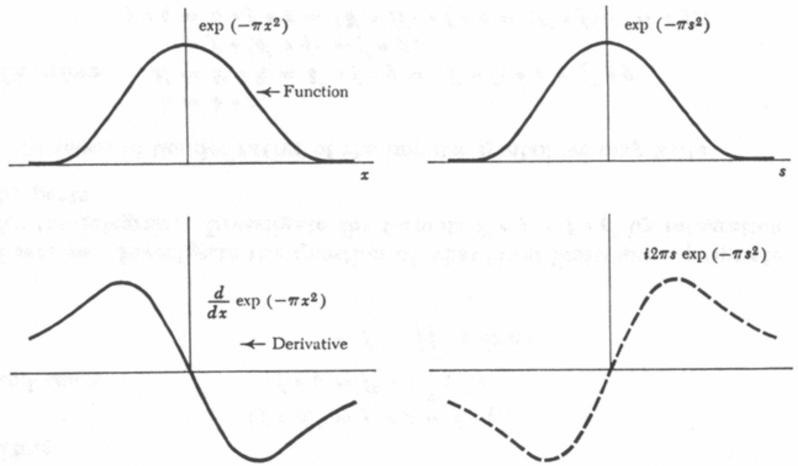


Fig. 6.13 Differentiation of a function incurs multiplication of the transform by i2πs.

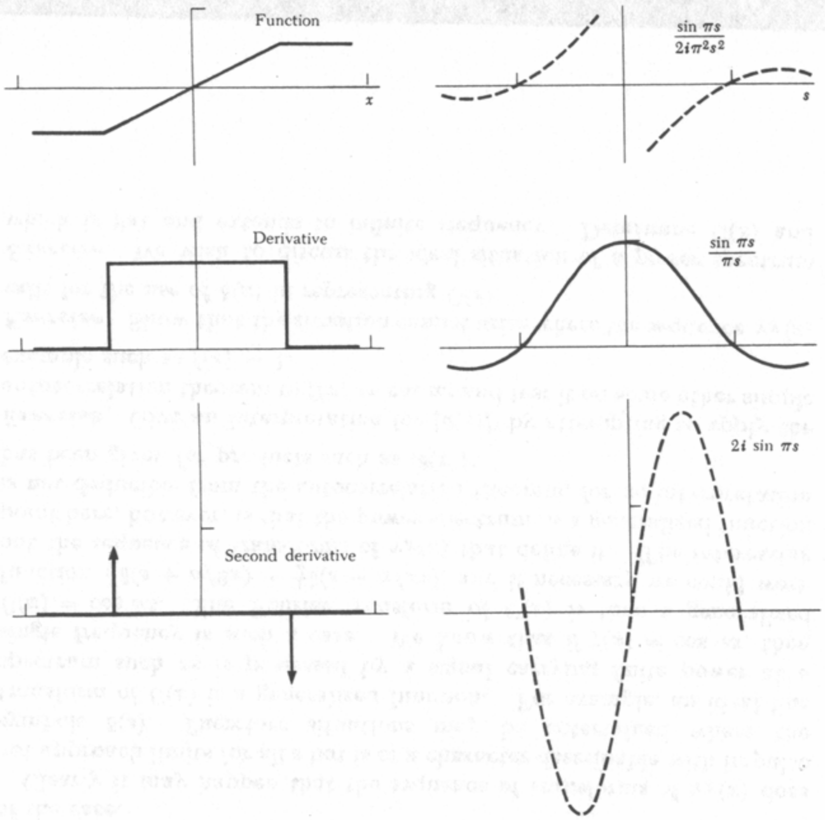


Fig. 6.14 Successive applications of the derivative theorem.

It happens quite frequently that the multiplication by $i2\pi s$ causes the integral of $|i2\pi sF(s)|$ to diverge. Correspondingly, the derivative $f'(x)$ will exhibit infinite discontinuities. Such situations are accommodated by the impulse symbol and its derivatives.

Derivative of a convolution integral

From the derivative theorem taken in conjunction with the convolution theorem, it follows that if

$$h = f * g,$$

then

$$h' = f' * g,$$

and also

$$h' = f * g'.$$

Derivation:

$$\frac{d}{dx}[f(x) * g(x)] \supset i2\pi s[F(s)G(s)]$$

$$f'(x) * g(x) \supset [i2\pi sF(s)]G(s)$$

$$f(x) * g'(x) \supset F(s)[i2\pi sG(s)].$$

These conclusions may be stated in a different form as follows:

The derivative of a convolution is the convolution of either of the functions with the derivative of the other.

Thus

$$(f * g)' = f' * g = f * g',$$

and again

$$f * g = f' * \int^x g\,dx$$

$$= f'' * \iint^x g\,dx\,dx$$

$$\cdots$$

Exercise Investigate the question of what lower limits are appropriate for the integrals. Investigate the formula $f' * g = f * g'$ by integration by parts.

In terms of the derivative of the impulse symbol we may write

$$h = \delta * h.$$

Therefore

$$h' = \delta' * h = \delta' * f * g = (\delta' * f) * g = f' * g$$

$$= f * (\delta' * g) = f * g',$$

$$f * g = \delta * f * g = (\delta' * H) * f * g = (\delta' * f) * (H * g).$$

The formulas quoted here are applicable to the evaluation of particular convolution integrals analytically or numerically but are principally of theoretical value (for example, in the deduction of the uncertainty relation and in deriving the formulas for the response of a filter in terms of its impulse and step responses).

The convenient algebra permitted by the δ notation in conjunction with the associative and other properties of convolution enables rapid generation of relations which may be needed for some problem under study. Many interesting possibilities arise. For example, starting from

$$\frac{d}{dx}(f * g) \supset i2\pi sFG$$

we may factor the right-hand side to get

$$\frac{d}{dx}(f * g) = \frac{d^{\frac{1}{2}}f}{dx^{\frac{1}{2}}} * \frac{d^{\frac{1}{2}}g}{dx^{\frac{1}{2}}} \supset (i2\pi s)^{\frac{1}{2}}F(i2\pi s)^{\frac{1}{2}}G.$$

The transform of a generalized function

Let $p(x)$ be a generalized function defined as in Chapter 5 by the sequence $p_\tau(x)$. Let the Fourier transforms of members of the defining sequence be $P_\tau(s)$, where

$$P_\tau(s) = \int_{-\infty}^{\infty} p_\tau(x)e^{-i2\pi sx}\,dx.$$

Perhaps this new sequence $P_\tau(s)$ defines a generalized function. We know that the members of the sequence are particularly well-behaved; we test the sequence for regularity by means of an arbitrary particularly well-behaved function $F(x)$ whose Fourier transform is $\bar{F}(s)$. From the energy theorem

$$\int_{-\infty}^{\infty} P_\tau(s)\bar{F}(s)\,ds = \int_{-\infty}^{\infty} p_\tau(x)F(-x)\,dx$$

it follows that

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} P_\tau(s)\bar{F}(s)\,ds = \lim_{\tau\to 0}\int_{-\infty}^{\infty} p_\tau(x)F(-x)\,dx,$$

and we know the latter limit exists. Hence $P_\tau(s)$ defines a generalized function, to which we give the symbol $P(s)$ and the meaning "Fourier transform of the generalized function $p(x)$."

The following statement can be made about $p(x)$ and $P(s)$,

$$\int_{-\infty}^{\infty} P(s)\bar{F}(s)\,ds = \int_{-\infty}^{\infty} p(x)F(-x)\,ds,$$

where $F(x)$ is any particularly well-behaved function and $\bar{F}(s)$ is its Fourier transform.

Let $\phi(x)$ be a function, such as a polynomial, that has derivatives of all orders at all points but whose behavior as $|x| \to \infty$ is not so stringently controlled as that of particularly well-behaved functions. We allow $\phi(x)$ to go infinite as $|x|^N$ where N is finite. Functions such as exp x and log x would be excluded. Then products of $\phi(x)$ and any particularly well-behaved function will eventually be overwhelmed by the latter as $|x| \to \infty$, since the particularly well-behaved factor dies out faster than $|x|^{-N}$. Furthermore, since the product has all derivatives at all points, the product itself is particularly well-behaved.

Consider a sequence

$$\phi(x)p_\tau(x).$$

It is particularly well-behaved, and

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} [\phi(x)p_\tau(x)]F(x)\,dx = \lim_{\tau\to 0}\int_{-\infty}^{\infty} p_\tau(x)[\phi(x)F(x)]\,dx,$$

which exists since $p_\tau(x)$ is a regular sequence and $[\phi(x)F(x)]$ is particularly well-behaved. The generalized function so defined we write as

$$\phi(x)p(x).$$

In practice we use this notation with functions $\phi(x)$ that have a *sufficient* number of derivatives and include exponentially increasing functions when, as is the case with the pulse sequence $\tau^{-1}\Pi(x/\tau)$, the behavior at infinity is inessential.

Nothing is introduced which could be called the product of two generalized functions; the product of two defining sequences is not necessarily a regular sequence and consequently does not in general define a generalized function.

Proofs of theorems

The numerous theorems of Fourier theory, which have proved so fruitful in the preceding sections, have shown themselves perfectly adaptable to the insertion of the impulse symbol $\delta(x)$, the shah symbol III(x), the duplicating symbol II(x), and other familiar nonfunctions. In the course of standard proofs of the theorems it is found necessary to eliminate these cases, and in the outcome we have conditions for the applicability of the theorems which we have found in practice need not be observed. This situation was dealt with by introducing the idea of a transform in the limit, and special ad hoc interpretations as limits were placed on expressions containing impulse symbols.

Having established the algebra for generalized functions, we can also give systematic proofs of the various theorems, free from the awkward conditions that arise when attention is confined to ordinary functions possessing regular transforms. The difficulties associated with functions that do not have derivatives disappear, for generalized functions possess derivatives of all orders. And more intolerable circumstances, such as the lack of a regular spectrum for direct current, also vanish.

Addition theorem The proof of this theorem can be supplied by the reader.

Similarity and shift theorems We prove these theorems simultaneously. Let $p(x)$ be a generalized function with Fourier transform $P(s)$. Then

$$p(ax+b) \supset \frac{1}{|a|}\,e^{i2\pi bs/a}P\left(\frac{s}{a}\right).$$

Proof: Since

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} p_\tau(ax+b)F(x)\,dx = \frac{1}{|a|}\lim_{\tau\to 0}\int_{-\infty}^{\infty} p_\tau(x)F\left(\frac{x-b}{a}\right)dx$$

exists, we have a meaning for $p(ax+b)$. Now

$$p_\tau(ax+b) \supset \frac{1}{|a|}\,e^{i2\pi bs/a}P_\tau\left(\frac{s}{a}\right)$$

by substitution of variables. Hence the two theorems follow.

Derivative theorem The Fourier transform of $p_\tau'(x)$ is $i2\pi sP_\tau(s)$. Hence the Fourier transform of $p'(x)$ is $i2\pi sP(s)$.

Power theorem Since no meaning has been assigned to the product of two generalized functions, the best theorem that can be proved is

$$\int_{-\infty}^{\infty} P(s)\bar{F}(s)\,ds = \int_{-\infty}^{\infty} p(x)F(-x)\,dx,$$

where $F(x)$ is a particularly well-behaved function and $p(x)$ is a generalized function. The theorem follows from the fact that

$$\lim_{\tau\to 0}\int_{-\infty}^{\infty} P_\tau(s)\bar{F}(s)\,ds = \lim_{\tau\to 0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\bar{F}(s)p(x)e^{-i2\pi sx}\,dx\,ds$$
$$= \lim_{\tau\to 0}\int_{-\infty}^{\infty} p_\tau(x)F(-x)\,dx.$$

Summary of theorems

The theorems discussed in the preceding pages are collected for reference in Table 6.1.

Table 6.1 Theorems for the Fourier transform

Theorem	$f(x)$	$F(s)$
Similarity	$f(ax)$	$\frac{1}{\lvert a\rvert}F\left(\frac{s}{a}\right)$
Addition	$f(x)+g(x)$	$F(s)+G(s)$
Shift	$f(x-a)$	$e^{-i2\pi as}F(s)$
Modulation	$f(x)\cos\omega x$	$\frac{1}{2}F\left(s-\frac{\omega}{2\pi}\right)+\frac{1}{2}F\left(s+\frac{\omega}{2\pi}\right)$
Convolution	$f(x)*g(x)$	$F(s)G(s)$
Autocorrelation	$f(x)*f^*(-x)$	$\lvert F(s)\rvert^2$
Derivative	$f'(x)$	$i2\pi sF(s)$
Derivative of convolution	$\frac{d}{dx}[f(x)*g(x)] = f'(x)*g(x) = f(x)*g'(x)$	
Rayleigh	$\int_{-\infty}^{\infty}\lvert f(x)\rvert^2\,dx = \int_{-\infty}^{\infty}\lvert F(s)\rvert^2\,ds$	
Power	$\int_{-\infty}^{\infty} f(x)g^*(x)\,dx = \int_{-\infty}^{\infty} F(s)G^*(s)\,ds$	
(f and g real)	$\int_{-\infty}^{\infty} f(x)g(-x)\,dx = \int_{-\infty}^{\infty} F(s)G(s)\,ds$	

Problems

1 Using the transform pairs given for reference, deduce the further pairs listed below by application of the appropriate theorem. Assume that A and σ are positive.

$$\frac{\sin x}{x} \supset \pi\Pi(\pi s) \qquad \left(\frac{\sin x}{x}\right)^2 \supset \pi\Lambda(\pi s)$$
$$\frac{\sin Ax}{Ax} \supset \frac{\pi}{A}\Pi\left(\frac{\pi s}{A}\right) \qquad \left(\frac{\sin Ax}{Ax}\right)^2 \supset \frac{\pi}{A}\Lambda\left(\frac{\pi s}{A}\right)$$
$$e^{-x^2} \supset \pi^{\frac{1}{2}}e^{-\pi^2s^2} \qquad \delta(ax) \supset \frac{1}{\lvert a\rvert}$$
$$e^{-Ax^2} \supset \left(\frac{\pi}{A}\right)^{\frac{1}{2}}e^{-\pi^2s^2/A} \qquad \delta(ax+b) \supset \frac{1}{\lvert a\rvert}e^{i2\pi bs/a}$$
$$e^{-x^2/2\sigma^2} \supset (2\pi)^{\frac{1}{2}}\sigma e^{-2\pi^2\sigma^2s^2} \qquad e^{ix} \supset \delta\left(s-\frac{1}{2\pi}\right)$$

2 Show that the following transform pairs follow from the addition theorem, and make graphs.

$$1+\cos\pi x \supset \delta(s)+\mathrm{II}(s)$$
$$1+\sin\pi x \supset \delta(s)+i\,\mathrm{I_I}(s)$$
$$\operatorname{sinc} x+\tfrac{1}{2}\operatorname{sinc}^2\tfrac{1}{2}x \supset \Pi(s)+\Lambda(2s)$$
$$A^{-\frac{1}{2}}e^{-\pi x^2/A}+A^{\frac{1}{2}}e^{-\pi Ax^2} \supset e^{-\pi s^2/A}+e^{-\pi As^2}$$
$$4\cos^2\pi x+4\cos^2\tfrac{1}{2}\pi x-3 \supset \delta(s+\tfrac{1}{2})+\delta(s-\tfrac{1}{2})+\delta(s)+\tfrac{1}{2}\delta(x-1)+\tfrac{1}{2}\delta(s+1).$$

3 Deduce the following transform pairs, using the shift theorem.

$$\frac{\cos\pi x}{\pi(x-\frac{1}{2})} \supset -e^{-i\pi s}\Pi(s)$$
$$\frac{\sin\pi x}{\pi(x-1)} \supset -e^{-i2\pi s}\Pi(s)$$
$$\Lambda(x-1) \supset e^{-i2\pi s}\operatorname{sinc}^2 s$$
$$\Pi(x-\tfrac{1}{2}) \supset e^{-i\pi s}\operatorname{sinc} s$$
$$\Pi(x)\operatorname{sgn} x \supset -i\sin\tfrac{1}{2}\pi s\operatorname{sinc}\tfrac{1}{2}s$$
$$\Pi\left(\frac{x-\frac{1}{2}a}{a}\right) \supset \lvert a\rvert e^{-i\pi as}\operatorname{sinc} as.$$

4 Use the convolution theorem to find and graph the transforms of the following functions: $\operatorname{sinc} x\operatorname{sinc} 2x$, $(\operatorname{sinc} x\cos 10x)^2$.

5 Let $f(x)$ be a periodic function with period a, that is, $f(x+a) = f(x)$ for all x. Since the Fourier transform of $f(x+a)$ is, by the shift theorem, equal to $\exp(i2\pi as)F(s)$, which must be equal to $F(s)$, what can be deduced about the transform of a periodic function?

6 Graph the transform of $f(x) \sin \omega x$ for large and small values of ω, and explain graphically how, for small values of ω, the transform of $f(x) \sin \omega x$ is proportional to the derivative of the transform of $f(x)$.

7 Graph the transform of $\exp(-x)H(x) \cos \omega x$. Is it an even function of s?

8 Show that a pulse signal described by $\Pi(x/X) \cos 2\pi f x$ has a spectrum

$$\tfrac{1}{2}X\{\operatorname{sinc}[X(s+f)] + \operatorname{sinc}[X(s-f)]\}$$

9 Show that a modulated pulse described by $\Pi(x/X)(1 + M \cos 2\pi F x) \cos 2\pi f x$ has a spectrum

$$\begin{aligned}\tfrac{1}{2}X\{\operatorname{sinc}[X(s+f)] + \operatorname{sinc}[X(s-f)]\} + \tfrac{1}{4}MX\{\operatorname{sinc}[X(s+f+F)] \\ + \operatorname{sinc}[X(s+f-F)] + \operatorname{sinc}[X(s-f+F)] + \operatorname{sinc}[X(s-f-F)]\}.\end{aligned}$$

Graph the spectrum to a suitably exaggerated scale for a case where there are 100 modulation cycles and 100,000 radio-frequency cycles in one pulse and the modulation coefficient M is 0.6. Show by dimensioning how the factors 100, 100,000, and 0.6 enter into the shape of the spectrum.

10 A function $f(x)$ is defined by

$$f(x) = \begin{cases} 0 & |x| > 2 \\ 2 - |x| & 1 < |x| < 2 \\ 1 & |x| < 1; \end{cases}$$

show that

$$f(x) = 2\Lambda\left(\frac{x}{2}\right) - \Lambda(x) = \Lambda(x) * [2\text{Ш}(x) + \delta(x)]$$

and hence that

$$F(s) = 4 \operatorname{sinc}^2 2s - \operatorname{sinc}^2 s = \operatorname{sinc}^2 s(1 + 2 \cos 2\pi s).$$

11 Prove that $f * g * h \supset FGH$ and hence that $f^{*n} \supset F^n$.

12 The notation f^{*n} meaning $f(x)$ convolved with itself $n - 1$ times, where $n = 2, 3, 4, \ldots$, suggests the idea of fractional-order self-convolution. Show that such a generalization of convolution is readily made and that, for example, one reasonable expression for $f(x)$ convolved with itself half a time would be

$$f^{*1\frac{1}{2}} \equiv \int e^{i2\pi sx}\left[\int e^{-i2\pi su} f(u)\, du\right]^{1\frac{1}{2}} ds.$$

13 Prove that

$$(f * g)(h * j) \supset (FG) * (HJ)$$

and that

$$(f + g) * (h + j) \supset FH + FJ + GH + GJ.$$

14 Use the convolution theorem to obtain an expression for

$$e^{-ax^2} * e^{-bx^2}.$$

15 Prove that

$$\int_{-\infty}^{\infty} f^*(u) g^*(x - u)\, du \supset F^*(-s)G^*(-s).$$

16 Prove that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^*(u) g^*(u - x) e^{-i2\pi xs}\, du\, dx = F^*(-s)G^*(s).$$

17 Show by Rayleigh's theorem that

$$\int_{-\infty}^{\infty} \operatorname{sinc}^2 x\, dx = 1$$

$$\int_{-\infty}^{\infty} \operatorname{sinc}^4 x\, dx = \int_{-\infty}^{\infty} [\Lambda(x)]^2\, dx = \tfrac{2}{3}$$

$$\int_{-\infty}^{\infty} [J_0(x)]^2\, dx = \infty$$

$$\int_{-\infty}^{\infty} \frac{dx}{(1 + x^2)^2} = \frac{\pi}{2}$$

18 Complete the following schemata for reference, including thumbnail sketches of the functions.

function	$\supset$	transform
autocorrelation $f \star f$	$\supset$	power spectrum $\|F(s)\|^2$

$\Pi(x)$	

$\Lambda(x)$	

$\delta(x)$	

sinc x	

$\Pi(x) \cos 2\pi f x$	

19 Show the fallacy in the following reasoning. "The Fourier transform of $\int_{-\infty}^{x} f(x)\, dx$ must be $F(s)/i2\pi s$ because the derivative of $\int_{-\infty}^{x} f(x)\, dx$ is $f(x)$, and hence by the derivative theorem the transform of $f(x)$ would be $F(s)$, which is true."

20 Establish an integral theorem for the Fourier transform of the indefinite integral of a function.

21 Use the derivative theorem to find the Fourier transform of $xe^{-\pi x^2}$.

22 Show that $2\pi x \Pi(x) \supset i \operatorname{sinc}' s$.

23 The following brief derivation appears to show that the area under a derivative is zero. Thus

$$\int_{-\infty}^{\infty} f'(x)\,dx = \overline{f'(x)}\Big|_0 = i2\pi s F(s)\Big|_0 = 0.$$

Confirm that this is so, or find the error in reasoning.

24 Show that

$$f(ax-b) \supset \frac{1}{|a|} e^{-i2\pi bs/a} F\left(\frac{s}{a}\right).$$

25 Show from the energy theorem that

$$\int_{-\infty}^{\infty} e^{-\pi x^2} \cos 2\pi a x\,dx = e^{-\pi a^2}.$$

26 Show from the energy theorem that

$$\int_{-\infty}^{\infty} \operatorname{sinc}^2 x \cos \pi x\,dx = \tfrac{1}{2}.$$

27 Show that the function whose Fourier transform is $|\operatorname{sinc} s|$ has a triangular autocorrelation function.

28 As a rule, the autocorrelation function tends to be more spread out than the function it comes from. But show that

$$\frac{1}{\pi x} * \frac{-1}{\pi x} = \delta(x).$$

Show that $(\pi x)^{-1}$ must have a flat energy spectrum, and from that deduce and investigate other functions whose autocorrelation is impulsive.

29 The Maclaurin series for $F(s)$ is

$$F(0) + sF'(0) + \frac{s^2}{2!}F''(0) + \cdots.$$

Consider the case of $F(s) = \exp(-\pi s^2)$, where the series is known to converge and to converge to $F(s)$. Thus, in this particular case,

$$F(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} F^{(n)}(0).$$

If $F(s)$ is the transform of $f(x)$, then transforming this equation we obtain

$$f(x) = \delta(x)\int_{-\infty}^{\infty} f(x)\,dx - \delta'(x)\int_{-\infty}^{\infty} xf(x)\,dx + \delta''(x)\int_{-\infty}^{\infty} \frac{x^2}{2!} f(x)\,dx + \cdots.$$

How do you explain this result?

Chapter 7 Doing transforms

A number of transforms were introduced earlier to illustrate the basic theorems of the Fourier transformation. Of course, one need not necessarily be aware of any particular Fourier transform pairs to appreciate the meaning of the theorems. Many of the chains of argument in which the Fourier transformation is important are independent of any knowledge of particular examples. Even so, carrying out a general argument with a special case in mind often serves as insurance against surprises.

The examples of Fourier transform pairs chosen for illustration were all introduced without derivation and asserted to be verifiable by evaluation of the Fourier integral. Obviously this does not help when it is necessary to generate new pairs. We therefore consider various ways of carrying out the Fourier transformation. Numerical methods based on the discrete Fourier transform and allowing for use of the fast Fourier transform algorithm are discussed in Chapter 18.

Starting from the given function $f(x)$ whose Fourier transform is to be deduced, one may first contemplate the integral

$$\int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx.$$

If this integral can be evaluated for all s, the problem is solved.

A number of different approaches from the standpoint of integration are discussed separately in subsequent sections.

In addition to this direct approach we have a powerful resource in the basic theorems which have now been established. Many of the theorems take the form "If f and F are a transform pair then g and G are also." Thus, if one knows a transform pair to begin with, others may be generated. It is indeed possible to build up an extensive dictionary of transforms by means of the theorems, starting from the beginnings already laid down. It may be surmised that some classes of function will never be stumbled on in this way; on the other hand, a variety of physically feasible functions prove to be accessible. Generation from theorems is taken up later with examples. Finally, there is the possibility of extracting a desired transform from tables.

Integration in closed form

It is a propitious circumstance if $f(x)$ is zero over some range of x and if in addition its behavior is simple where it is nonzero. Thus if

$$f(x) = \Pi(x),$$

then

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)e^{-i2\pi xs}\,dx &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-i2\pi xs}\,dx \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos 2\pi xs\,dx \\ &= \frac{\sin \pi s}{\pi s} \\ &= \operatorname{sinc} s.\end{aligned}$$

If $f(x)$ is nonzero on several segments of the abscissa and constant within each segment, the integration in closed form can likewise be done. Thus if we let

$$f(x) = a\Pi\left(\frac{x-b}{c}\right),$$

then

$$\begin{aligned}F(s) &= \int_{b-\frac{1}{2}c}^{b+\frac{1}{2}c} ae^{-i2\pi xs}\,dx \\ &= a\int_{-\frac{1}{2}c}^{\frac{1}{2}c} e^{-i2\pi(u+b)s}\,du \\ &= ae^{-i2\pi bs}\int_{-\frac{1}{2}c}^{\frac{1}{2}c} \cos 2\pi us\,du \\ &= ace^{-i2\pi bs}\operatorname{sinc} cs.\end{aligned}$$

It follows that if

$$f(x) = \sum_n a_n\Pi\left(\frac{x-b_n}{c_n}\right),$$

then

$$F(s) = \sum_n a_n c_n e^{-i2\pi b_n s}\operatorname{sinc} c_n s.$$

Functions built up segmentally of rectangle functions include staircase functions and functions suitable for discussing Morse code, teleprinter signals, and on/off servomechanisms. They not only occur frequently in engineering but can also simulate, as closely as may be desired, any kind of variation.

An even simpler case arises if $f(x)$ is zero almost everywhere; for example, suppose that $f(x)$ is a set of impulses of various strengths at various values of x:

$$f(x) = \sum_n a_n\,\delta(x-b_n).$$

Then only the values of $a_n \exp(-i2\pi xs)$ at the points $x = b_n$ can matter. By the sifting theorem for the impulse symbol,

$$\begin{aligned}F(s) &= \int_{-\infty}^{\infty} \sum a_n\,\delta(x-b_n)e^{-i2\pi xs}\,dx \\ &= \int_{b_1-}^{b_1+} a_1\,\delta(x-b_1)e^{-i2\pi xs}\,dx + \int_{b_2-}^{b_2+} a_2\,\delta(x-b_2)e^{-i2\pi xs}\,dx + \cdots \\ &= a_1e^{-i2\pi b_1 s} + a_2e^{-i2\pi b_2 s} + \cdots.\end{aligned}$$

If $f(x)$ has some special functional form, it may prove possible to perform the integration, but no general rules can be given. Some simple examples follow.

1. Let $f(x) = \operatorname{sinc} x$. Then

$$\begin{aligned}F(s) &= \int_{-\infty}^{\infty} \operatorname{sinc} x\, e^{-i2\pi xs}\,dx \\ &= \int_{-\infty}^{\infty} \frac{\sin \pi x \cos 2\pi xs}{\pi x}\,dx \\ &= \int_{-\infty}^{\infty}\left[\frac{\sin(\pi x + 2\pi xs)}{2\pi x} + \frac{\sin(\pi x - 2\pi xs)}{2\pi x}\right]dx \\ &= \int_{-\infty}^{\infty}\left\{\frac{1+2s}{2}\operatorname{sinc}[(1+2s)x] + \frac{1-2s}{2}\operatorname{sinc}[(1-2s)x]\right\}dx \\ &= \frac{1+2s}{2|1+2s|} + \frac{1-2s}{2|1-2s|} \\ &= \Pi(s).\end{aligned}$$

We have used the result that

$$\int_{-\infty}^{\infty} \operatorname{sinc} ax\,dx = \frac{1}{|a|}.$$

2. Let $f(x) = e^{-|x|}$. Then

$$\begin{aligned}F(s) &= \int_{-\infty}^{\infty} e^{-|x|}e^{-i2\pi xs}\,dx \\ &= 2\int_0^{\infty} e^{-x}\cos 2\pi xs\,dx \\ &= 2\operatorname{Re}\int_0^{\infty} e^{-x}e^{i2\pi xs}\,dx \\ &= 2\operatorname{Re}\int_0^{\infty} e^{(i2\pi s-1)x}\,dx \\ &= 2\operatorname{Re}\frac{-1}{i2\pi s-1} \\ &= \frac{2}{4\pi^2s^2+1}.\end{aligned}$$

3. Let $f(x) = e^{-\pi x^2}$. Then

$$\begin{aligned} F(s) &= \int_{-\infty}^{\infty} e^{-\pi x^2} e^{-i2\pi xs}\,dx \\ &= \int_{-\infty}^{\infty} e^{-\pi(x^2+i2xs)}\,dx \\ &= e^{-\pi s^2}\int_{-\infty}^{\infty} e^{-\pi(x+is)^2}\,dx \\ &= e^{-\pi s^2}\int_{-\infty}^{\infty} e^{-\pi(x+is)^2} d(x+is) \\ &= e^{-\pi s^2}. \end{aligned}$$

In this example we have used the known result that the infinite integral of $\exp(-\pi x^2)$ is unity. The next case illustrates a contour integral.

4. Let $f(x) = x^{-1}$. The infinite discontinuity at the origin causes the standard Fourier integral to diverge; hence we consider instead

$$\lim_{\epsilon\to 0}\left(\int_{-\infty}^{-\epsilon} \frac{e^{-i2\pi sx}}{x}\,dx + \int_{\epsilon}^{\infty}\right).$$

Now consider

$$\int_C \frac{e^{-i2\pi sz}}{z}\,dz,$$

where the contour C is a semicircle of radius R in the complex plane of z, whose diameter lies along the real axis and has a small indentation of radius ϵ at the origin. The contour integral is zero since no poles are enclosed. Thus

$$\int_{-R}^{-\epsilon} \frac{e^{-i2\pi sx}}{x}\,dx + \int_{\pi}^{0} ie^{-i2\pi s\epsilon e^{i\theta}}\,d\theta + \int_{\epsilon}^{R} \frac{e^{-i2\pi sx}}{x}\,dx + \int_{0}^{\pi} ie^{-i2\pi sRe^{i\theta}}\,d\theta = 0.$$

As $R \to \infty$, the fourth integral vanishes and the second equals $\pm i\pi$ according to the sign of s, so that we have

$$\lim_{\epsilon\to 0}\left(\int_{-\infty}^{-\epsilon} \frac{e^{-i2\pi sx}}{x}\,dx + \int_{\epsilon}^{\infty}\right) + i\pi \operatorname{sgn} s = 0.$$

Hence the desired transform pair in the limit is

$$\frac{1}{x} \supset -i\pi \operatorname{sgn} s.$$

5. Let $f(x) = \operatorname{sgn} x$. This is the previous example in reverse, and it is seen that in this case the Fourier integral fails to exist in the standard sense because the function does not possess an absolutely convergent integral. Therefore consider a sequence of transformable functions which approach sgn x as a limit, for example, the sequence $\exp(-\tau|x|) \operatorname{sgn} x$ as $\tau \to 0$. The transform will be

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-\tau|x|} \operatorname{sgn} x\, e^{-i2\pi xs}\,dx &= \int_{-\infty}^{0} -e^{(\tau-i2\pi s)x}\,dx + \int_{0}^{\infty} e^{-(\tau+i2\pi s)}\,dx \\ &= -\frac{1}{\tau - i2\pi s} + \frac{1}{\tau + i2\pi s}. \end{aligned}$$

As $\tau \to 0$ this expression has a limit $1/i\pi s$. Hence we have the Fourier transform pair in the limit,

$$\operatorname{sgn} x \supset \frac{1}{i\pi s}.$$

Numerical Fourier transformation

If the values of a function have been obtained by physical measurement and it is necessary to find its Fourier transform, various possibilities exist.

First of all, certain limitations of physical data will influence the result, and we shall begin with this aspect of numerical transformation.

The data will be given at discrete values of the independent variable x. The interval Δx may be so fine that there is no concern about interpolating intermediate values, but in any case we can take the view that the data can hardly contain significant information about Fourier components with periods less than $2\ \Delta x$. Therefore it is not necessary to take the computations to frequencies higher than about $(2\ \Delta x)^{-1}$.

In addition, observational data are given for a finite range of x, let us say for $-X < x < X$. By a corresponding argument, the Fourier transform need not be calculated for values of s more closely spaced than $(2X)^{-1}$. If there were any significant fine detail in $F(s)$ that required a finer tabulation interval than $(2X)^{-1}$ for its description, then measurements of $f(x)$ would have to be extended beyond $x = X$ to reveal it.

These simple but important facts for the computer programmer may be summarized by saying that the function $f(x)$ tabulated at interval Δx over a range $2X$ possesses $2X/\Delta x$ degrees of freedom, and this should be comparable with the number of data computed for the transform. In Chapters 10 and 18 the underlying thought is explored in detail.

A further property of physical data is to contain errors. There is therefore a limit to the precision that is warranted in calculating values of the Fourier transform. This limit is expressed concisely by the power spectrum of the error component (or, what is equivalent, the autocorrelation function of the error component). Sometimes only the magnitude of the errors is available, and not their spectrum; and sometimes the magnitude is not known either. Nevertheless, the errors set a limit to the number of physically significant decimal places in a computed value of the transform.

Let $f(x)$ be represented by values at $x = n$, where the integer n ranges from $-N$ to N. Before the Fourier transform of $f(x)$ can be computed, it is necessary to have information about $f(x)$ over the whole infinite range of x. Therefore we call on our physical knowledge and make some assumption about the behavior outside the range of measurement. In this case suppose that $f(x)$ is zero where $|x| > N$. Then the sum

$$\sum_{-N}^{N} f(n)e^{i2\pi sn}$$

will be an approximation to $F(s)$. In practice one computes the real and imaginary parts separately, forming the sums

$$\sum_{-N}^{N} f(n) \cos 2\pi sn \qquad \text{and} \qquad \sum_{-N}^{N} f(n) \sin 2\pi sn.$$

This summation, over $2N + 1$ terms, must be done for each chosen value of s. For this reason it is convenient in using a desk calculator to possess tables of cosine and sine prepared for suitable values of s; instead of being limited to one quadrant they should run on and on, showing negative values as they occur.

Generation of transforms by theorems

A wide variety of functions, especially those occurring in theoretical work, can be transformed if some property can be found that permits a simplifying application of a theorem. For example, consider a polygonal function, as in Fig. 7.1. If we perceive that it can be expressed as the convolution of the triangle function and set of impulses, then we can handle

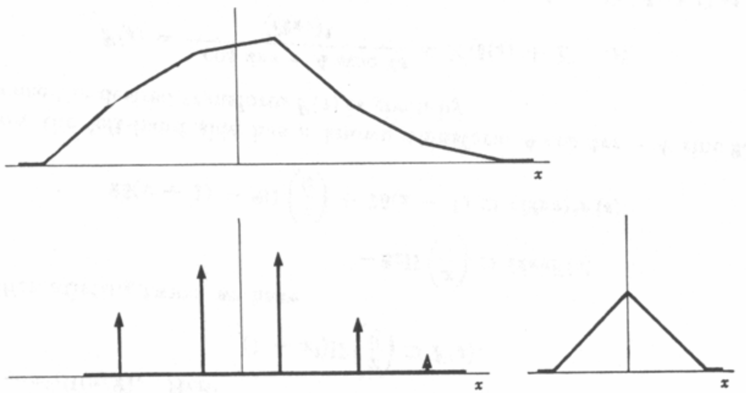


Fig. 7.1 A polygonal function (above) that can be regarded as the convolution of a set of impulses and a triangle function (below).

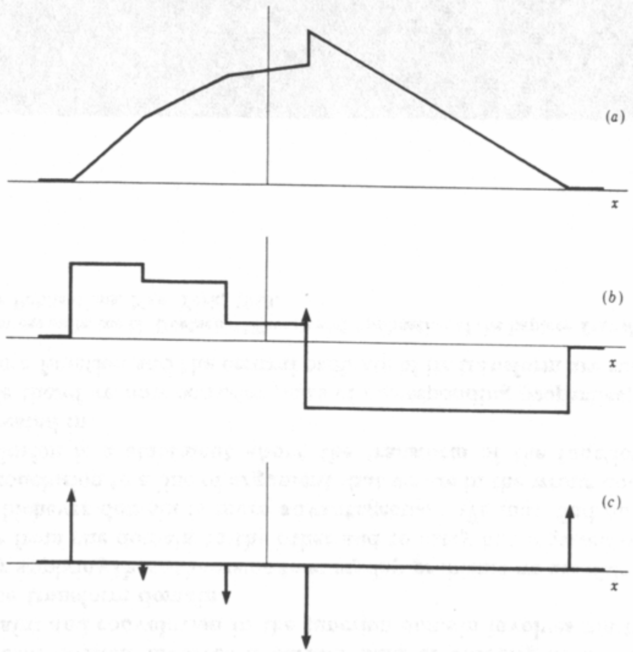


Fig. 7.2 Technique of reduction to impulses by continued differentiation.

the impulses as described above and multiply their transform by the transform of the triangle function. Thus

$$\Lambda(x) * \sum_n A_n\, \delta(x - a_n) \supset \operatorname{sinc}^2 s \sum_n A_n e^{-i2\pi a_n s}.$$

As an example take the trapezoidal pulse $\Lambda(x) * \Pi(x)$. Evidently

$$\Lambda(x) * \Pi(x) \supset \operatorname{sinc}^2 s \cos \pi s.$$

In practice the convolution theorem is frequently applicable for the generation of derived transforms, and many examples of the use of the convolution theorem and other theorems will be given in the problems for this chapter.

Application of the derivative theorem to segmented functions

There is a special application of the derivative theorem that has wide use in connection with switching waveforms. Consider a segmentally linear function such as that of Fig. 7.2a. The first derivative, shown in b, contains an impulse. Since the transform of the impulse is known, we remove the impulse and differentiate again. This time there remains only a set of impulses $\Sigma C_n\,\delta(x - c_n)$. If the first derivative contained, instead of a single impulse as in the example, a set of impulses $\Sigma B_n\,\delta(x - b_n)$, and if the original function $f(x)$ contained impulses $\Sigma A_n\,\delta(x - a_n)$, then evidently the transform $F(s)$ is given by

$$(i2\pi s)^2 F(s) = (i2\pi s)^2 \Sigma A_n e^{-i2\pi a_n s} + i2\pi s \Sigma B_n e^{-i2\pi b_n s} + \Sigma C_n e^{-i2\pi c_n s}.$$

Clearly this technique extends immediately to functions composed of segments of polynomials, in which case further continued differentiation is required. As a simple example let us consider the parabolic pulse $(1 - x^2)\Pi(x/2)$. Here

$$(1 - x^2)\Pi\left(\frac{x}{2}\right) \supset F(s).$$

Differentiating twice, we have

$$-2x\Pi\left(\frac{x}{2}\right) \supset i2\pi s F(s)$$

$$2\delta(x + 1) - 2\Pi\left(\frac{x}{2}\right) + 2\delta(x - 1) \supset (i2\pi s)^2 F(s).$$

Now the left-hand side has a known transform $4 \cos 2\pi s - 4 \operatorname{sinc} 2s$. Hence the desired transform $F(s)$ is given by

$$F(s) = \frac{4 \cos 2\pi s - 4 \operatorname{sinc} 2s}{(i2\pi s)^2} + K_1\delta(s) + K_2\delta'(s),$$

where K_1 and K_2 are integration constants arising from the fact that a constant K_1 may be added to the original parabolic pulse without changing its first derivative, and similarly for the second derivative. Integration of the given pulse shows that no additive constant or linear ramp is present, so in this case $K_1 = K_2 = 0$ and

$$F(s) = -\frac{\cos 2\pi s}{\pi^2 s^2} + \frac{\sin 2\pi s}{2\pi^3 s^3}.$$

Chapter 8 The two domains

We may think of functions and their transforms as occupying two domains, sometimes referred to[1] as the upper and the lower, as if functions circulated at ground level and their transforms in the underworld. There is a certain convenience in picturing a function as accompanied by a counterpart in another domain, a kind of shadow which is associated uniquely with the function through the Fourier transformation, and which changes as the function changes. In the illustrations given here the uniform practice is to keep the functions on the left and the transforms on the right.

The theorems given earlier can be regarded as a list of pairs of corresponding simultaneous operations, one in the function domain and the other in the transform domain. For example, compression of the abscissas in the function domain means expansion of the abscissas plus contraction of the ordinates in the transform domain; translation in the function domain involves a certain kind of twisting in the transform domain; and convolution in the function domain involves multiplication in the transform domain.

By applying these theorems to everyday problems we are able at will to cross from one domain to the other and to carry out required operations in whichever domain is more advantageous. We may find on reaching the conclusion to a line of argument that we are in the wrong domain; our conclusion is a statement about the transform of the function we are interested in.

We therefore now consider pairs of corresponding properties; the area under a function and the central ordinate of its transform are such a pair.

[1] For example, see G. Doetsch, "Theory and Application of the Laplace Transformation," Dover Publications, New York, 1943.

longer duration than the input waveform but that the amount of stretching, as measured by equivalent width, is short compared with the duration of the impulse response.

5 The (complex) electrical length of a uniform transmission line is θ. Show that the transfer factor relating the output voltage to the input voltage is given by $T(f) = \operatorname{sech} \theta$ when no load impedance is connected.

6 A filter consists of a tee-section whose series impedances are Z_1 and Z_2 and the shunt impedance is Z_3. Show that the voltage transfer function is given by

$$T(f) = \frac{Z_3}{Z_1 + Z_3}.$$

7 In the tee-section of the previous problem, let $Z_2 = 0$ and $Z_3 = R$. The element Z_1 is an open-circuited length of loss-free transmission line of characteristic impedance R, so that we may write $Z_1 = -iR \cot 2\pi Tf$, where T is a constant. Show that the output voltage response to an input voltage step is a rectangle function of time, and hence that in general the output voltage is the finite difference of the input voltage.

8 A very large number of identical passive two-port networks are cascaded, and a voltage impulse is applied to the input, causing a disturbance to propagate down the chain. The disturbance at a distant point is expressible by repeated self-convolution of the impulse response of a single network. Does the disturbance approach Gaussian form?

9 Show that a *linear* system can be imagined whose response to a modulated signal $(1 + M \cos \omega t) \cos \Omega t$ is proportional to the audio signal $M \cos \omega t$.

Chapter 10 Sampling and series

Suppose that we are presented with a function whose values were chosen arbitrarily, and suppose that no connection exists between the neighboring values chosen for the dependent variable. Thus if the independent variable were time, we would have to expect jumps of arbitrary magnitude and sign from one instant to the next.

In nature such a function would never be observed. Because of limitations of the measuring instrument, or for other reasons, there is always a limit to the rate of change. One might thus surmise that there is at least a little interdependence between waveform values at neighboring instants, and consequently that it might be possible to predict from past values over a certain brief time interval. This suggests that it might be possible to dispense with the values of a function for intervals of the same order and yet preserve essentially all the information by noting a set of values spaced at fine, but not infinitesimal, intervals. From this set of samples it would seem reasonable that the intervening values could be recovered. if only to some degree of approximation.

Sampling theorem

The sampling theorem states that, under a certain condition, it is in fact possible to recover the intervening values with full accuracy; in other words, the sample set can be fully equivalent to the complete set of function values. The condition is that the function should be "band-limited"; that is, its Fourier transform is nonzero over a finite range of the transform variable.

Clearly, the interval between samples is crucial in deciding the utility of the theorem; if the samples had to be very close together, not much would be gained. As an illustration of the fineness of sampling we take a simple band-limited function, namely sinc x, whose spectrum is flat up to a cutoff value $s = \frac{1}{2}$. The sampling interval, deduced as explained below, is 1. Figure 10.1a shows the sample values that define sinc x, and it will be seen that the interval is extremely coarse in comparison, for example, with the interval that would be chosen for numerical integration. The sampling intervals indicated by this theorem often seem, at first, to be surprisingly wide. Also shown in Fig. 10.1b are a set of samples for $\text{sinc}^2 \frac{1}{2}x$, a function whose spectrum cuts off at the same point ($s = \frac{1}{2}$) as that of sinc x. Figures 10.1c and 10.1d provide some samples for experiment.

With any given waveform, there is always a frequency beyond which spectral contributions are negligible for some purposes. However, on the other hand, the transform probably never cuts off absolutely; consequently, in applications of the sampling theorem, the error incurred by taking a given waveform to be band-limited must always be estimated.

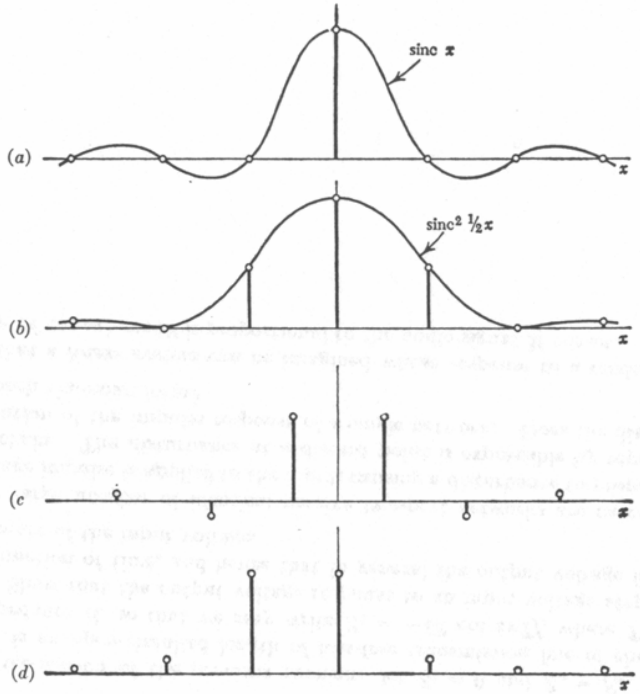
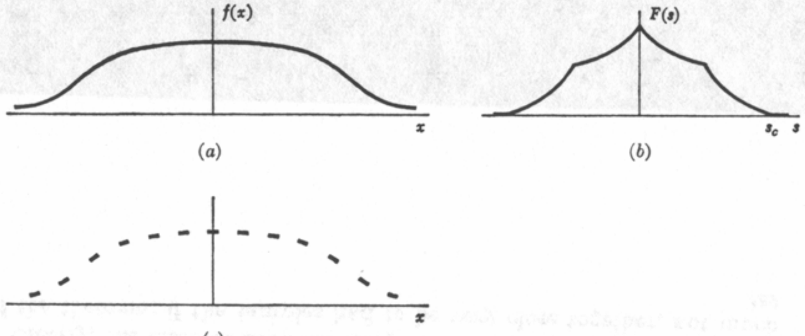


Fig. 10.1 Two functions and their samples, and two sets of samples for practice.



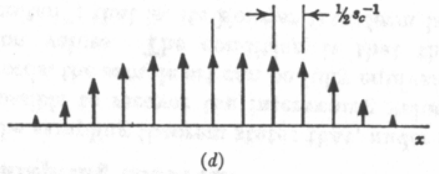


Fig. 10.2 (a) A band-limited function $f(x)$ with a cutoff spectrum (b). The samples (c) suffice to reconstitute $f(x)$ accurately in full detail and are equivalent in content to the train of impulses (d).

Consider a function $f(x)$, whose Fourier transform $F(s)$ is zero where $|s| > s_c$ (see Fig. 10.2). Evidently $f(x)$ is a band-limited function; in this case the band to which the Fourier components are limited is centered on the origin of s. Such a function is representative of a wide class of physical distributions which have been observed with equipment of limited resolving power. We shall refer to such transforms as "cutoff transforms" and describe them as being cut off beyond the "cutoff frequency" s_c.

In general a cutoff transform is of the form $\Pi(s/2s_c)G(s)$, where $G(s)$ is arbitrary, and therefore the general form of functions whose transforms are cut off is

$$\text{sinc } 2s_c x * g(x),$$

where $g(x)$ is arbitrary.

Of course, if the original function is cut off, then it is the transform which is band-limited.

With the exception noted below, band-limited functions have the peculiar property that they are fully specified by values spaced at equal intervals not exceeding $\frac{1}{2}s_c^{-1}$ (see Fig. 10.2c).

In the derivation that follows, the introduction of the *shah* symbol proves convenient, because multiplication by III(x) is equivalent to sampling, in the sense that information is retained at the sampling points and abandoned in between.

As an additional bonus, the *shah* symbol, as a result of its replicating property under convolution, enables us to express compactly the kind of repetitive spectrum that arises in sampling theory.

In the course of the argument we use the relation

$$\mathrm{III}(x) \supset \mathrm{III}(s),$$

which is discussed at greater length at the end of this chapter.

Consider the function

$$\tau^{-1}f(x)\,\mathrm{III}\left(\frac{x}{\tau}\right) = \sum_n f(n\tau)\,\delta(x - n\tau)$$

shown in Fig. 10.2c. Information about $f(x)$ is conserved only at the

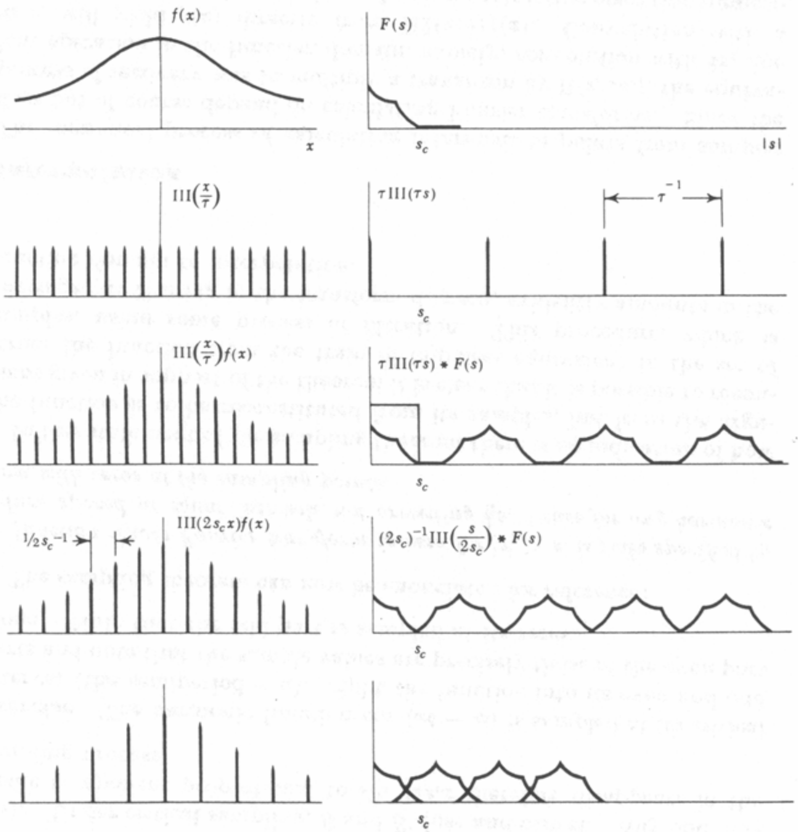


Fig. 10.3 Demonstrating the sampling theorem.

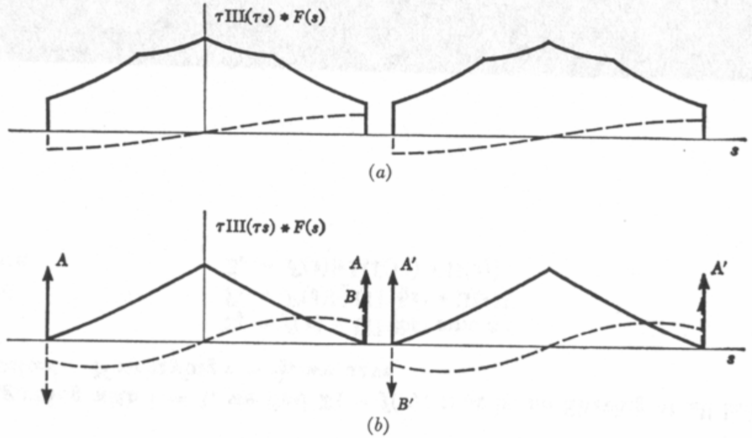


Fig. 10.4 Critical sampling.

sampling points where x is an integral multiple of the sampling interval τ. The intermediate values of $f(x)$ are lost. Therefore, if $f(x)$ can be reconstructed from $f(x)\mathrm{III}(x/\tau)$, the theorem is proved. The transform of $\mathrm{III}(x/\tau)$ is $\tau\mathrm{III}(\tau s)$, Fig. 10.3b, which is a row of impulses at spacing τ^{-1}. Therefore

$$\overline{f(x)\mathrm{III}(x/\tau)} = \tau\mathrm{III}(\tau s) * F(s),$$

and we see that multiplication of the original function by $\mathrm{III}(x/\tau)$ has the effect of replicating the spectrum $F(s)$ at intervals τ^{-1} (see Fig. 10.3c). We can reconstruct $f(x)$ if we can recover $F(s)$, and this can evidently be done, in the case illustrated in Fig. 10.3c, by multiplying $\tau\mathrm{III}(\tau s) * F(s)$ by $\Pi(s/2s_c)$. Except for cases of singular behavior at $s = s_c$ (to be considered below), this is sufficient to demonstrate the sampling theorem.

At the same time a condition for sufficiently close sampling becomes apparent, for recovery will be impossible if the replicated islands overlap as shown in Fig. 10.3e, and this will happen if the spacing of the islands τ^{-1} becomes less than the width of an island $2s_c$. Hence the sampling interval τ must not exceed $\frac{1}{2}s_c^{-1}$, the semiperiod of a sinusoid of frequency s_c, and for critical sampling we shall have the islands just touching, as in Fig. 10.3d.

A small refinement must now be considered before the sampling theorem can be enunciated with strictness. In the illustration $F(s_c)$ is shown equal to zero. If $F(s_c)$ is not zero, the islands have cliffs which, under conditions of sampling at precisely the critical interval, make butt contact. In Fig. 10.4a this is on the point of happening, but careful examination, taking into account also the imaginary part of $F(s)$, reveals that multiplication by $\Pi(s/2s_c)$ permits exact recovery of $F(s)$. However, if $F(s)$

behaves impulsively at $s = s_c$, that is, if $f(x)$ contains a harmonic component of frequency s_c, then there is more to be said.

In Fig. 10.4*b*, which shows such a case, consider first that $F(s)$ is even; that is, ignore the odd imaginary part shown dotted. Then as the sampling interval approaches the critical value, the impulses A and A' tend to fuse at $s = s_c$ into a single impulse of double strength, and multiplication by $\Pi(s/2s_c)$ taken equal to $\frac{1}{2}$ at $s = s_c$, restores the impulse to its proper value, thus permitting exact recovery of $F(s)$. The impulses A represent, of course, an even, or cosinusoidal, harmonic component: $2A \cos 2\pi s_c x$. Now consider the impulses contained in the odd part of $F(s)$. Under critical sampling, B and B' fuse and cancel. Any odd harmonic component proportional to $\sin 2\pi s_c x$ therefore disappears in the sampling process.

Exercise The harmonic function $\cos(\omega t - \phi)$ is sampled at its critical interval (the semiperiod π/ω). Split the function into its even and odd parts and note that the sample values are precisely those of the even part alone. Note that the odd part is sampled at its zeros.

The sampling theorem can now be enunciated for reference:

A function whose Fourier transform is zero for $|s| > s_c$ is fully specified by values spaced at equal intervals not exceeding $\frac{1}{2}s_c^{-1}$ save for any harmonic term with zeros at the sampling points.

In this statement of the sampling theorem there is no indication of how the function is to be reconstituted from its samples, but from the argument given in support of the theorem it is clear that it is possible to reconstruct the function from the train of impulses equivalent to the set of samples, using some process of filtration. This procedure, which is envisaged as filtering in the transform domain, evidently amounts in the function domain to interpolation.

Interpolation

The numerical process of calculating intermediate points from samples does not of course depend on calculating Fourier transforms. Since the process of recovery was to multiply a transform by $\Pi(s/2s_c)$, the equivalent operation in the function domain, namely, convolution with $2s_c \operatorname{sinc} 2s_c x$, will yield $f(x)$ directly from $\text{III}(x/\tau)f(x)$. Convolution with a function consisting of a row of impulses is an attractive operation numerically because the convolution integral reduces exactly to a summation (serial product).

For midpoint interpolation we can permanently record a table (see

Table 10.1 Midpoint interpolation

$\|x\|$	sinc x	$\|x\|$	sinc x	$\|x\|$	sinc x	$\|x\|$	sinc x
$\frac{1}{2}$	0.6366	$9\frac{1}{2}$	−0.0335	$18\frac{1}{2}$	0.0172	$27\frac{1}{2}$	−0.0116
$1\frac{1}{2}$	−0.2122	$10\frac{1}{2}$	0.0303	$19\frac{1}{2}$	−0.0163	$28\frac{1}{2}$	0.0112
$2\frac{1}{2}$	0.1273	$11\frac{1}{2}$	−0.0277	$20\frac{1}{2}$	0.0155	$29\frac{1}{2}$	−0.0108
$3\frac{1}{2}$	−0.0909	$12\frac{1}{2}$	0.0255	$21\frac{1}{2}$	−0.0148	$30\frac{1}{2}$	0.0104
$4\frac{1}{2}$	0.0707	$13\frac{1}{2}$	−0.0236	$22\frac{1}{2}$	0.0141	$31\frac{1}{2}$	−0.0101
$5\frac{1}{2}$	−0.0579	$14\frac{1}{2}$	0.0220	$23\frac{1}{2}$	−0.0135	$32\frac{1}{2}$	0.0098
$6\frac{1}{2}$	0.0490	$15\frac{1}{2}$	−0.0205	$24\frac{1}{2}$	0.0130	$33\frac{1}{2}$	−0.0095
$7\frac{1}{2}$	−0.0424	$16\frac{1}{2}$	0.0193	$25\frac{1}{2}$	−0.0125	$34\frac{1}{2}$	0.0092
$8\frac{1}{2}$	0.0374	$17\frac{1}{2}$	−0.0182	$26\frac{1}{2}$	0.0120	$35\frac{1}{2}$	−0.0090

Table 10.1) of suitably spaced values of sinc x, and it proves practical when further interpolation is required to repeat the midpoint process, using the same array.

Rectangular filtering

Suppose that it is required to remove from a function spectral components whose frequencies exceed a certain limit, that is, to multiply the transform by a rectangle function, which we shall take to be $\Pi(s)$. We are assuming that $s_c = \frac{1}{2}$ and that the critical sampling interval is 1. This is just the operation which has already been carried out for the purpose of interpolation. However, in general the function to be filtered will not consist solely of impulses, and the convolution integral giving one filtered value does not reduce exactly to a summation. However, when it is evaluated numerically it will have to be approximated by a summation,

$$\Sigma_\tau = f(x) * \left[\text{III}\left(\frac{x}{\tau}\right) \operatorname{sinc} x\right],$$

and we may ask how coarse the tabulation interval may be and still sufficiently approximate the desired integral

$$f(x) * \operatorname{sinc} x.$$

Beginning with $\tau = 1$, we find $\Sigma_1 = f(x)$; that is, no filtering at all has resulted. Now trying $\tau = \frac{1}{2}$, we have

$$\Sigma_{\frac{1}{2}} = f(x) * \text{III}(2x) \operatorname{sinc} x$$

and

$$\Sigma_{\frac{1}{2}} = F(s)[\tfrac{1}{2}\text{III}(\tfrac{1}{2}s) * \Pi(s)],$$

since

$$\Sigma_\tau = F(s)[\tau\text{III}(\tau s) * \Pi(s)].$$

Hence $\Sigma_{\frac{1}{2}}$ consists of a central part $F(s)\Pi(s)$ plus remoter parts. For many purposes this simple operation would suffice (for example, when the components to be rejected lie chiefly just beyond the central region).

Adopting an idea from the method of interpolation, where we economize on interpolating arrays by repeated use of the one midpoint interpolation array, we now consider the effect of repeated approximate filtering of the one kind. By using the same filtering array at $\tau = \frac{1}{4}$ we have the filter characteristic $\frac{1}{4}\text{III}(s/4) * \Pi(s/2)$, which when multiplied by $\frac{1}{2}\text{III}(s/2) * \Pi(s)$ gives the bottom line of Fig. 10.5. In other words, repeated application of the process has pushed down more of the outer islands of response. The same result is obtained by taking $\tau = \frac{1}{4}$ initially.

To summarize, approximate rectangular filtering with a cutoff at s_c is carried out by reading off $f(x)$ at half the critical sampling interval (that is, at intervals of $\frac{1}{4}s_c^{-1}$) and taking the convolution (more precisely, the serial product) with $2s_c \operatorname{sinc} 2s_cx$, where $2s_cx$ assumes all half-integral values, including 0. This filtering array (see Table 10.2) contains precisely the values tabulated for interpolation plus interleaved zeros and a central value of unity.

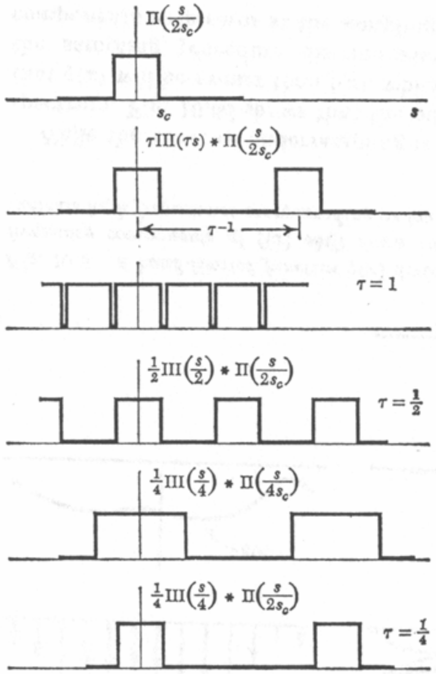


Fig. 10.5 Numerical procedure for achieving the desired filter characteristic $\Pi(s)$.

Table 10.2 Array for approximate rectangular filtering

.
.
.
−0.0909
0
0.1273
0
−0.2122
0
0.6366
⟶1.0000
0.6366
0
−0.2122
0
0.1273
0
−0.0909
.
.
.

Undersampling

Suppose that $f(x)$ (see Fig. 10.6*a*) is read off at intervals corresponding to a desired cutoff for rectangular filtering. Then the band-limited function $g(x)$ (see Fig. 10.6*c*) defined by this set of values (see Fig. 10.6*b*) has a cutoff spectrum of the desired extent and may at first sight appear to be a product of rectangular filtering. But the process is not the same as rectangular filtering, since the result depends on high-frequency components in $f(x)$; for example, one of the sample values may fall at the peak of a narrow spike; furthermore, the *phase* of the coarse sampling points will clearly affect the result. However, the effect may often be a good approximation to rectangular filtering.

By examining the process in terms of Fourier transforms, we see that the band-limited function $g(x)$ derived from too-coarse sampling contains contributions from high-frequency components of $f(x)$, impersonating low frequencies in a way described by reflection of the high-frequency part in the line $s = s_c$. The effect has been referred to as "aliasing." If this high-frequency tail is not too important, then the coarse sampling procedure gives a fair result.

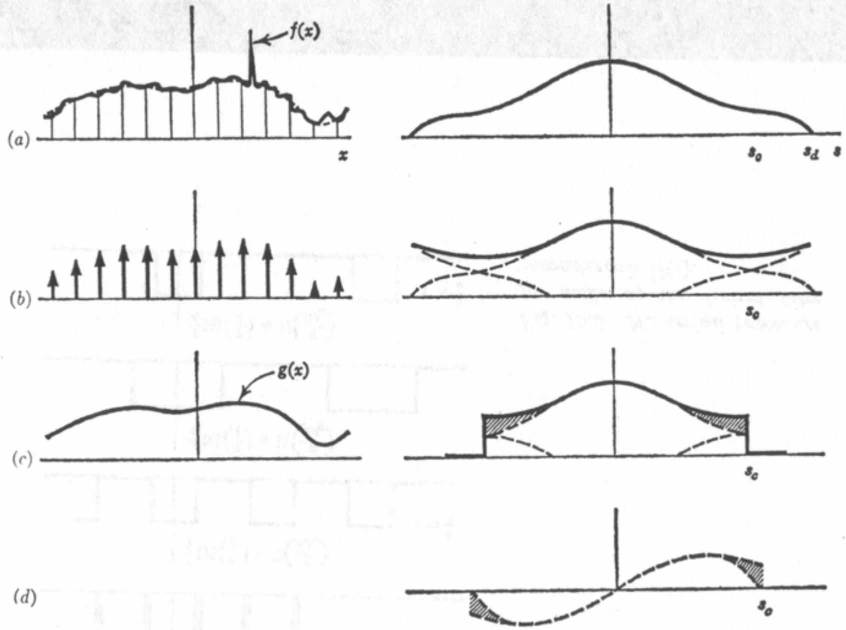


Fig. 10.6 A band-limited function $g(x)$ derived from $f(x)$ by undersampling; high-frequency components of $f(x)$ shift down inside the band limits. Shaded areas indicate high frequencies masquerading as low.

While the effect of undersampling is to reinforce the even part of the spectrum, Fig. 10.6*d* shows that the odd part is diminished. It follows that $g(x)$ will be evener than $f(x)$, which is, of course, to be expected, for the sampling procedure discriminates against the (necessarily odd) components with zeros at the sampling points.

Ordinate and slope sampling

Let $f(x)$ be a band-limited function that is fully specified by ordinates at a spacing of 0.5; but suppose that only IIIf is given; that is, only every second ordinate is given. Then from the overlapping islands (see Fig. 10.7) composing $\overline{\text{III}f}$, it would not be possible to recover $F(s)$ or, consequently, $f(x)$. But if further partial information were given, it might become possible.

If the slope is given, in addition to the ordinate, recovery proves to be possible. Thus, given IIIf and IIIf', one can express $f(x)$ as a combina-

tion of linear functionals of IIIf and IIIf':

$$f(x) = a(x) * (\text{III}f) + b(x) * (\text{III}f'),$$

where $a(x)$ and $b(x)$ are solving functions that have to be found. Just as the sinc function, which is the solving function for ordinary sampling, must be zero at all its sample points save the origin, where it must be unity, so must $a(x)$. And $b(x)$ must be zero at all sample points but have unit slope at the origin.

To find $a(x)$ and $b(x)$, note that in the interval $-1 \leqslant s \leqslant 1$

$$\overline{\text{III}f} = F(s+1) + F(s) + F(s-1)$$

and that

$$\overline{\text{III}f'} = i2\pi(s+1)F(s+1) + i2\pi sF(s) + i2\pi(s-1)F(s-1).$$

These two equations can be solved for $F(s)$, for although there appear to be three unknowns, namely, $F(s+1)$, $F(s)$, $F(s-1)$, in fact, for any value of s, one of them is always known to be zero. Thus for positive s we have $F(s+1) = 0$, and on eliminating $F(s-1)$ we have

$$i2\pi F(s) = \overline{\text{III}f'} - i2\pi(s-1)\overline{\text{III}f},$$

and for negative s we have

$$-i2\pi F(s) = \overline{\text{III}f'} - i2\pi(s+1)\overline{\text{III}f}.$$

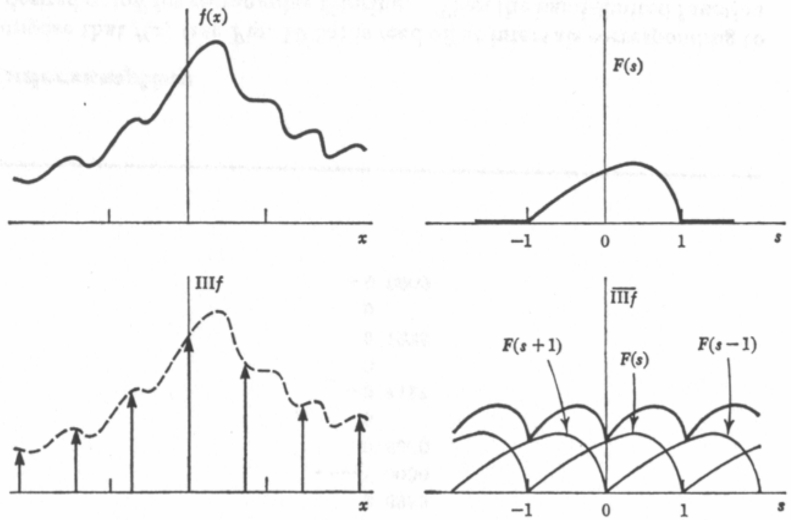


Fig. 10.7 Ordinate sampling at half the rate necessary for full definition.

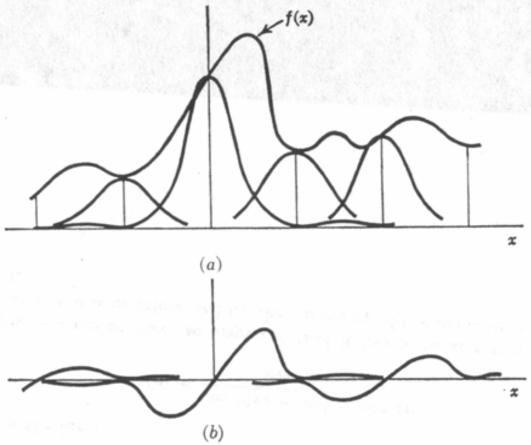


Fig. 10.8 (a) The ordinate-dependent constituents $f(n)a(x-n)$; (b) The slope-dependent constituents $f'(n)b(x-n)$.

For all s, both positive and negative, we have concisely

$$F(s) = \frac{i}{2\pi}\Lambda'(s)\overline{\mathrm{III}f'} + \Lambda(s)\overline{\mathrm{III}f}.$$

Hence $$f(x) = \operatorname{sinc}^2 x * (\mathrm{III}f) + x \operatorname{sinc}^2 x * (\mathrm{III}f').$$

The solving functions are

$$a(x) = \operatorname{sinc}^2 x$$
$$b(x) = x \operatorname{sinc}^2 x,$$

and the convolution integrals reduce to a sum of spaced a's and b's of suitable amplitudes,

$$\begin{aligned} f(x) &= \sum_{-\infty}^{\infty} f(n)a(x-n) + \sum_{-\infty}^{\infty} f'(n)b(x-n) \\ &= \sum_{-\infty}^{\infty} f(n)\operatorname{sinc}^2(x-n) + \sum_{-\infty}^{\infty} f'(n)(x-n)\operatorname{sinc}^2(x-n). \end{aligned}$$

We see that each $a(x-n)$ is zero at all sampling points (integral values of x) save where $x = n$, and that each has zero slope at all sampling points (see Fig. 10.8a). Each $b(x-n)$ has zero ordinate at all sampling points and likewise zero slope, save where $x = n$ (see Fig. 10.8b).

Interlaced sampling

As in the previous example, let IIIf represent every second ordinate of the set that is necessary to specify $f(x)$, and let a supplementary set $\mathrm{III}(x-a)f(x)$, interlaced with the first as illustrated in Fig. 10.9, also be available. Will it be possible to reconstitute $f(x)$? It is known that equispaced samples, separated by just more than the critical interval, do not suffice, and in the case of interlaced sampling every second jump exceeds this critical interval. On the other hand, it has been shown that ordinate-and-slope sampling suffices, and this is clearly equivalent to extreme interlacing as a approaches zero.

If there is a solution, it should be in the form of a sum of two linear functionals of IIIf and $\mathrm{III}_a f$, where $\mathrm{III}_a \equiv \mathrm{III}(x-a)$:

$$f(x) = a(x) * (\mathrm{III}f) + b(x) * (\mathrm{III}_a f).$$

The solving function $a(x)$ must be equal to unity at $x = 0$, and zero at all other sampling points, and $b(x)$ must be the mirror image of $a(x)$; that is, $b(x) = a(-x)$.

In the interval $-1 < s < 1$,

$$\overline{\mathrm{III}f} = F(s+1) + F(s) + F(s-1)$$
and
$$\overline{\mathrm{III}_a f} = e^{i2\pi a}F(s+1) + F(s) + e^{-i2\pi a}F(s-1).$$

For positive s we have $F(s+1) = 0$, and on eliminating $F(s-1)$ we have

$$F(s) = -\frac{e^{-i2\pi a}}{1-e^{-i2\pi a}}\overline{\mathrm{III}f} + \frac{1}{1-e^{-i2\pi a}}\overline{\mathrm{III}_a f},$$

and for negative s we have

$$F(s) = -\frac{e^{i2\pi a}}{1-e^{i2\pi a}}\overline{\mathrm{III}f} + \frac{1}{1-e^{i2\pi a}}\overline{\mathrm{III}_a f}.$$

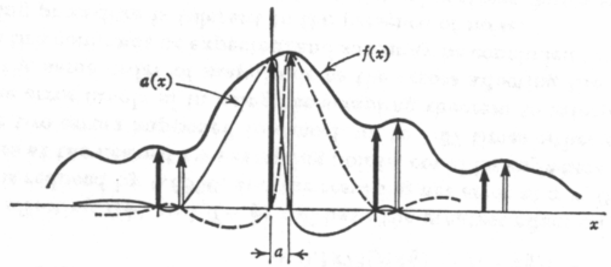


Fig. 10.9 Interlaced samples.

For all s we have

$$F(s) = A(s)\overline{\text{III}f} + A^*(s)\overline{\text{III}_a f},$$

where
$$A(s) = \begin{cases} -\dfrac{e^{-i2\pi a}}{1 - e^{-i2\pi a}} & 0 < s < 1 \\ -\dfrac{e^{i2\pi a}}{1 - e^{i2\pi a}} & -1 < s < 0 \\ 0 & |s| > 1 \end{cases}$$

$$= \tfrac{1}{2}\Pi\left(\frac{s}{2}\right) + \tfrac{1}{2}i \cot a\pi \Lambda'(s).$$

Hence $$a(x) = \operatorname{sinc} 2x - (\pi \cot a\pi)x \operatorname{sinc}^2 x,$$

as graphed in Fig. 10.9.[1]

It may seem strange that the equidistant samples may be regrouped in pairs, even to the extreme of close spacing. However, it is also possible to bunch the samples in groups of any size separated by such wide intervals as maintain the original average spacing. The bunching may be indefinitely close; see Linden for a proof that the ordinates and first n derivatives, at points spaced by $n + 1$ times the usual spacing, suffice to specify a band-limited function. In the limit, as n approaches infinity, the formula for reconstituting the function becomes the Maclaurin series. This introduces doubt of the practical applicability of higher-order sampling theorems, for it is well known that the Maclaurin series

$$f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \cdots + \frac{x^n}{n!}f^{(n)}(0)$$

does not usually converge to $f(x)$. (Consider the functions $\Pi(ax)$, which all have the same Maclaurin series.)

In practice, higher-order sampling breaks down at some point because small amounts of noise drastically affect the determination of high-order derivatives or finite differences. In the total absence of noise, trouble would still be expected to set in at some stage because of the impossibility of ensuring perfectly band-limited behavior. In applications of sampling theorems, the claim of a given function to be band-limited must always be scrutinized, and any error resulting must be estimated.

[1] For a discussion of sampling theorems see D. A. Linden, A Discussion of Sampling Theorems, *Proc. IRE.*, vol. 47, p. 1219, July, 1959. In this paper the solving function $a(x)$ is given in the form

$$a(x) = \frac{\cos(2\pi x - a\pi) - \cos a\pi}{2\pi x \sin a\pi}.$$

The numerator is a cosine wave so displaced that it has zeros at $x = n$ and $x = n + a$, but the zero at $x = 0$ is counteracted by the vanishing of the denominator in such a way that $a(0) = 1$.

Sampling in the presence of noise

Suppose that the samples $f(n)$ of a certain function $f(x)$ cannot be obtained without some error being made; that is, the observable quantity is

$$f(n) + \text{error}.$$

Then when an attempt is made to reconstruct $f(x)$ from the observed samples by applying the same procedure used for true samples, the reconstructed values will differ from the true values of $f(x)$.

Consider the case of midpoint interpolation, supposing that samples have been taken at $x = \pm\frac{1}{2}, \pm 1\frac{1}{2}, \pm 2\frac{1}{2}, \ldots$. Then

$$f(0) = 0.6366[f(\tfrac{1}{2}) + f(-\tfrac{1}{2})] - 0.2122[f(1\tfrac{1}{2}) + f(-1\tfrac{1}{2})] + 0.1273[f(2\tfrac{1}{2}) + f(-2\tfrac{1}{2})] - \cdots .$$

The errors affecting $f(\frac{1}{2})$ and $f(-\frac{1}{2})$ will have the greatest effect on $f(0)$; each error is reduced by 0.6366, and the resulting net error at $x = 0$, due to the errors at the nearest two sampling points, could be anywhere from zero, if the two errors happened to cancel, up to 1.27 times either error. Clearly, the error involved in using the sampling theorem to interpolate will be of the same order of magnitude as the errors affecting the data. More than this could not be expected, and so it may be concluded that the interpolating procedure is tolerant to the presence of noise.

It is not the purpose here to go into statistical matters, but a simple result should be pointed out that arises when the errors are of such a nature that they are independent from one sample to the next, and all come from a population with zero mean value and a certain variance σ^2. Then the variance of the error contributed at $x = 0$ by the error at $x = \frac{1}{2}$ is $(0.6366)^2\sigma^2$, and the variance of the total error at $x = 0$ due to the errors at all the sampling points is

$$[\cdots + (0.1273)^2 + (0.2122)^2 + (0.6366)^2 + (0.6366)^2 + (0.2122)^2 + (0.1273)^2 + \cdots]\sigma^2.$$

Now the terms of the series within the brackets are values of $\operatorname{sinc}^2 x$ at unit intervals of x and hence add up to unity. Therefore, in this simple error situation, the interpolated value is subject to precisely the same error as the data.

Now we apply the same reasoning to interlaced sampling, especially to the extreme situation where the narrow interval is small compared with the wide one, that is, where $a \ll 1$. At the point $x = \frac{1}{2}a + \frac{1}{2}$, which is in the middle of the wide interval, the four nearest sample values enter with coefficients

$$a(\tfrac{1}{2}a + \tfrac{1}{2}),\ b(-\tfrac{1}{2}a + \tfrac{1}{2}),\ a(\tfrac{1}{2}a - \tfrac{1}{2}),\ b(-\tfrac{1}{2}a - \tfrac{1}{2}).$$

The first and last of these are positive and the others negative, and the presence of the factor cot $a\pi$ in the formula for $a(x)$ shows that the numerical values may be large. Thus the interpolated value may result from the cancellation of large terms, and the total error may be large.

In another way of looking at this, a pair of terms

$$f(0)a(x) + f(a)b(x - a)$$

may be reexpressed in the form

$$[a(x) + b(x - a)]\left[\frac{f(0) + f(a)}{2}\right] + \frac{b(x - a) - a(x)}{2}[f(a) - f(0)].$$

Here the first term represents the mean of a sample pair multiplied by a certain coefficient, and the second represents the difference between two close-spaced samples multiplied by a certain other coefficient. In the limit as $a \to 0$, the solving functions for the ordinate-and-slope sampling theorem would result. Now the coefficient of the difference term can be large; for example, when $a < 0.2$, the value adopted in the illustration of interlaced sampling, the coefficient exceeds unity. Thus errors in the difference term may be amplified.

It thus appears that interlaced sampling, where the sample spacing is alternately narrow and wide, is not tolerant to errors, and therefore the magnitude of the errors would have to be estimated carefully in an application. In a full study it would be essential to take account of any correlation between the errors in successive samples since it is clear that the error in $f(a) - f(0)$ would be reduced if both $f(a)$ and $f(0)$ were subject to about the same error.

Fourier series

It is well known that a periodic waveform, such as the acoustical waveform associated with a sustained note of a musical instrument, is composed of a fundamental and harmonics. Exploration of such an acoustical field by means of tunable resonators reveals that the energy is concentrated at frequencies which are integral multiples of the fundamental frequency. There is nothing here that should exclude this case from treatment by the Fourier transform methods so far used. However, insistence on strict periodicity, a physically impossible thing, will clearly lead to an impulsive spectrum and to the refined considerations that are needed in connection with impulses. We now proceed to do this, using the *shah* symbol for convenient handling of the sets of impulses that arise, in connection both with the replication inherent in periodicity and with the sampling associated with harmonic spectra.

The Fourier series will be exhibited as an extreme situation of the Fourier transform, even though the opposite procedure, taking the Fourier series as a point of departure for developing the Fourier transform, is more usual.

For reference let it be stated that the Fourier series associated with the periodic function $g(x)$, with frequency f and period T, is

$$a_0 + \sum_1^\infty (a_n \cos 2\pi n f x + b_n \sin 2\pi n f x),$$

where
$$a_0 = \frac{1}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(x)\,dx$$
$$a_n = \frac{2}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(x)\cos 2\pi n f x\,dx$$
$$b_n = \frac{2}{T}\int_{-\frac{1}{2}T}^{\frac{1}{2}T} g(x)\sin 2\pi f x\,dx.$$

It is necessary for $g(x)$ to have been chosen so that the integrals exist; otherwise $g(x)$ is arbitrary.

The purpose of a good deal of theory dealing with the Fourier series has been to show that the series associated with a periodic function $g(x)$ does in fact often converge, and furthermore, that when it converges, it often converges to

$$\tfrac{1}{2}[g(x + 0) + g(x - 0)].$$

The rigorous development of this topic was initiated by Dirichlet in 1829, following a controversial period dating back to D. Bernoulli's success in 1753 in expressing the form of a vibrating string as a series

$$y = A_1 \sin x \cos at + A_2 \sin 2x \cos 2at + \cdots.$$

Euler, who had been working on this problem and had just obtained the general solution in terms of traveling waves, said that if Bernoulli was right, an arbitrary function could be expanded as a sine series. This, he said, was impossible. In 1807, when Fourier made this same claim in his paper presented to the Paris Academy, Lagrange rose and said it was impossible. This exciting subject led to many important developments in pure mathematics, including the invention of the Riemann integral. It must be remembered that the expressions for the Fourier constants a_0, a_n, b_n were given long before modern analysis developed.

For the present purpose let us take $T = 1$ and note that the complex constant $a_n - ib_n$ is related to the one-period segment $g(x)\Pi(x)$ by the Fourier transform

$$a_n - ib_n = 2\int_{-\frac{1}{2}}^{\frac{1}{2}} g(x)e^{-i2\pi n x}\,dx$$
$$= 2\int_{-\infty}^{\infty} g(x)\Pi(x)e^{-i2\pi n x}\,dx.$$

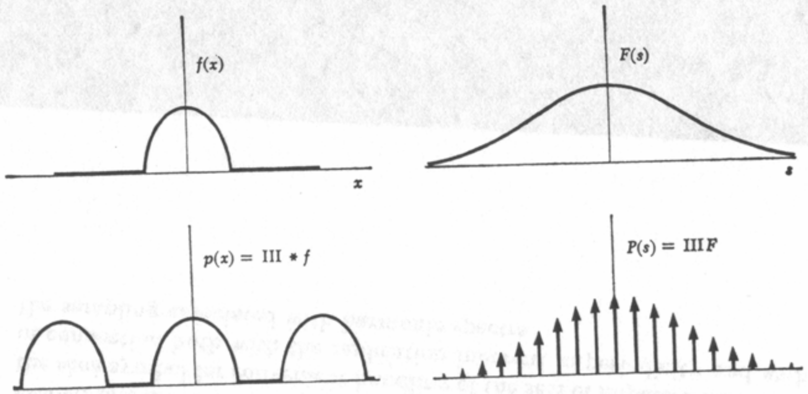


Fig. 10.10 The transform in the limit of a periodic function.

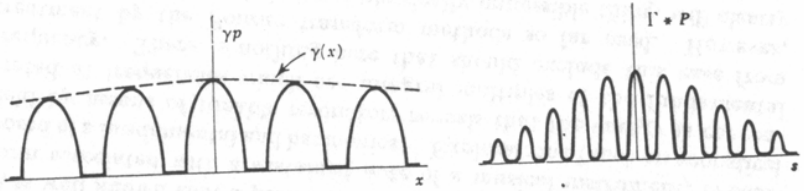


Fig. 10.11 The convergence factor γ applied to a periodic function p(x) renders its infinite integral convergent while the convolvent Γ removes infinite discontinuities from a line spectrum P(s).

We now recover this result by considering the Fourier transform of a periodic function.

Let $f(x)$ be a function that possesses a regular Fourier transform $F(s)$ (see Fig. 10.10). Then its convolution with the replicating symbol $\mathrm{III}(x)$ will be the periodic function $p(x)$, defined by

$$p(x) = \mathrm{III}(x) * f(x) = \sum_{-\infty}^{\infty} f(x-n) \qquad n \text{ integral.}$$

The convergence of the summation is guaranteed by the absolute integrability of $f(x)$, which was requisite for the possession of a Fourier transform. The period of $p(x)$ is unity; that is,

$$p(x+1) = p(x)$$

for all x.

There will be no regular Fourier transform for $p(x)$ because

$$\int_{-\infty}^{\infty} |p(x)|\, dx$$

cannot converge, save in the trivial case where $p(x)$ is identically zero. Hence we multiply $p(x)$ by a factor $\gamma(x)$ that dies out to zero for large values of x, both positive and negative. In effect we are bringing the strictly periodic function $p(x)$ back to the realm of the physically possible, but only barely so (see Fig. 10.11). Let

$$\gamma(x) = e^{-\pi\tau^2x^2};$$

then the Fourier transform $\Gamma(s)$ of $\gamma(x)$ will be given by

$$\Gamma(s) = \tau^{-1}e^{-\pi s^2/\tau^2}.$$

The function $\gamma(x)p(x)$ will possess a Fourier transform,

$$\begin{aligned}\gamma(x)p(x) &= \gamma(x)\sum_{-\infty}^{\infty} f(x-n)\\ &\supset \Gamma(s) * \sum_{-\infty}^{\infty} e^{-i2\pi ns}F(s)\\ &= \sum_{-\infty}^{\infty} F(n)\Gamma(s-n).\end{aligned}$$

In *shah*-symbol notation,

$$\sum_{-\infty}^{\infty} F(n)\Gamma(s-n) = \Gamma(s) * [\mathrm{III}(s)F(s)].$$

Now let $$P(s) = \mathrm{III}(s)F(s).$$

This entity $P(s)$ is a whole set of equidistant impulses of various strengths, as given by samples of $F(s)$ at integral values of s, and has the property that

$$\gamma p \supset \Gamma * P.$$

By the convolution theorem we see that P is a suitable symbol to assign as the Fourier transform of $p(x)$, and indeed as $\tau \to 0$ and γp runs through a sequence of functions having p as limit, the sequence $\Gamma * P$ defines an entity P of such a type that, as agreed, we may call $p(x)$ and $P(s)$ a Fourier transform pair in the limit.

Taking $f(x)$ to be the one-period segment $g(x)\Pi(x)$, we see that the spectrum of a periodic function is a set of impulses whose strengths are given by equidistant samples of $F(s)$, the Fourier transform of the one-period segment. Now

$$\begin{aligned}\int_{-\infty}^{\infty} \text{III}(s)F(s)e^{+i2\pi sx}\,ds &= \sum_{-\infty}^{\infty}\int_{n-0}^{n+0} F(s)e^{+i2\pi sx}\,ds \\ &= \sum_{-\infty}^{\infty} F(n)e^{+i2\pi nx} \\ &= F(0) + \sum_{1}^{\infty}[F(n)e^{+i2\pi nx} + \text{conjugate}] \\ &= F(0) + 2\sum_{1}^{\infty}(\text{Re}\,F\cos 2\pi nx - \text{Im}\,F\sin 2\pi nx) \\ &= a_0 + \sum_{1}^{\infty}(a_n\cos 2\pi nx + b_n \sin 2\pi nx)\end{aligned}$$

if $a_n - ib_n = 2F(n)$. And this is precisely the value that was quoted earlier for the complex coefficient $a_n - ib_n$.

The fact that rigorous deliberations on Fourier series generally are more complex than those encountered with the Fourier integral is essentially connected with the infinite energy of periodic functions (the integrals of which are not absolutely convergent). It is therefore very natural physically to regard a periodic function as something to be approached through functions having finite energy, and to consider *line* spectra as something to be approached via continuous spectra with finite energy density. The strange thing is that the physically possible functions and spectra are often presented as elaborations of the physically impossible. Some people cannot see how a *line* spectrum, no matter how closely the lines are packed, can ultimately become a continuous spectrum. This order of presentation is inherited from the historical precedence of the theory of trigonometric series, and runs as follows (in our terminology). The periodic function $\text{III} * f$ has a line spectrum $\text{III}F$ (set of Fourier coefficients). If the repetition period is lengthened to τ, the lines of the spectrum are packed τ times more closely and are τ times weaker (note compensating change in ordinate scale in Fig. 10.12). Now let the period become infinite, so that the pulse f does not recur. Then the trigonometric sum which represented its periodic predecessors passes into an infinite integral, and the finite integral which specified the series coefficients does likewise. These two integrals are the plus-i and minus-i Fourier integrals.

On the view described here line spectra and periodic functions are regarded as included in the theory of Fourier transforms, to be handled

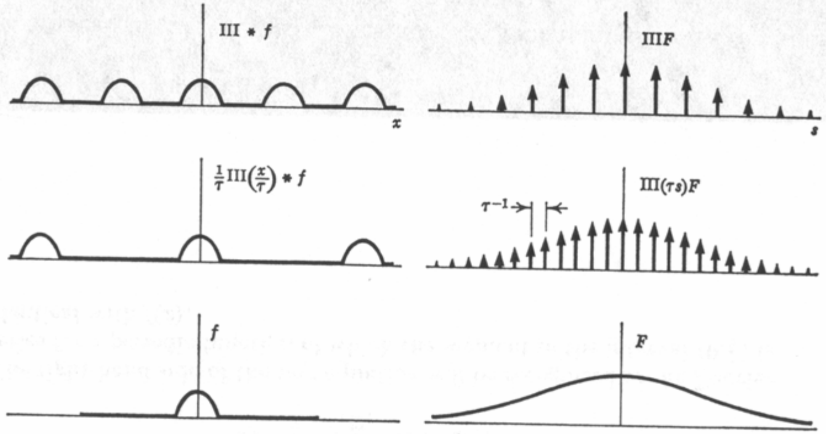


Fig. 10.12 Deriving the Fourier integral from the Fourier series.

exactly as other transforms by means of the III symbology, and with the same caution accorded to other transforms in the limit.

Gibbs phenomenon One of the classical topics of the theory of Fourier series may be studied profitably from the point of view that we have developed. In situations where periodic phenomena are analyzed to determine the coefficients of a Fourier series, which is then used to predict, it is a matter of practical importance to know how many terms of the series to retain. Various considerations enter into this, but one of them is the phenomenon of overshoot associated with discontinuities, or sharp changes, in the periodic function to be represented.

It is quite clear that by omitting terms beyond a certain limiting frequency we are subjecting the periodic function $p(x)$ to low-pass filtering. Thus, if the fundamental frequency is s_0, and frequencies up to ns_0 are retained, it is as though the spectrum had been multiplied by a rectangle function $\Pi(s/2s_c)$, where s_c is a cutoff frequency between ns_0 and $(n+1)s_0$. It makes no difference precisely where s_c is taken; for convenience it may be taken at $(n+\frac{1}{2})s_0$. Multiplication of the spectrum by $\Pi[s/(2n+1)s_0]$ corresponds to convolution of the original periodic function $p(x)$ with $(2n+1)s_0 \operatorname{sinc}[(2n+1)s_0x]$. Therefore, when the series is summed for terms up to frequencies ns_0 only, the sum will be

$$p(x) * (2n+1)s_0 \operatorname{sinc}[(2n+1)s_0x].$$

The convolving function has unit area, so in places where $p(x)$ is slowly varying, the result will be in close agreement with $p(x)$.

We now wish to study what happens at a discontinuity, and so we choose a periodic function which is equal to sgn x for a good distance to each side of $x = 0$ (see Fig. 10.13). What it does outside this range will not matter, as long as it is periodic, because we are going to focus attention on what happens near $x = 0$. Near $x = 0$ the result will be approximately

$$(2n + 1)s_0 \operatorname{sinc} [(2n + 1)s_0 x] * \operatorname{sgn} x.$$

We know that

$$\operatorname{sinc} x * \operatorname{sgn} x = 2 \int_0^x \operatorname{sinc} t \, dt,$$

a function closely related to the sine integral Si(x). In fact

$$2 \int_0^x \operatorname{sinc} t \, dt = \frac{2}{\pi} \operatorname{Si}(\pi x).$$

This function oscillates about -1 for large negative values of x, oscillates with increasing amplitude as the origin is approached, passes through zero at $x = 0$, shoots up to a maximum value of 1.18, and then settles down to decaying oscillations about a value of $+1$. If we change the scale factors of the sinc function, compressing it by a factor $N = (2n + 1)s_0$, and

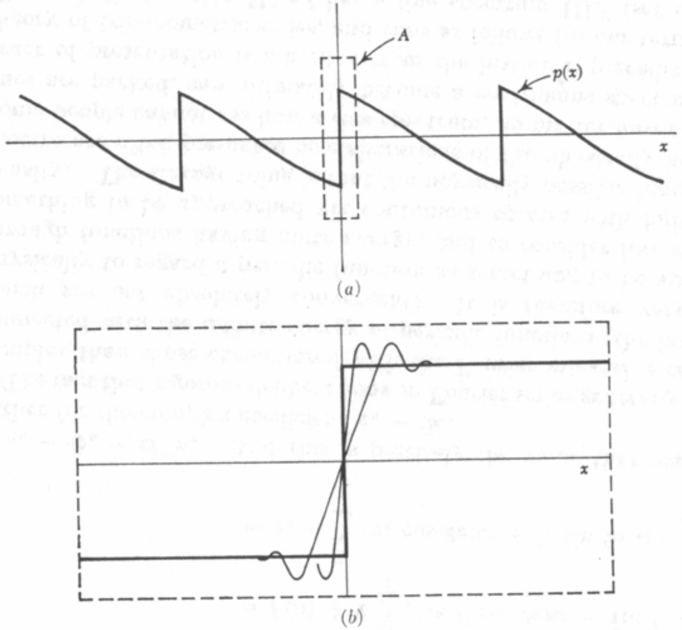


Fig. 10.13 (a) A periodic function $p(x)$; (b) an enlargement of area A.

strengthening it by a factor N, so as to retain its unit area, then convolution with sgn x will result in oscillations about -1 and then about $+1$ that are faster but have the same amplitude, that is,

$$N \operatorname{sinc} Nx * \operatorname{sgn} x = \frac{2}{\pi} \operatorname{Si} (N\pi x).$$

The overshoot, amounting to 9 per cent of the amount of the discontinuity, remains at 9 per cent, but the maximum is reached nearer to the discontinuity. The same applies to the minimum that occurs on the negative side.

Now we see precisely what happens when a Fourier series is truncated. There is overshoot on both sides of any discontinuity, amounting to about 9 per cent, regardless of the inclusion of more and more terms. At any given point to one side of the discontinuity the oscillations decrease indefinitely in amplitude as N increases, so that in the limit the sum of the series approaches the value of the function of which it is the Fourier series (and at the point of discontinuity, the sum of the series approaches the midpoint of the jump). In spite of this, the maximum departure of the sum of the series from the function remains different from zero, and as the maximum moves in close to the step, it approaches the precise value of 9 per cent that was derived for sgn x, because the parts of the function away from the discontinuity have indeed become irrelevant.

This behavior is reminiscent of that of $x\,\delta(x)$, which is zero for all x, even though sequences defining it have nonvanishing maxima and minima.

Finite Fourier transforms In problems where the range of the independent variable is not from $-\infty$ to ∞, advantages accrue from the introduction of finite transforms; for example,

$$F(s,a,b) = \int_a^b f(x) e^{-i2\pi xs} \, dx.$$

With such a definition one can work out an inversion formula, a convolution theorem, theorems for the finite transform of derivatives of functions, and so on. As a particular example of an inversion formula we have

$$F_s(s,0,\tfrac{1}{2}) = \int_0^{\frac{1}{2}} f(x) \sin 2\pi xs \, dx,$$

$$f(x) = 4 \sum_{s=1}^{\infty} F_s(s,0,\tfrac{1}{2}) \sin 2\pi xs.$$

The right-hand side of the last equation will be recognized as the Fourier series for a periodic function of which the segment in the interval $(0,\frac{1}{2})$ is identical with $f(x)$.

It will be evident that the theory of finite transforms will be the same as the theory of Fourier series and that the principal advantage of their use will lie in the approach. We have seen the convenience of embracing Fourier series within the scope of Fourier transforms through the concept of transforms in the limit, and we shall therefore also include finite transforms. Thus we may write the foregoing example in terms of ordinary sine transforms as follows:

$$\int_0^{\frac{1}{2}} f(x) \sin 2\pi xs \, dx = 2 \int_0^{\infty} [\tfrac{1}{2}\Pi(x)f(x)] \sin 2\pi xs \, dx$$

and in the general case

$$\int_a^b f(x)e^{-i2\pi xs} \, dx = \int_{-\infty}^{\infty} \left[\Pi\left(\frac{x - \frac{1}{2}(b + a)}{b - a}\right) f(x) \right] e^{-i2\pi xs} \, dx.$$

In other words, we substitute for integration over a finite range infinite integration of a function which is zero outside the old integration limits.

All the special properties of finite transforms then drop out. For example, the derivative of a function will (in general) be impulsive at the points a and b where it cuts off, and therefore the Fourier transform of the derivative will contain two special terms proportional to the jumps at a and b. It is not necessary to make explicit mention of this property of the transformation when stating the theorem that the Fourier transform of the derivative of a function is $i2\pi s$ times the transform of the function; for example, the Fourier transform of $\Pi'(x)$ is $i2\pi s.\text{sinc } s = 2i \sin \pi s$. However, the derivative theorem for finite transforms contains these additive terms explicitly. Thus

$$\int_a^b f'(x)e^{-i2\pi xs} \, dx = i2\pi sF(s,a,b) + f(a)e^{-i2\pi as} - f(b)e^{-i2\pi bs}.$$

Fourier coefficients If we consider the usual formula for the series coefficients a_n and b_n for a periodic function $p(x)$ of unit period, namely,

$$a_n - ib_n = 2 \int_{-\frac{1}{2}}^{\frac{1}{2}} p(x)e^{-i2\pi nx} \, dx,$$

we note that the integral has the form of a finite Fourier transform. Thus in spite of the fact that $p(x)$ is a function of a continuous variable x, whereas the Fourier series coefficients depend on a variable n which can assume only integral values, Fourier transforms as we have been studying them enter directly into the determination of series coefficients. The finite transform can, we know, be expressed as a standard transform of a slightly different function $\Pi(x)p(x)$ as follows:

$$\int_{-\infty}^{\infty} \Pi(x)p(x)e^{-i2\pi nx} \, dx.$$

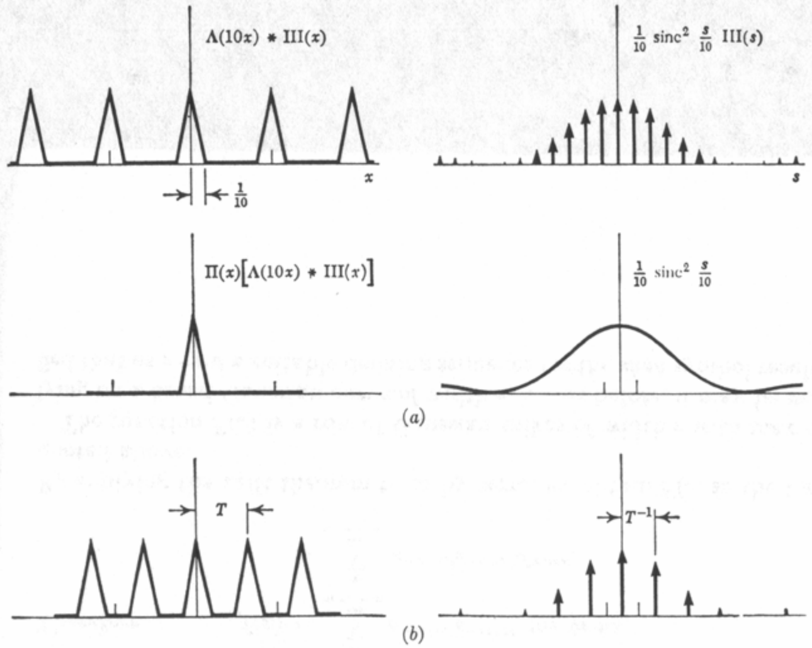


Fig. 10.14 Obtaining Fourier series coefficients.

Our ability to handle transforms can thus be freely applied to the determination of Fourier series coefficients.

As an example consider a periodic train of narrow triangular pulses shown in Fig. 10.14a,

$$\sum_{n=-\infty}^{\infty} \Lambda[10(x - n)].$$

With the aid of the *shah* notation this train can also be expressed in the form

$$\Lambda(10x) * \text{III}(x),$$

whose transform is

$$\frac{1}{10} \text{sinc}^2 \frac{s}{10} \text{III}(s).$$

We note that the $\frac{1}{10}$ sinc² $(s/10)$ part of this expression came from

$$\int_{-\infty}^{\infty} \Lambda(10x)e^{-i2\pi sx} \, dx,$$

which, in terms of the periodic train $\Lambda(10x) * \text{III}(x)$, is the same as

$$\int_{-\infty}^{\infty} \Pi(x)[\Lambda(10x) * \text{III}(x)]e^{-i2\pi sx} \, dx.$$

Thus by taking the Fourier transform of a single pulse of the train we obtain precisely the expression which arose in connection with the series. It is true that s is a continuous variable while n is not, but

$$\frac{1}{10}\operatorname{sinc}^2\frac{s}{10}\,\text{III}(s)$$

is zero everywhere save at discrete values of s, and at these values the strength of the impulse is equal to the corresponding series coefficient.

In the general case of a function

$$f(x) * \frac{1}{T}\,\text{III}\left(\frac{x}{T}\right)$$

of any period T, the transform

$$F(s)\,\text{III}(Ts)$$

shows that the coefficients are obtained by reading off the same $F(s)$ at different intervals $s = T^{-1}$. This is illustrated in Fig. 10.14*b*.

The shah symbol is its own Fourier transform

The *shah* symbol $\text{III}(x)$ is defined by

$$\text{III}(x) = \sum_{-\infty}^{\infty} \delta(x-n)$$

and therefore, in accordance with the approach being adopted here, is to be considered in terms of defining sequences. If, as asserted, its Fourier transform proves to be $\text{III}(s)$, then it too will be considered in terms of sequences.

We proceed therefore to construct a sequence of regular Fourier transform pairs of ordinary functions such that one sequence is suitable for defining $\text{III}(x)$, and we then see whether the other sequence defines $\text{III}(s)$.

Consider the function

$$f(x) = \tau^{-1}e^{-\pi\tau^2x^2}\sum_{n=-\infty}^{\infty} e^{-\pi\tau^{-2}(x-n)^2}.$$

For a given small value of τ (which we shall later allow to vary to generate a sequence), the function $f(x)$ represents a row of narrow Gaussian spikes of width τ, the whole multiplied by a broad Gaussian envelope of width τ^{-1} (see Fig. 10.15). As $\tau \to 0$, each spike narrows in on an integral value of x and increases in height. For any value of x not equal to an integer, we can show that $f(x) \to 0$ as $\tau \to 0$ and, in addition, the area under each

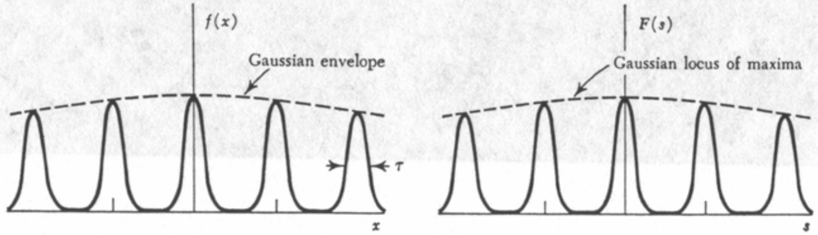


Fig. 10.15 A transform pair for discussing the shah *symbol.*

spike approaches unity. The sequence is therefore a suitable one for defining a set of equal unit impulses situated at integral values of x.

The function $f(x)$ possesses a regular Fourier transform, since $|f(x)|$ is integrable, and there are no discontinuities. The Fourier transform $F(s)$ is given by

$$F(s) = \tau^{-1}\sum_{m=-\infty}^{\infty} e^{-\pi\tau^2m^2}e^{-\pi\tau^{-2}(s-m)^2}.$$

One way of establishing this is to note that the factor

$$\tau^{-1}\sum_{n=-\infty}^{\infty} e^{-\pi\tau^{-2}(x-n)^2}$$

is periodic, with unit period, and hence may be expressed as a Fourier series. The theory of Fourier series is a well-established branch of mathematics that we are entitled to draw upon here, as long as we do not attempt to make the self-transforming property of the *shah* symbol a basis for reestablishing the theory of Fourier series. The Fourier series is

$$\sum_{m=-\infty}^{\infty} e^{-\pi\tau^2m^2}\cos 2\pi mx.$$

Therefore

$$f(x) = \sum_{m=-\infty}^{\infty} e^{-\pi\tau^2m^2}e^{-\pi\tau^2x^2}\cos 2\pi mx$$
$$= \sum_{m=-\infty}^{\infty} e^{-\pi\tau^2m^2}e^{-\pi\tau^2x^2}e^{2\pi imx}.$$

By applying the shift theorem term by term, we obtain $F(s)$ in the form quoted above.

The function $F(s)$ is a row of Gaussian spikes of width τ with maxima lying on a broad Gaussian curve of width τ^{-1}. As before, it may be verified that as $\tau \to 0$ a suitable defining sequence for the *shah* symbol results.

Problems

1 Show that a periodic function $p(x)$ with unit period can always be expressed in the form $\mathrm{III}(x) * f(x)$ in infinitely many ways, and relate this to the fact that infinitely many different functions can share the same infinite set of equidistant samples.

2 Express the periodic pulse train

$$\sum_{n=-\infty}^{\infty} \Pi[10(x-n)]$$

in the form $\mathrm{III}(x) * f(x)$ in three distinct ways.

3 Determine the Fourier series coefficients for the functions of period equal to unity which, in the interval $-\frac{1}{2} < x < \frac{1}{2}$, are defined as follows:

a. $\cos \pi x$, $\Lambda(2x)$, $\Pi(2x) - \frac{1}{2}$,
b. $\Lambda(8x - 1)$, $(1 - 4x)H(x)$,
c. $e^{-\pi x^2}$, $1 - 4x^2$, $e^{-|x|}$, $e^{-x}H(x)$.

4 The following sample set defines a band-limited function:

$$\ldots\ 0,\ 10,\ 30,\ 50,\ 50,\ -40,\ -35,\ -10,\ -5,\ 0,\ \ldots$$

All samples omitted are zero. Establish the form of the function by numerical interpolation. What is the minimum value assumed by the function?

5 A certain function is approximately band-limited, that is, a small fraction μ of its power spectrum in fact lies beyond the nominal cutoff frequency. It is sampled at the nominal sampling interval and reconstituted by the usual rules. Use the inequality

$$f(x) \leqslant \int_{-\infty}^{\infty} |F(s)|\, ds$$

to examine how great the discrepancy between the original and reconstituted functions can become.

6 The approximately band-limited function mentioned above is subjected to ordinate-and-slope sampling. Use the inequality

$$f'(x) \leqslant 2\pi \int_{-\infty}^{\infty} |sF(s)|\, ds$$

to show that the discrepancy between the original and reconstituted functions can be serious, even when μ is small.

7 A little noise is added to a band-limited function. Before sampling, one subjects the noisy function to filtering that eliminates the noise components beyond the original cutoff. However, the function is still contaminated with a little band-limited noise. Examine the relative susceptibilities of ordinate-and-slope sampling and of ordinary ordinate sampling to the presence of the noise.

8 A finite sequence of equispaced impulses of finite strength

$$a_0\, \delta(x) + a_1\, \delta(x-1) + \cdots + a_n\, \delta(x-n)$$

is convolved with itself many times, and the result is another sequence of impulses

$$\Sigma \alpha_i\, \delta(x-i).$$

Combine the central-limit theorem with the sampling theorem to show that a graph of α_i against i will approach a normal distribution. State the simple condition that the coefficients $a_0, a_1, \ldots, a_n$ must satisfy.

9 In the previous problem, practice with simple cases that violate the condition, and develop the theory that suggests itself.

10 Show that the overshoot quoted as around 9 per cent in the discussion of the Gibbs phenomenon is given exactly by

$$-\int_1^{\infty} \operatorname{sinc} x\, dx.$$

11 A carrier telephony channel extends from a low-frequency limit $f_l = 95$ kilocycles per second to a high-frequency limit $f_h = 105$ kilocycles per second. The signal is sampled at a rate of 21 kilocycles per second, and a new waveform is generated consisting of very brief pulses at the sampling instants with strength proportional to the sample values. Make a graph showing the spectral bands occupied by the new waveform, and verify that the original waveform could be reconstructed. Show that in general the critical sampling rate is $2f_h$ divided by the largest integer not exceeding $f_h/(f_h - f_l)$.

12 In the previous problem, show that $2(f_h - f_l)$ is in general too slow a sampling rate to suffice to reconstitute a carrier signal. Show that by suitably interlacing two trains of equispaced samples the average sampling rate can, however, be brought down to twice the bandwidth.

13 A band-limited signal $X(t)$ passes through a filter whose impulse response is $I(t)$. The input and output signals $X(t)$ and $Y(t)$ can be represented by sample values X_i and Y_i. Show that

$$\{Y_i\} = \{I_i\} * \{X_i\}$$

and explain how to derive the sequence $\{I_i\}$.

14 In the previous problem the input signal samples are $\{X_i\} = \{1\ 2\ 3\ 4\ 5\ \ldots\}$, and the sequence $\{I_i\}$ describing the filter is $\{1\ 2\ 1\}$. Show that the output sequence $\{Y_i\}$ is $\{1\ 4\ 8\ 12\ \ldots\}$ and that it is possible to work back from the known output and determine the input by evaluating

$$\{1\ -2\ 3\ -4\ 5\ \ldots\} * \{1\ 4\ 8\ 12\ \ldots\}.$$

15 In the previous problem, verify numerically that a particular output signal sample can be expressed in terms of the history of the *output* plus a knowledge

of the most recent input signal sample, that is,

$$Y_i = \alpha_1 Y_{i-1} + \alpha_2 Y_{i-2} + \cdots + \beta X_i.$$

Show that the coefficients $\alpha_1, \alpha_2, \ldots$ and β are given in terms of the reciprocal sequence

$$\{K_0\ K_1\ K_2\ \ldots\} = \{I_i\}^{-1}$$

by

$$\{\alpha_1\ \alpha_2\ \alpha_3\ \ldots\} = -\frac{1}{K_0}\{K_1\ K_2\ K_3\ \ldots\}$$

and

$$\beta = \frac{1}{K_0}.$$

16 The input signal to a filter ceases, but the output continues, with each new sample deducible from the previous ones by the relation

$$Y_i = \alpha_1 Y_{i-1} + \alpha_2 Y_{i-2} + \cdots$$

Show that

$$Y(t) = [\alpha_1\, \delta(t-1) + \alpha_2\, \delta(t-2) + \cdots] * Y(t).$$

In a particular case, there are only two nonzero coefficients:

$$Y_i = 1.65Y_{i-1} - 0.9Y_{i-2}.$$

The first two output samples immediately after the input ceased were each 100; calculate and graph enough subsequent output values to determine the general character of the behavior. Show that a damped oscillation $Y(t) = e^{-\sigma t} \cos[\omega(t-a)]$ satisfies the convolution relation given above when

$$\alpha_1 = 2e^{-\sigma} \cos \omega$$

and

$$\alpha_2 = -e^{-2\sigma}$$

Is this band-limited behavior?

17 In the previous problem, show that the series for Y_i contains only a finite number of terms if the filter is constructed of a finite number of inductors, capacitors, and resistors. Hence show that the output due to any input signal is deducible, for a filter whose internal construction is unknown, after a certain number of consecutive sample measurements have been made at its terminals. How would you know that enough samples had been taken?

18 The input voltage $X(t)$, and output voltage $Y(t)$ of an electrical system are sampled simultaneously at regular intervals with the following results.

$X(t)$	15	10	6	2	1	0	0	0	0	0
$Y(t)$	15	15	7.5	−2.75	−2.5					

Calculate the missing values of $Y(t)$, and also calculate what the output would be if, after some time had elapsed, $X(t)$ began to rise linearly.

Chapter 11 The Laplace transform

Hitherto we have taken the variables x and s to be real variables. Now, however, let t be a real variable and p a complex one, and consider the integral

$$\int_{-\infty}^{\infty} f(t)e^{-pt}\,dt.$$

This is known as the (two-sided) Laplace transform of $f(t)$ and will be seen to differ from the Fourier transform merely in notation. When the real part of p is zero, the identity with the Fourier transform (as interpreted with noncomplex variables) is complete. In spite of this, however, there is a profound difference in application between the two transforms.

Alternative definitions of the Laplace transform include

$$\int_{0+}^{\infty} f(t)e^{-pt}\,dt,$$

which may be referred to as the one-sided Laplace transform, and

$$p\int f(t)e^{-pt}\,dt,$$

which may be referred to as the p-multiplied form.

The one-sided transform arises in the analysis of transients, where $f(t)$ comes into existence following the throwing of a switch at $t = 0$. However, if we deem that $f(t) = 0$ for $t < 0$, such cases are seen to be included in the two-sided definition. It is not always stressed that the lower limit of the integral defining the one-sided Laplace transform is $0+$; indeed in practice it is normally written as 0. One must remember that

$$\int_0^{\infty} f(t)e^{-pt}\,dt \quad \text{usually means} \quad \lim_{h\to 0}\int_{|h|}^{\infty} f(t)e^{-pt}\,dt.$$

Chapter 18 The discrete Fourier transform

If one wishes to obtain the Fourier transform of a given function, it may happen that the function is defined in terms of a continuous independent variable, as is most often the case in books, especially in lists of transform pairs. But it may also happen that function values are given only at discrete values of the independent variable, as with physical measurements made at regular time intervals. Regardless of the form of the given function, if the transform is evaluated by numerical computing, the values of the transform will be available only at discrete intervals. We often think of this as though an underlying function of a continuous variable really exists and we are approximating it. From an operational viewpoint, however, it is irrelevant to talk about the existence of values other than those given and those computed (the input and output). Therefore, it is desirable to have a mathematical theory of the actual quantities manipulated.

The discrete transform formula

Questions of discreteness also arise in connection with periodic functions. A periodic function is describable by a sequence of coefficients at discrete intervals (of frequency), but this situation *may* be viewed as a special case of continuous frequency. The transform is then regarded as a string of equally spaced delta functions (Fig. 18.1) of strengths given by the coefficients. What if this transform were itself periodic? Then the original function would also reduce to a string of delta functions. With both function and transform now being periodic, all the information about both would be limited to two finite sets of coefficients: the strengths of the delta functions.

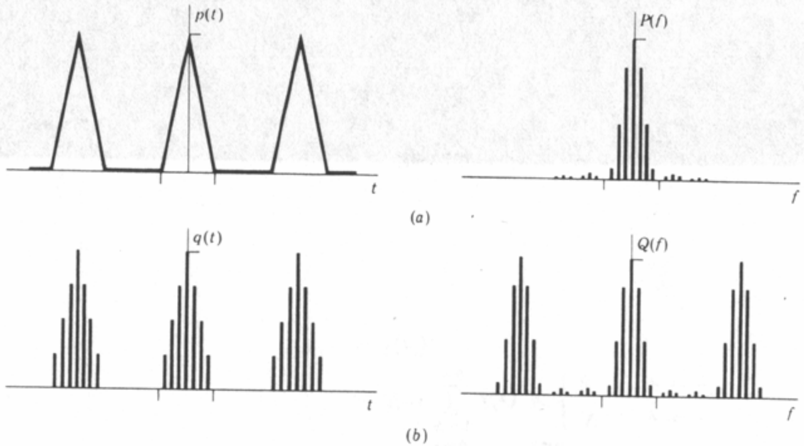


Fig. 18.1 (a) A periodic function $p(t)$ has a transform $P(f)$ which is a string of equispaced delta functions; (b) a periodic function $q(t)$, which is a string of equispaced delta functions, has a transform $Q(f)$ which is also a string of equispaced delta functions and the string is periodic.

Thus the practical situation where a finite set of values is given and a finite set is computed does actually lie within the continuous theory. However, it pays to start afresh rather than to force the theory of the discrete Fourier transform into the continuous framework. The reason is that the discrete notation is concise and reasonably standardized.

To retain some physical ties, let us think in terms of a signal that is a function of time; but to recognize the discreteness of the independent variable, let us use the symbol τ, which we agree can assume only a finite number N of consecutive integral values. Furthermore, we agree that τ cannot be negative. Thus, before entering into the realm of the discrete Fourier transform, we first make, if necessary, a change of scale and a change of origin. For example, suppose that a voltage waveform $v(t)$ is half a period of a cosinusoid of period 1 second (Fig. 18.2):

$$v(t) = \begin{cases} \cos 2\pi t & -0.25 < t < 0.25 \\ 0 & \text{otherwise} \end{cases}$$

and that samples are taken at intervals of 100 milliseconds. Table 18.1 shows signal values as a function of t but Table 18.2 shows how it is to be converted into a function $f(\tau)$ of discrete time τ before proceeding. This is illustrated in Fig. 18.2. In general, if the sampling interval is T and the first sample of interest occurs at $t = t_0$, then

$$f(\tau) = v(t_0 + \tau T), \qquad \tau = 0, 1, 2, \ldots, N - 1.$$

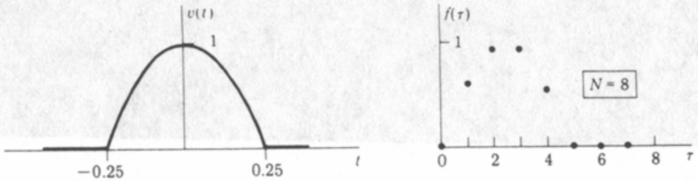


Fig. 18.2 A function of the continuous variable t and one way of representing it by eight sample values.

In what follows, $f(\tau)$ forms the point of departure. It will be noticed that no provision is made for cases where there is no starting point, as with a function such as $\exp(-t^2)$. This is in keeping with the practical character of the discrete transform, which does not contemplate data trains dating back to the indefinitely remote past. A second feature to note is that the finishing point must occur after a finite time. However, it need not come at $\tau = 5$ as in Table 18.2; one might choose to let τ run on to 15 and assign values of zero to the extra samples. This is a conscious choice that must always be made. It may be important; for example, Table 18.2 does not convey the information given in the equation preceding it—that following the half-period cosine, the voltage remains zero. The table remains silent on that point, and if it is important, the necessary number of zeros would need to be appended.

By *definition*, $f(\tau)$ possesses a discrete Fourier transform $F(\nu)$ given by

$$F(\nu) = N^{-1} \sum_{\tau=0}^{N-1} f(\tau) e^{-i2\pi(\nu/N)\tau}. \tag{1}$$

The quantity ν/N is analogous to frequency measured in cycles per sampling interval. The correspondence of symbols may be summarized as follows:

	Time	*Frequency*
Continuous case	t	f
Discrete case	τ	ν/N

The symbol ν has been chosen in the discrete case, instead of f, to emphasize that the frequency integer ν is *related* to frequency but is not the *same* as frequency f. For example, if the sampling interval is 1 second and there are eight samples ($N = 8$), then the component of frequency f will be found at $\nu = 8f$; conversely, the frequency represented by a frequency integer $\nu = 1$ will be $\frac{1}{8}$ hertz.

Table 18.1

t (milliseconds)	$v(t)$
−250	0
−150	0.588
−50	0.951
50	0.951
150	0.588
250	0

Table 18.2

τ	$f(\tau)$
0	0
1	0.588
2	0.951
3	0.951
4	0.588
5	0

Given the discrete transform $F(\nu)$, one may recover the time series $f(\tau)$ with the aid of the inverse relationship

$$f(\tau) = \sum_{\nu=0}^{N-1} F(\nu) e^{i2\pi(\nu/N)\tau}. \tag{2}$$

To see that this is so, we first verify the fact that

$$\sum_{\nu=0}^{N-1} e^{-i2\pi(\nu/N)(\tau-\tau')} = \begin{cases} N & \tau = \tau' \\ 0 & \text{otherwise.} \end{cases}$$

One way is to picture the summation as a closed polygon on the complex plane, except when $\tau = \tau'$, in which case the polygon becomes a straight line composed of N unit vectors end to end. (The variable τ' is taken to assume values $0, 1, \ldots, N-1$; if larger values were allowed, for example if $\tau - \tau'$ could become equal to $N, 2N, \ldots$, the summation would equal N for $\tau = \tau' \bmod N$.) Another way to think of this is by analogy with the orthogonality relation of sinusoids of different frequencies, a viewpoint that is aided by rewriting it

$$\sum_{\tau=0}^{N-1} e^{-i2\pi(\nu/N)\tau} e^{i2\pi(\nu'/N)\tau} = \begin{cases} N & \nu = \nu' \\ 0 & \text{otherwise.} \end{cases}$$

To establish the inverse discrete transform, introduce a dummy variable τ' for convenience, and substitute (1) into the right-hand side of (2):

$$\begin{aligned}
\sum_{\nu=0}^{N-1} F(\nu) e^{i2\pi(\nu/N)\tau'} &= \sum_{\nu=0}^{N-1} N^{-1} \sum_{\tau=0}^{N-1} f(\tau) e^{-i2\pi(\nu/N)\tau} e^{i2\pi(\nu/N)\tau'} \\
&= N^{-1} \sum_{\tau=0}^{N-1} f(\tau) \sum_{\nu=0}^{N-1} e^{-i2\pi(\nu/N)(\tau-\tau')} \\
&= N^{-1} \sum_{\tau=0}^{N-1} f(\tau) \times \begin{cases} N & \tau = \tau' \\ 0 & \tau \neq \tau' \end{cases} \\
&= f(\tau').
\end{aligned}$$

From the inverse transform we see that only N integral values of the frequency integer ν are needed and that they range from 0 to $N - 1$ just as with the discrete time τ. It certainly sounds reasonable that a function defined by N measurements should be representable after transformation by just N parameters. Even when the values of f are real, the values of F are in general complex and, as will be seen below, one must be careful how to count complex numbers. The fact that ν ranges over integral values, whereas the frequency ν/N is fractional, is the reason for introducing the integer ν; mathematical convenience takes priority over physical significance.

In order to regain our physical feeling for numerical orders of magnitude, let us consider a record consisting of 1,024 samples separated by 1-second intervals. We expect this to be representable by a Fourier series consisting of a constant term and multiples of a certain fundamental frequency. The fundamental period should be 1,024 seconds, corresponding to a fundamental frequency 1/1,024 hertz. The highest frequency needed will be 0.5 hertz, which has two samples per period. This will be the 512th harmonic. The reason that ν assumes 1,024 values, whereas the number of frequencies is only 512, is as follows. If the values of $f(\tau)$ are real, as is usual in records of physical quantities, there are 1,024 real data values. Now the transform $F(\nu)$ has 1,024 complex values which would require 2,048 real numbers to specify except that $F(0)$ and $F(N/2)$ have no imaginary part [see (Eq. 1)] and half the remaining values of $F(\nu)$ are complex conjugates of the other half. This is because $f(\tau)$ is real (Chapter 2). If $f(\tau)$ were complex, there would be 2,048 real data values and $F(\nu)$ would require 2,048 real numbers for its specification.

Since the highest frequency reached is 0.5 hertz, it is apparent that ν/N, which reaches a maximum value of 1,023/1,024, is not a strict analogue of frequency. For instance, where $\nu = 724$, $\nu/N = 0.707$ hertz and, as we have seen, frequencies above 0.5 hertz are neither required nor can be represented by samples at 1-second intervals. Rather, $\nu = 724$ corresponds to a negative frequency $-300/1{,}024$ hertz.

This anomaly is a distinct impediment to the visualization of the connection between the Fourier transform and the discrete Fourier transform. One way to bring the two into harmony would be to redefine the discrete transform in terms of summation over negative and positive values, although it might be objected that negative subscripts are permitted only in some computer languages (for example, ALGOL). On the other hand, some widely used languages do not even allow indexing to start at *zero*, and so an index shift is required anyhow. Another objection asserts the desirability of making N an even number. Figure 18.3b

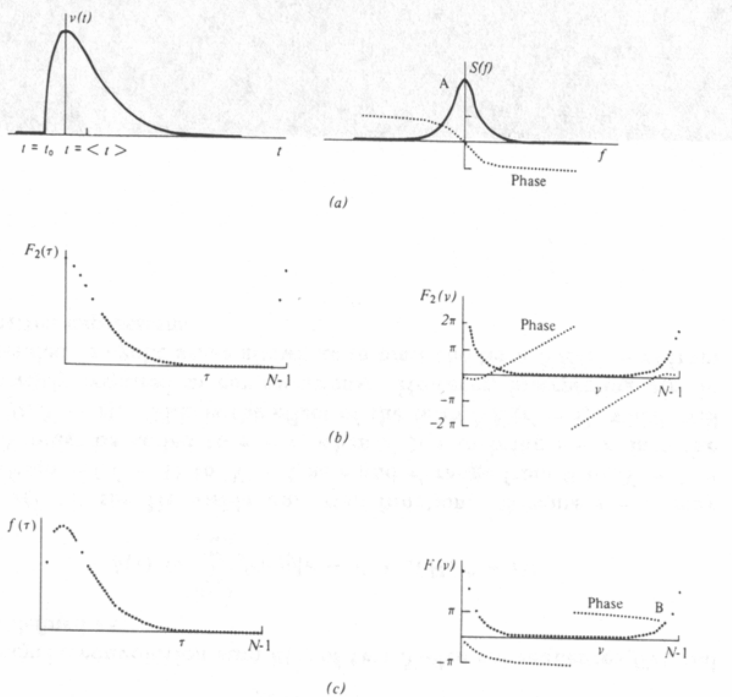


Fig. 18.3 (a) A function and its Fourier transform; (b) and (c) two way of representing the function by N samples and the corresponding discrete Fourier transforms, shown by modulus (dots) and phase (small dots).

shows how a signal $v(t)$ and its transform $S(f)$ might be made to correspond with $f(\tau)$ and its discrete transform $F(\nu)$. Alternatively, as in Fig. 18.3c, it could be arranged that $\tau = 0$ corresponds to $t = t_0$; this would not prevent the small negative values of frequency labeled A from corresponding to large positive values of ν labeled B.

Finally, another way of looking at this, which is customary, is to lift the restriction of τ and ν to values from 0 to $N - 1$ and to allow all integral values, including negative values. The function values assigned are on the basis that $f(\tau)$ and $F(\nu)$, in their extended sense, are periodic, with period N. Thus

$$f(\tau) = f(\tau \pm N) = f(\tau \pm 2N) = \ldots$$
$$F(\nu) = F(\nu \pm N) = F(\nu \pm 2N) = \ldots .$$

This plan will be adopted here. Under this plan, the basic transform and its inverse may be written

$$F(\nu) = N^{-1} \sum_{\tau=-(1/2)N}^{(1/2)N-1} f(\tau)e^{-i2\pi(\nu/N)\tau} \tag{3}$$

$$f(\tau) = \sum_{\nu=-(1/2)N}^{(1/2)N-1} F(\nu)e^{i2\pi(\nu/N)\tau} \tag{4}$$

and ν/N may be identified with frequency measured in cycles per sampling interval over the range $-\frac{1}{2}N \le \nu \le \frac{1}{2}N$. If the sampling interval is T, the frequency measured in hertz is ν/NT.

In Fig. 18.4 we see a way of visualizing $f(\tau)$ and $F(\nu)$ as having cyclic dependence on τ and ν. The upper type of diagram is helpful in making a connection with our previous experience [e.g., to take the autocorrelation of a sequence, we imagine $f(\tau)$ to be rotated, in the top left-hand diagram, relative to itself]. Multiplication of corresponding values followed by summation gives a value of the autocorrelation and we can see how the result ties in with our previous knowledge that the autocorrelation of a rectangle function is a triangle function. This illuminates the lower diagrams where the original form of indexing, starting from zero, obscures the simple shapes.

Cyclic convolution

If we convolve two sequences, one having m elements and the other n, then the convolution sum, or serial product, will have $m + n - 1$ elements (p. 32). In particular, if we deal with sequences having N elements, then the convolution of two such sequences will not itself be containable in an N-element sequence. Figure 18.4 (top center) coped with the desire to exhibit the triangular autocorrelation of two rectangles by packing the sequence expressing the rectangle with extra zeros. Three-fourths of the elements shown at the top left are zeros, and at the top center there are still plenty of zeros left to witness to the isolated nature of the triangular island.

Suppose now that N is kept constant while the number of nonzero elements at the top left is increased. When the rectangle extends more than halfway around the circle, the outer ends of the triangle overlap. The result will be a flat, nonzero segment in the region of overlap. Clearly, the result is wrong if we are looking for the triangular autocorrelation. However, the operation exists in its own right, and may be defined as cyclic convolution.

The cyclic convolution integral $h(\theta)$ of two periodic functions $f(\cdot)$ and

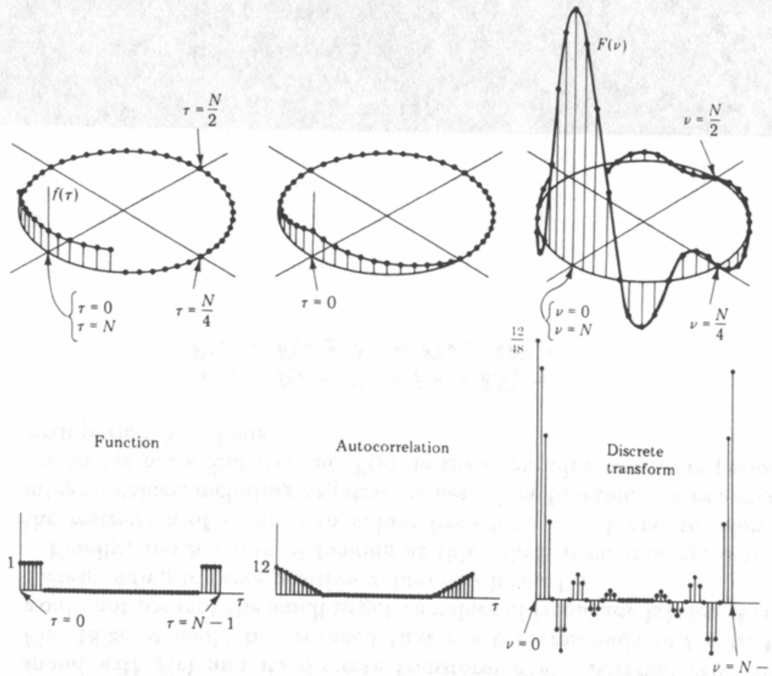


Fig. 18.4 Cyclic (above) and standard (below) representations of a sampled rectangle function, its triangular autocorrelation, and its transform.

$g(\cdot)$ with period 2π is defined as

$$h(\theta) = \int_0^{2\pi} f(\theta')g(\theta - \theta')\,d\theta'.$$

The cyclic convolution sum $h(\tau)$ of two N-element sequences $f(\tau)$ and $g(\tau)$ is defined as

$$h(\tau) = \sum_{\tau'=0}^{N-1} f(\tau')g[\tau - \tau' + NH(\tau' - \tau)],$$

where $H(\cdot)$ is the Heaviside unit step function. Because $\tau - \tau'$ may range from $-(N-1)$ to $N-1$ as τ and τ' range from 0 to $N-1$, a term N must be added to $\tau - \tau'$ when $\tau' > \tau$ to bring $\tau - \tau'$ into the range $[0, N-1]$. This is the effect of the term $NH(\tau' - \tau)$, which will be explicitly required in computations. However, interpreting $g(\tau)$ in its extended or cyclic sense allows us to omit the term $NH(\tau' - \tau)$ from the written expressions.

Examples of discrete Fourier transforms[1]

Certain short, discrete transform pairs are often needed, and a small stock of transform pairs can be helpful for checking. Here is a reference list.

$N = 2$:

$$\{1\ 0\} \supset \tfrac{1}{2}\{1\ 1\}$$
$$\{1\ 1\} \supset \tfrac{1}{2}\{2\ 0\}$$
$$\{0\ 1\} \supset \tfrac{1}{2}\{1\ \ -1\}$$
$$\{1\ \ -1\} \supset \tfrac{1}{2}\{0\ 2\}.$$

$N = 4$:

$$\{1\ 0\ 0\ 0\} \supset \tfrac{1}{4}\{1\ 1\ 1\ 1\}$$
$$\{0\ 1\ 0\ 0\} \supset \tfrac{1}{4}\{1\ \ -i\ \ -1\ \ i\}$$
$$\{0\ 0\ 1\ 0\} \supset \tfrac{1}{4}\{1\ \ -1\ \ 1\ \ -1\}$$
$$\{0\ 0\ 0\ 1\} \supset \tfrac{1}{4}\{1\ \ i\ \ -1\ \ -i\}$$
$$\{1\ 1\ 0\ 0\} \supset \tfrac{1}{4}\{2\ \ 1-i\ \ 0\ \ 1+i\}$$
$$\{0\ 0\ 1\ 1\} \supset \tfrac{1}{4}\{2\ \ -1+i\ \ 0\ \ -1-i\}$$
$$\{1\ 1\ 1\ 1\} \supset \tfrac{1}{4}\{4\ 0\ 0\ 0\}$$
$$\{1\ 1\ 0\ \ -1\} \supset \tfrac{1}{4}\{1\ \ 1-2i\ \ 1\ \ 1+2i\}$$
$$\{1\ 0\ 0\ 1\} \supset \tfrac{1}{4}\{2\ \ 1+i\ \ 0\ \ 1-i\}.$$

$N = 8$:

$$\{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\} \supset \tfrac{1}{8}\{8\ 0\ 0\ 0\ 0\ 0\ 0\ 0\}$$
$$\{1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\} \supset \tfrac{1}{8}\{1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\}$$
$$\{0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\} \supset \tfrac{1}{8}\{1\ \ e^{-i2\pi/8}\ \ e^{-i2\pi(2/8)}\ \ e^{-i2\pi(3/8)}\ \ -1\ \ e^{-i2\pi(5/8)}\ \ e^{-i2\pi(6/8)}\ \ e^{-i2\pi(7/8)}\}$$
$$\{0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\} \supset \tfrac{1}{8}\{1\ \ -i\ \ -1\ \ i\ \ 1\ \ -i\ \ -1\ \ i\}.$$

Reciprocal property

Just as the Fourier transformation applied twice in succession, with alternating sign of i, gives back the original function, so there is a corresponding property for the discrete Fourier transform. But owing to the use of a scaling factor N to make the index ν integral, the discrete transform is not strictly reciprocal, even allowing for the sign of i. Thus

$$N^{-1}\left[\sum_{\nu=0}^{N-1} N^{-1} \sum_{\tau'=0}^{N-1} f(\tau')e^{-i2\pi(\nu/N)\tau'}\right] e^{i2\pi(\nu/N)\tau} = N^{-1}f(\tau).$$

If the DFT is applied twice in succession without changing the sign

[1] As there is little risk of confusion, we may use the sign $\supset$ to stand for "has DFT." Thus Eq. (1) could be written $f(\tau) \supset F(\nu)$ and Eq. (2) could be written $F(\nu) \subset f(\tau)$, where $\subset$ means "has inverse DFT."

of i, we will get $N^{-1}f(-\tau)$ on the right-hand side. Expressing this differently, if

$$f(\tau) \supset F(\nu),$$

then

$$F(\nu) \supset N^{-1}f(-\tau).$$

This property and others about to be mentioned and several theorems will not be derived. Instead, emphasis will be placed on interpreting and illustrating them and presenting them in a form suitable for reference.

Oddness and evenness

By definition, $f(\tau)$ is even if

$$f(-\tau) = f(\tau)$$

and odd if

$$f(-\tau) = -f(\tau).$$

Figure 18.5 illustrates the following odd and even sequences of 16 elements, where τ runs from 0 to 15.

Even. $\{5\ 4\ 3\ 2\quad 1\ 0\ 0\ 0\quad 0\ 0\ 0\ 0\quad 1\ 2\ 3\ 4\}$.
Odd: $\{0\ 8\ 7\ 6\quad 5\ 4\ 3\ 2\quad 0\ \ -2\ \ -3\ \ -4\quad -5\ \ -6\ \ -7\ \ -8\}$.

In these examples, the elements have been grouped in fours to help display the nature of the symmetry. Even and odd sequences of four elements are as follows:

Even: $\{a\ b\ c\ b\}$.
Odd: $\{0\ b\ 0\ \ -b\}$.

Clearly the rules for extending the range of τ to negative integers are required here, in particular the special case

$$f(-\tau) = f(N - \tau).$$

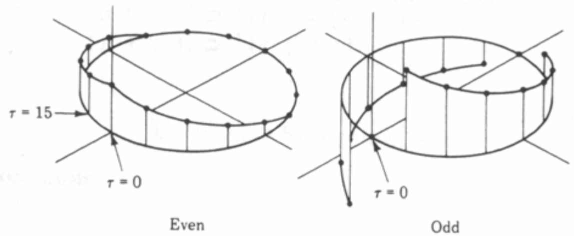


Fig. 18.5 Even and odd sequences shown cyclically.

A similar relation holds for $F(\nu)$,

$$F(-\nu) = F(N - \nu).$$

Examples with special symmetry

Symmetry rules with respect to oddness and evenness assume the same form for the discrete transform as given in Chapter 2 for the continuous transform.

$$\begin{aligned}
\text{real} &\supset \text{hermitian}\\
\text{imaginary} &\supset \text{antihermitian}\\
\text{real and even} &\supset \text{real and even}\\
\text{real and odd} &\supset \text{imaginary and odd}\\
\text{imaginary and even} &\supset \text{imaginary and even}\\
\text{imaginary and odd} &\supset \text{real and odd}\\
\text{even} &\supset \text{even}\\
\text{odd} &\supset \text{odd.}
\end{aligned}$$

The following examples illustrate these rules, which are of major importance in practical computing

$$\begin{aligned}
\{1\ 2\ 3\ 4\} &\supset \tfrac{1}{4}\{10\ \ -2+2i\ \ -2\ \ -2-2i\}\\
i\{1\ 2\ 3\ 4\} &\supset \tfrac{1}{4}\{10i\ \ -2-2i\ \ -2\ \ 2-2i\}\\
\{1\ 0\ 0\ 0\} &\supset \tfrac{1}{4}\{1\ 1\ 1\ 1\}\\
\{0\ 1\ 0\ -1\} &\supset \tfrac{1}{4}\{0\ -2i\ 0\ 2i\}\\
i\{4\ 2\ 1\ 2\} &\supset \tfrac{1}{4}i\{9\ 3\ 1\ 3\}\\
i\{0\ 1\ 0\ -1\} &\supset \tfrac{1}{4}\{0\ 2\ 0\ -2\}\\
\{1+4i\ \ 2i\ \ i\ \ 2i\} &\supset \tfrac{1}{4}\{1+9i\ \ 1+3i\ \ 1+i\ \ 1+3i\}\\
\{0\ \ 1+i\ \ 0\ \ -1-i\} &\supset \tfrac{1}{4}\{0\ \ 2-2i\ \ 0\ \ -2-2i\}.
\end{aligned}$$

Complex conjugates

The discrete transform of the conjugate is the conjugate of the transform, reversed:

$$f^*(\tau) \supset F^*(-\nu).$$

By "reversal," we mean changing the sign of the independent variable.

Reversal property

If the sign of τ is changed, that is, if $f(\tau)$ is reflected in the line $\tau = 0$, the sign of ν is changed:

$$f(-\tau) \supset F(-\nu).$$

It is worth noting that this operation on $f(\tau)$ is also produced by reflection in the line $\tau = \frac{1}{2}N$.

Addition theorem

$$f_1(\tau) + f_2(\tau) \supset F_1(\nu) + F_2(\nu).$$

Example:

If $\{2\ 0\ 0\ 0\} \supset \frac{1}{4}\{2\ 2\ 2\ 2\}$
and $\{0\ 1\ 0\ 0\} \supset \frac{1}{4}\{1\ \ -i\ \ -1\ \ i\}$,
then $\{2\ 1\ 0\ 0\} \supset \frac{1}{4}\{3\ \ 2-i\ \ 1\ \ 2+i\}$.

Shift theorem

$$f(\tau - T) \supset e^{-i2\pi T(\nu/N)}F(\nu).$$

Example:

$$\begin{aligned}
\{1\ 0\ 0\ 0\} &\supset \tfrac{1}{4}\{1\ 1\ 1\ 1\}\\
\{0\ 1\ 0\ 0\} &\supset \tfrac{1}{4}\{1\ \ -i\ \ (-i)^2\ \ (-i)^3\}\\
\{0\ 0\ 1\ 0\} &\supset \tfrac{1}{4}\{1\ \ (-i)^2\ \ (-i)^4\ \ (-i)^6\}\\
\{0\ 0\ 0\ 1\} &\supset \tfrac{1}{4}\{(1\ \ (-i)^3\ \ (-i)^6\ \ (-i)^9\}.
\end{aligned}$$

Sometimes a shift in the frequency domain is required; one version of what could be called the inverse shift theorem is

$$e^{i2\pi(\nu_0/N)\tau}f(\tau) \supset F(\nu - \nu_0).$$

Convolution theorem

The cyclic convolution of two sequences $\{f_1(\tau)\}$ and $\{f_2(\tau)\}$ was defined by

$$f_1(\tau) * f_2(\tau) = \sum_{\tau'=0}^{N-1} f_1(\tau')f_2(\tau - \tau').$$

Remember that $f_2(\cdot)$ has to be understood in its extended cyclic sense. To emphasize the distinction between this discrete operation and the convolution *integral*, we may use the term convolution *sum*, but where there is no risk of confusion, it may be called simply the convolution of the sequences $\{f_1\}$ and $\{f_2\}$. The theorem is

$$f_1(\tau) * f_2(\tau) \supset NF_1(\nu)F_2(\nu).$$

Example:

Let $\{1\ 1\ 0\ 0\} \supset \frac{1}{4}\{2\ \ 1-i\ \ 0\ \ 1+i\}$.

Then $\{1\ 1\ 0\ 0\} * \{1\ 1\ 0\ 0\} = \{1\ 2\ 1\ 0\} \supset \frac{1}{4}\{4\ \ -2i\ \ 0\ \ 2i\}$.

Also, $\{1\ 1\ 0\ 0\} * \{0\ 0\ 2\ 2\} = \{2\ 0\ 2\ 4\} \supset \frac{1}{4}\{8\ \ 4i\ \ 0\ \ 4i\}$.

Product theorem

The inverse of the convolution theorem, which applies to products in the τ domain, or convolution in the ν domain, is

$$f_1(\tau)f_2(\tau) \supset \sum_{\nu'=0}^{N-1} F_1(\nu')F_2(\nu-\nu').$$

Example:

$$f_1 = f_2 = f_1f_2 = \{1\ 1\ 0\ 0\} \supset \tfrac{1}{4}\{2\ \ 1-i\ \ 0\ \ 1+i\}.$$

Cross-correlation

$$\sum_{\tau'=0}^{N-1} f_1(\tau')f_2(\tau'+\tau) \supset NF_1(\nu)F_2(-\nu).$$

Autocorrelation

$$\sum_{\tau'=0}^{N-1} f_1(\tau')f_1(\tau'+\tau) \supset N|F_1(\nu)|^2.$$

Example:

$$\{1\ 1\ 0\ 0\} \star \{1\ 1\ 0\ 0\} = \{2\ 1\ 0\ 1\} \supset \tfrac{1}{4}\{4\ \ 2\ \ 0\ \ 2\}.$$

Sum of sequence

$$\sum_{\tau=0}^{N-1} f(\tau) = NF(0).$$

Example:

$$\{1\ 0\ 5\ 0\} \supset \{1.5\ \ -1\ \ 1.5\ \ -1\}.$$

We see that $\Sigma f = 6$, $F(0) = 1.5$, $N = 4$, and $NF(0) = 6$. If we follow the practice of writing N^{-1} as a first factor of $F(\nu)$, the theorem means that the sum of the sequence is equal to the first term after the opening brace on the right-hand side.

First value

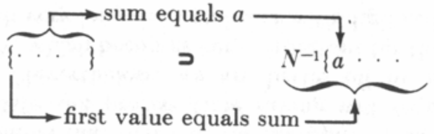
This is the inverse of the preceding:

$$f(0) = \sum_{\nu=0}^{N-1} F(\nu).$$

Example:

$$f(0) = 1 \quad\text{and}\quad \Sigma F = 1.5 - 1 + 1.5 - 1 = 1.$$

We see that the sum of a sequence equals N times the first value of its DFT, but conversely the sum of the DFT exactly equals the first value of the sequence.



Generalized Parseval-Rayleigh theorem

$$\sum_{\tau=0}^{N-1} |f(\tau)|^2 = N\sum_{\nu=0}^{N-1} |F(\nu)|^2.$$

Example:

$$\{1\ 1\ 0\ 0\} \supset \tfrac{1}{4}\{2\ \ 1-i\ \ 0\ \ 1+i\}.$$

We see that $\Sigma f^2 = 2$ and that $N\Sigma F^2 = 4 \times 0.5 = 2$.

Packing theorem

The packing operator Pack_K packs a given N-member sequence $f(\tau)$ with trailing zeros so as to increase the number of elements to KN.

$$\mathrm{Pack}_K\ \{f(\tau)\} = \{g(\tau)\},$$

where $$g(\tau) = \begin{cases} f(\tau) & 0 \le \tau \le N-1 \\ 0 & N \le \tau \le KN-1. \end{cases}$$

Thus $\mathrm{Pack}_2\ \{1\ 2\ 3\ 4\} = \{1\ 2\ 3\ 4\ 0\ 0\ 0\ 0\}$.

The theorem is

$$\mathrm{Pack}_K\ \{f(\tau)\} \supset G(\nu),$$

where $$G(\nu) = \frac{1}{K}F\left(\frac{\nu}{K}\right), \qquad \nu = 0, K, 2K, \ldots, KN-K.$$

The intermediate values of $G(\nu)$, not given by this relation, can be determined by sinc-function interpolation between the known values [e.g., by midpoint interpolation (p. 195) when $K = 2$], but for a better method, see Problem 18.8.

Similarity theorem

To have an analogy with expansion or contraction of the scale of continuous time, we must supply sufficient zero elements, either at the end, as with packing, so that the sequence may expand, or between elements, so that there is room for contraction. The operation of inserting zeros between elements so as to increase the total number of elements by a factor K will be denoted by the stretch operator Stretch_K.

$$\text{Stretch}_K \{f(\tau)\} = \{g(\tau)\},$$

where
$$g(\tau) = \begin{cases} f(\tau/K) & \tau = 0, K, 2K, \ldots, (N-1)K \\ 0 & \text{otherwise.} \end{cases}$$

Example:

$$\text{Stretch}_2 \{1\ 2\ 3\ 4\} = \{1\ 0\ 2\ 0\ 3\ 0\ 4\ 0\}.$$

The theorem is, if $\{g\} \supset \{G\}$,

$$G(\nu) = \begin{cases} \dfrac{1}{K} F(\nu) & \nu = 0, \ldots, N-1 \\ \dfrac{1}{K} F(\nu - N) & \nu = N, \ldots, 2N-1 \\ \cdots\cdots & \cdots\cdots \\ \dfrac{1}{K} F(\nu - \overline{K-1}N) & \nu = (K-1)N, \ldots, KN-1. \end{cases}$$

Thus, stretching by a factor K in the τ domain results in K-fold repetition of $F(\nu)$ in the ν domain; the frequency scale is not compressed by a factor K.

The fast Fourier transform

In 1965 a method of computing discrete Fourier transforms suddenly became widely known (J. W. Cooley and J. W. Tukey, *Math. Comput.*, vol. 19, April 1965, pp. 297–301), which revolutionized many fields where onerous computing was an impediment to progress. A good source of detailed information is a set of papers published in the *IEEE Transactions on Audio and Electroacoustics*, vol. AU-2, June 1967. Another source is G. D. Bergland, *Spectrum*, vol. 6, July 1969, pp. 41-52.

There are various ways of understanding this fast Fourier transform (FFT). One way, which will appeal to certain people, is in terms of factorization of the transform matrix. From the definition, we can write the DFT relation (for $N = 8$) in the form of a matrix product,

$$\begin{bmatrix} F(0)\\F(1)\\F(2)\\F(3)\\F(4)\\F(5)\\F(6)\\F(7) \end{bmatrix} = \begin{bmatrix} 1&1&1&1&1&1&1&1\\ 1&W&W^2&W^3&W^4&W^5&W^6&W^7\\ 1&W^2&W^4&W^6&W^8&W^{10}&W^{12}&W^{14}\\ 1&W^3&W^6&W^9&W^{12}&W^{15}&W^{18}&W^{21}\\ 1&W^4&W^8&W^{12}&W^{16}&W^{20}&W^{24}&W^{28}\\ 1&W^5&W^{10}&W^{15}&W^{20}&W^{25}&W^{30}&W^{35}\\ 1&W^6&W^{12}&W^{18}&W^{24}&W^{30}&W^{36}&W^{42}\\ 1&W^7&W^{14}&W^{21}&W^{28}&W^{35}&W^{42}&W^{49} \end{bmatrix} \times \begin{bmatrix} f(0)\\f(1)\\f(2)\\f(3)\\f(4)\\f(5)\\f(6)\\f(7) \end{bmatrix}, \qquad (1)$$

where $W = \exp(-i2\pi/N)$. The quantity W is an Nth root of unity, since $W^N = \exp(-i2\pi) = 1$. It may be thought of as a complex number whose modulus is unity and whose phase is $-(1/N)$ turns.

$$\begin{bmatrix} F(0)\\F(1)\\F(2)\\F(3)\\F(4)\\F(5)\\F(6)\\F(7) \end{bmatrix} = \begin{bmatrix} 1&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&1&0\\ 0&1&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&1 \end{bmatrix} \times \begin{bmatrix} 1&1&0&0&0&0&0&0\\ 1&W^4&0&0&0&0&0&0\\ 0&0&1&1&0&0&0&0\\ 0&0&1&W^4&0&0&0&0\\ 0&0&0&0&1&1&0&0\\ 0&0&0&0&1&W^4&0&0\\ 0&0&0&0&0&0&1&1\\ 0&0&0&0&0&0&1&W^4 \end{bmatrix}$$

$$\times \begin{bmatrix} 1&0&1&0&0&0&0&0\\ 0&1&0&W^2&0&0&0&0\\ 1&0&W^4&0&0&0&0&0\\ 0&1&0&W^6&0&0&0&0\\ 0&0&0&0&1&0&1&0\\ 0&0&0&0&0&1&0&W^2\\ 0&0&0&0&1&0&W^4&0\\ 0&0&0&0&0&1&0&W^6 \end{bmatrix} \times \begin{bmatrix} 1&0&0&0&1&0&0&0\\ 0&1&0&0&0&W&0&0\\ 0&0&1&0&0&0&W^2&0\\ 0&0&0&1&0&0&0&W^3\\ 1&0&0&0&W^4&0&0&0\\ 0&1&0&0&0&W^5&0&0\\ 0&0&1&0&0&0&W^6&0\\ 0&0&0&1&0&0&0&W^7 \end{bmatrix} \times \begin{bmatrix} f(0)\\f(1)\\f(2)\\f(3)\\f(4)\\f(5)\\f(6)\\f(7) \end{bmatrix}. \qquad (2)$$

This factorization leaves only two nonzero elements in each row. In (1) there are N^2 multiplications but there are only $2N$ multiplications per factor if we use (2), and the number of factors M is given by $2^M = N$ if we do not count the first factor, which merely represents a rearrangement. Thus the multiplications total $2N \log_2 N$. Examination of the factors shows that many of the multiplications are trivial, and therefore to calculate the precise time saving will require careful attention to details. Nevertheless, we are better off by a factor of the order of $N/\log_2 N$, which becomes very important for the large values of N which arise with very long data trains or with digitized two-dimensional images such as photographs, for example.

Here is another method of understanding the fast Fourier transform. A sequence of N elements may be divided into two shorter sequences of $N/2$ elements each by placing the even-numbered elements into the first sequence and the odd-numbered ones into the second. For example, $\{8\ 7\ 6\ 5\ 4\ 3\ 2\ 1\}$ can be split into $\{8\ 6\ 4\ 2\}$ and $\{7\ 5\ 3\ 1\}$. Each of these possesses a DFT. From these two DFT's how could one obtain the DFT of the longer sequence? The answer is obtained by writing

$$\{8\ 7\ 6\ 5\ 4\ 3\ 2\ 1\} = \{8\ 0\ 6\ 0\ 4\ 0\ 2\ 0\} + \{0\ 7\ 0\ 5\ 0\ 3\ 0\ 1\}.$$

We see that the desired DFT can be obtained by using the stretching and shift theorems. From the stretching theorem we know that if

$$\{8\ 6\ 4\ 2\} \supset \{A\ B\ C\ D\},$$

then

$$\{8\ 0\ 6\ 0\ 4\ 0\ 2\ 0\} \supset \{A\ B\ C\ D\ A\ B\ C\ D\}, \tag{3}$$

a phenomenon that may be familiar from Fourier series coefficients for periodic functions.

Likewise, if

$$\{7\ 5\ 3\ 1\} \supset \{P\ Q\ R\ S\},$$

then

$$\{7\ 0\ 5\ 0\ 3\ 0\ 1\ 0\} \supset \{P\ Q\ R\ S\ P\ Q\ R\ S\}.$$

Now we apply the shift theorem to find that

$$\{0\ 7\ 0\ 5\ 0\ 3\ 0\ 1\} \supset \{P\ WQ\ W^2R\ W^3S\ W^4P\ W^5Q\ W^6R\ W^7S\}. \tag{4}$$

Multiplication by W means rotation by one Nth of a revolution in the complex plane, so the effect of the shift is to apply a phase delay that increases progressively along the sequence of elements $\{P\ Q\ R\ S\ P\ Q\ R\ S\}$. Adding (3) and (4) gives the DFT of the long sequence. Thus the transformation with $N = 8$ has been broken down into two

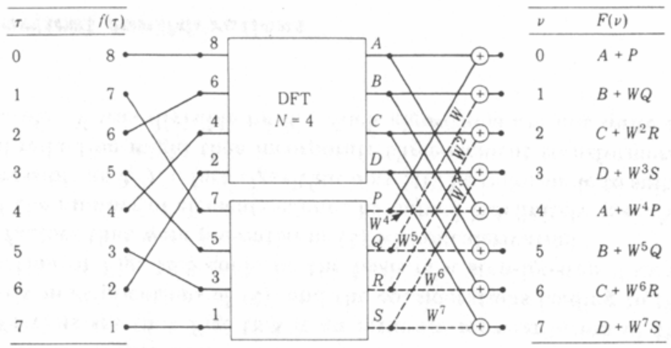


Fig. 18.6 Reduction of eight-element DFT to two four-element DFTs.

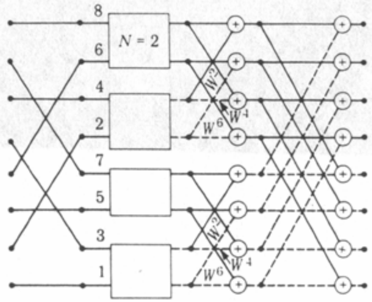


Fig. 18.7 Reduction to four two-element DFTs.

transformations with $N = 4$, which potentially represents a 50 per cent time saving, since the number of multiplications in a DFT performed according to (1) goes as N^2. To see how this breaking down can be taken even further, we refer to Fig. 18.6. Starting with the given sequence on the left, we rearrange it into the two short sequences $\{8\ 6\ 4\ 2\}$ and $\{7\ 5\ 3\ 1\}$ that form the inputs to two transformers with $N = 4$ whose outputs are $\{A\ B\ C\ D\}$ and $\{P\ Q\ R\ S\}$, respectively. The unbroken flow lines show that A, B, C, and D are transferred to the output nodes to deliver $\{A\ B\ C\ D\ A\ B\ C\ D\}$. The broken flow lines are tagged with factors that cause the delivery of P, WQ, W^2R, etc., as in (4) to the same output nodes, where addition takes place. Figure 18.7 now shows a further reduction of each four-element transformer to two two-element transformers, and Fig. 18.8 shows the full reduction to multiplications and additions.

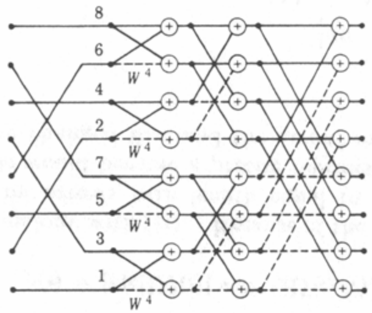


Fig. 18.8 Reduction of eight-element DFT to 3×16 multiplications and 3×8 additions. In the foregoing three figures, multiplication by unity is shown by a full line and the broken lines are associated with the factors shown. [Adapted from W. T. Cochran et al., IEEE Trans., AU-15 (1967), p. 45.]

Finally, the steps may be summarized as follows. First, we rearrange the given sequence into {8 4 6 2 7 3 5 1}, an operation corresponding exactly to multiplication by the first square matrix of (2) and sometimes loosely referred to as bit reversal. Then eight new numbers are calculated as linear combinations of various pairs of the rearranged data, exactly as indicated by the second square matrix of (2). These numbers are the outputs from the left-hand column of adders in Fig. 18.8. There are two more similar stages, making a total of three such operations in all (or M, in general, where $2^M = N$). Of course, not all the 48 multiplications are significant. There are 32 multiplications by unity and 7 multiplications by W^4, which is simply a sign reversal. In addition, W^2 and W^6 are rather simple to handle.

We thus see that Fig. 18.8 is an intimate representation of the four matrix multiplications of (2), and the considerations leading to the construction of Fig. 18.8 could be the basis of a step-by-step discovery of the factors that were presented in (2) without derivation.

If the number of elements cannot be halved indefinitely (i.e., N is not expressible as 2^M), a fast algorithm may still be tailormade to suit. The final reduction might then incorporate three-element transformers if, for example, N was divisible by 3. Such algorithms are not quite as fast.

Practical considerations

Many practical considerations have been incorporated into the programs available for performing the fast Fourier transform. For some applications, speed is an overriding consideration; for others, convenience. If N is not a power of 2, convenience says pack the data with zeros; speed says choose a modified program taking advantage of such factors as N possesses. Some users never require complex output; others do, but not in the form of real and imaginary parts. Some require two and three dimensions. Some normally have to segment their data because N exceeds the capacity of their computer. Questions of this kind, though important, can be studied through the literature using the sources given above or by examination of existing programs. Examination of documentation can be particularly important, because some software packages do not implement the DFT at all but rather some modification possessing more or less convenience.

Here we shall look into the very common situation where a data string of, say, 60 elements is to be transformed by a general-purpose program which allows a choice of 64, 128, or other power of 2 for N. Adding four zeros will fit the data to the program. Will it make a difference whether the zeros trail, precede, or are placed in twos at start and finish? From

Fig. 18.4 and remembering the shift theorem, we see that $|F(\nu)|$ will be unaffected but that the effective shifts of origin will introduce phase differences. If phase is important, as it might be if the data string has a natural origin, the shift theorem will supply the appropriate phase correction factor.

Will it make a difference if 68 trailing zeros are added and the $N = 128$ program is used? Surprisingly to some, the answer can be yes. To understand this, consider the extended sense in which $f(\tau)$ and $F(\nu)$ are regarded as periodic with period N.

Let $v(t)$ be a function of the continuous variable t, which at integral values of t in the range just embracing 0 to $N - 1$ agrees with $f(\tau)$ and is zero outside, as in Fig. 18.9a. Then, $v(t)\mathrm{III}(t)$ is a string of impulses (Fig. 18.9b) that contains precisely the same information as $f(\tau)$ but does not have the property of repeating with period N. Periodic character is, however, exhibited by the expression

$$p(t) = [v(t)\mathrm{III}(t)] * N^{-1}\mathrm{III}(t/N),$$

which is in strict analogy with $f(\tau)$. Because of the convention of representing impulses by arrows with length equal to the strength of the impulse, Fig. 18.9c would become a precise representation of $f(\tau)$ if the abscissa label were changed to τ and the arrowheads were changed to blobs.

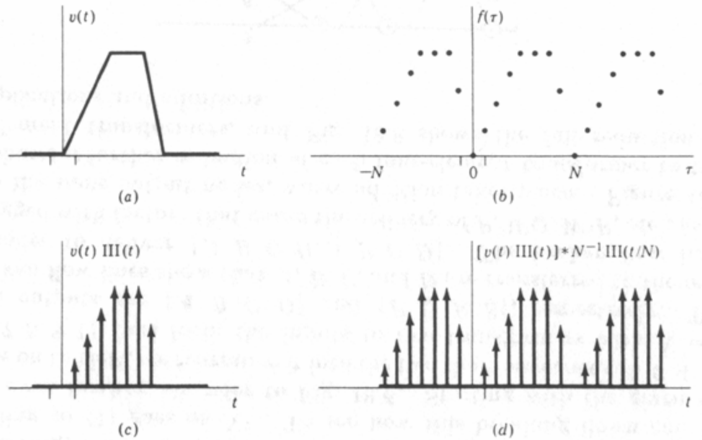


Fig. 18.9 A function (a) and its discrete representation (b) on a cyclic basis. Parts (c) and (d) show two pulse waveforms, as functions of continuous time, that are equivalent to the discrete representation.

If the Fourier transform of $v(t)$ is $S(f)$, then we may use the convolution theorem and knowledge that the *shah* function III(·) transforms into itself to obtain

$$P(f) = [S(f) * \mathrm{III}(f)]\mathrm{III}(Nf).$$

This expression has the same exact correspondence with the DFT $F(\nu)$ that $p(t)$ has with $f(\tau)$, provided that we use the relation $f = \nu/N$ to translate between f and ν. The factor N^{-1} in (3) is accounted for. Thus the rather simple algebra developed in Chapter 10 includes as a special case discrete situations that appear at first to be outside the scope of the integral transform.

If we now replicate the same data string $v(t)\mathrm{III}(t)$ with a period $2N$, we have $[v(t)\mathrm{III}(t)] * (2N)^{-1}\mathrm{III}(t/2N)$ and the transform becomes

$$[S(f) * \mathrm{III}(f)]\mathrm{III}(2Nf).$$

The only difference is that the same expression $S(f) * \mathrm{III}(f)$ is sampled twice as closely in frequency. How, then, could an apparently improved sampling be perceived as different? The answer is that $S(f)$ may be oscillatory and indeed normally will be, unless $f(\tau)$ is free from large jumps such as those that often occur at the beginning and end of data strings. Of course, because of the cyclic character of $f(\tau)$, a large initial value $f(0)$ will not count as a large jump if the final value $f(N - 1)$ is approximately equal. But if such a data string of 64 elements were extended to 128 elements by the addition of trailing zeros, there *would* be jumps. Rough structure would then appear in $F(\nu)$. Likewise, if four consecutive elements of the relatively smooth 64-element string were put to zero, oscillations would appear in $F(\nu)$. This suggests that adding zeros might not always be the best way to pack a data string out to 64 elements. A result more in keeping with expectation might result from packing with dummy values having more plausibility as data than zeros would have.

Is the discrete Fourier transform correct?

While the theory of the DFT is precise and self-consistent and exactly describes the manipulations performed on actual data samples when a Fourier transform is to be computed, the question remains to what degree the DFT approximates the Fourier transform of the function underlying the data samples. Clearly, the DFT can be only an approximation, since it provides only for a finite set of discrete frequencies. But will these discrete values themselves be correct? It is easy to show simple

cases where they are not. Discussion of the question can be based on the sampling theorem and the phenomenon of aliasing. If the initial samples are not sufficiently closely spaced to represent high-frequency components present in the underlying function, then both the DFT values and a smooth curve passing through them will be falsified by aliasing. If the underlying function is known, then the error associated with a given choice of sampling interval is calculable. If one takes the operational viewpoint that the measured data samples may be the only knowledge we have, then avoidance of error will depend on experimental factors such as prior knowledge or experience. For example, a run with twice the number of samples in the same time could confirm the presence or absence of higher frequencies.

A further important source of error in the DFT lies in the truncation of data strings. It is, of course, unavoidable that truncation of a function will result in an erroneous Fourier transform (the result obtained will be the convolution of the true Fourier transform with a certain sinc function), and so truncation error is not specific to the DFT. However, the error committed will be different. To see this, imagine a case where the sampling interval is quite fine enough to cope with the highest frequencies present in the data, so that there is no aliasing error. Now truncate the data. The effect on the DFT will be to convolve it with samples of the sinc function corresponding to the width of the rectangle-function factor describing the truncation. But this time we are convolving with an entity such as $Q(f)$ in Fig. 18.1. In addition to the smoothing out we now have the prospect of the left and right islands of $Q(f)$ leaking into the central island. The truncation effect thus comprises both smoothing error, or reduction of fine detail in the DFT, and leakage error. Leakage error may be reduced, at the expense of increased smoothing error, by use of a tapered truncation factor in place of the rectangle-function factor. The best compromise must depend on the case; leakage error tends to falsify the "higher" frequencies (ν in the neighborhood of $N/2$), whereas smoothing error is distributed differently.

Applications of the FFT

In some subjects, such as X-ray diffraction and radio interferometry, the observational data require Fourier transformation in order to be presented in customary ways, such as a molecular shape, a crystal structure, or a brightness distribution map of a celestial source. In these fields the introduction of the FFT merely speeds up what was already practiced.

In other applications, one takes the Fourier transform in order to perform some operation on it and then retransforms. For example, if we had a photographic enlargement that was particularly grainy (i.e., finely speckled because of the grain structure within the photographic emulsion) we might subject the photograph to two-dimensional low-pass filtering. First we would digitize it into a two-dimensional array of numbers, although it might already be in digitized form (e.g., if it had been received by radio telemetry from a space probe). Then we would take the two-dimensional DFT and remove or reduce the higher spatial frequencies by multiplication with a suitable low-pass transfer function. Finally, we would invert the DFT. Of course this would be equivalent to convolving the digitized photograph with the appropriate point-source response (inverse DFT of the transfer function). For desk calculation one convolution may be more attractive than two DFT's and one set of multiplications. But with the large quantities of data that a photograph normally contains, a larger computer would be required, and it would then be found that the DFT route is quicker if the FFT is used. The reason is that, if there are N elements in the array of data, the number of multiplications required is of the order of N^2, whereas as we have seen, the FFT requires far fewer if N is large.

Thus convolution in general, including cross correlation and autocorrelation, is now best performed by taking the two DFT's, multiplying, and inverting the DFT. Some special considerations arise. Consider first a case where (Fig. 18.10) the two inputs f and g to be convolved have the same number of elements, as happens with autocorrelation. The output function will have twice as many elements as the input functions. So if one simply multiplies F and G term by term and retransforms, the output will be only half the correct length. The result of this procedure can be visualized in terms of cyclic convolution as in the example of autocorrelation shown in Fig. 18.4. What will happen is that the output sequence will close around the circle and overlap itself. Clearly this can be avoided, as in the figure, by packing the given functions with zeros; enough to double the length of the given sequences will suffice. Figure 18.11 brings out these practical points in a way that is glossed over in Fig. 18.9.

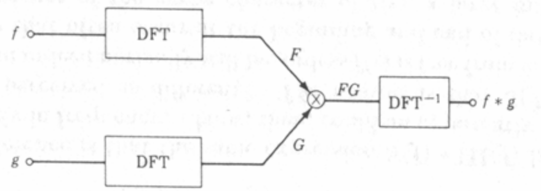


Fig. 18.10 Flow diagram for convolving.

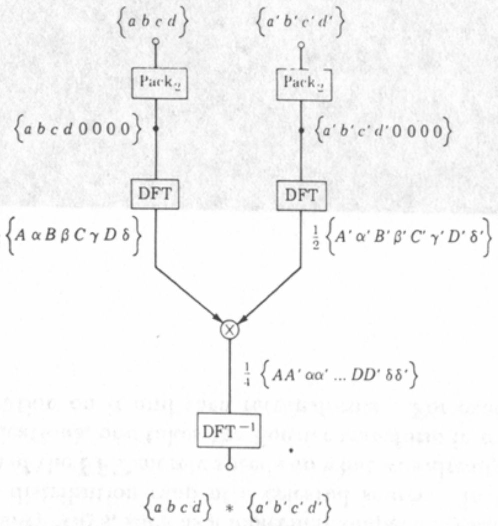


***Fig. 18.11** Convolution of two four-element sequences performed by using the DFT.*

Two-dimensional data

Let us compare the standard form of the two-dimensional Fourier transform

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-i2\pi(ux+vy)}\,dx\,dy$$

with the two-dimensional discrete Fourier transform

$$F(\mu,\nu) = M^{-1}N^{-1}\sum_{\sigma=0}^{M-1}\sum_{\tau=0}^{N-1} f(\sigma,\tau)e^{-i2\pi(\mu\sigma/M+\nu\tau/N)}.$$

The integers σ and τ may be connected with the (x,y) plane as follows. If the sampling intervals are X and Y, and $x_{\min}$ and $y_{\min}$ are the minimum values of x and y to be considered, then

$$\sigma = \frac{x - x_{\min}}{X}$$

$$\tau = \frac{y - y_{\min}}{Y}.$$

Since $M - 1$ and $N - 1$ are the largest values reached by σ and τ, respectively, it follows that

$$x_{\max} = x_{\min} + (M-1)X$$
$$y_{\max} = y_{\min} + (N-1)Y.$$

The spatial frequency integers μ and ν are such that μ/N and ν/M are spatial frequencies measured in cycles per sampling interval of x and y and μ/NX and ν/NY are spatial frequencies measured in cycles per unit of x and y. This discussion pictures $f(\sigma,\tau)$ as a function that possesses values in between its discrete samples but presumes those values to be unavailable. That situation often arises, which is why a connection with the integral transform has been established here. But we also understand that it is not necessary to regard $f(\sigma,\tau)$ as other than a function of integer pairs only and that μ and ν need not be regarded as frequencies. In fact, as previously noted in one dimension, great care has to be taken in interpreting μ and ν as frequencies.

Whereas the integral transform covers positive and negative areas of the (x,y) plane, the discrete transform does not require negative values of σ and τ. Consequently, a simple object in the (x,y) plane as in Fig. 18.12*a* becomes carved up in a strange way on the (σ,τ) plane when the shifts of origin are made. It is very helpful to have the topology of this figure in mind when handling two-dimensional data. For example, the idea of surrounding the object with a guard zone of zeros, shown cross-

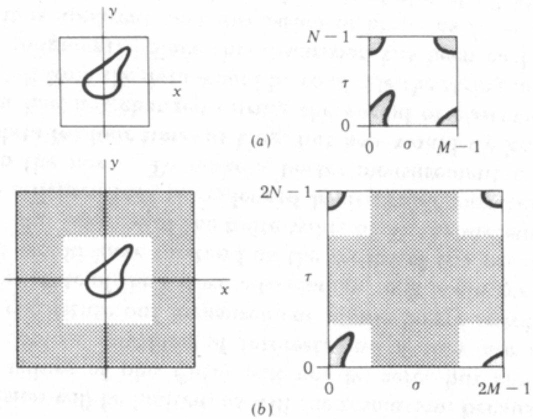


Fig. 18.12 (a) When $f(x,y)$ is forced into an $M \times N$ array with positive subscripts starting from zero, it appears on the (σ,τ) plane dissected as shown; (b) a surround of zeros in the (x,y) plane (shown shaded) is not entered as a surround in the (σ,τ) plane.

hatched, a trivial move in the (x,y) plane, not requiring reassignment of existing data, calls for a rather tricky maneuver on the (σ,τ) matrix as shown in Fig. 18.12*b*.

If, by oversight or design, the σ and τ axes are not taken to coincide with the nominal axes x and y, then the transform obtained will be affected. For example, an object symmetrical with respect to the x and y axes will have a real-valued transform, but complex values will result if the σ and τ axes are shifted. If, not realizing this, one reads out the real transform values only, they will be wrong. However, if one reads out the complex values they will differ only in a trivial way from the nominal transform. The modulus will be correct and the phase will advance linearly with σ and τ in the way controlled by the shift theorem.

Power spectra

In many situations where transform phase is unimportant or unknowable, one could deal with $|F(\nu)|$, but it is customary to deal with $|F(\nu)|^2$, which is equivalent, and to refer to $|F(\nu)|^2$ as the power spectrum. Values of $|F(\nu)|^2$ may indeed represent a number of watts in some applications, but even where the physical significance is not power or where there is no physical significance at all, the term "power spectrum" is in common use. The term is also used in connection with the Fourier transform $S(f)$ of a function of continuous time (pp. 113, 115) but there is a distinction. A power spectrum $|S(f)|^2$ would be measured in units of watts per hertz or something more complicated (such as ohm-watts per hertz as in Rayleigh's theorem on p. 183) but never in plain watts.

At first sight it might seem unnecessary to give special computational attention to the power spectrum, which is, after all, included within the larger concept of the (complex) Fourier transform. But, in fact, a great deal has been written under the heading of power spectra, where *power* spectra per se have not actually been of the essence. The literature referred to might equally well have been entitled spectra of random functions of time. The reason for the terminology is that random functions or noise waveforms present one of the important situations where phase loses meaning and the power spectrum becomes the natural entity to work with.

Although the power spectrum has a wider range of application than just to random processes or to deterministic signals of random origin, these applications are nevertheless important. The power spectrum of a random *process* is quite often defined as the Fourier transform of the autocovariance function of the process. (The autocovariance is what results when any d-c component is subtracted before autocorrelation is performed. But the distinction in terminology is not universally observed because it is commonly understood that any nonzero mean level is to be subtracted before calculating the autocorrelation, since the calculation is impossible otherwise.)

A definition of power spectrum in terms of autocorrelation (or autocovariance) seems indirect to many students but does permit the solution of problems involving random processes that are specified in the time domain by probabilities. When computation is required, however, one is never dealing with a random *process* but with an actual data string, possibly of random origin in some sense. (There may, in addition, be random errors of measurement.)

The curious fact that makes computation of power spectra so interesting is this. Suppose one takes the DFT of a data string of N elements. To be concrete, let the N data values be the height of the sea surface at a certain point, taken at 10-second intervals. Naturally, the values of $F_N(\nu)$ ought to show in what frequency bands the wave power resides, but the precision will be limited, as will the resolution, because N is only finite. The values of pha $F_N(\nu)$ will not be zero, but can hardly be expected to contain anything of interest, and if they are abandoned, $|F_N(\nu)|^2$ will constitute our measurement of the power spectrum of the waves. If the state of the sea were to change, as it is always doing, that measurement would have to stand as the record of the sea spectrum at that epoch. But because of the finite value of N, the measurements are imperfect to a degree that is evidenced by irregular variation from one value of ν to the next. To make a better measurement next time we might take data for four times as long, but how would we know that the sea spectrum had not changed during the period of observation? The only way to tell from the data would be to divide the string into segments and make a judgment. Since this discussion has been cast in terms of sea waves, it is apparent that discussion of limits as $N \rightarrow \infty$ would be inappropriate, but so would it be in the case of almost all kinds of data. One can conceive of exceptions such as determination of the power spectrum of the data string constituted by the consecutive digits of the decimal expansion of π, but in the physical world things change if an observation takes too long. Yet experience might suggest that quadrupling N, thus staying far short of infinity, would double our precision or nearly so (i.e., the irregular variation might be halved). The strange thing is that the precision is not improved at all by increasing N. Thus, even in theory, the idea of defining the power spectrum as a limit as $N \rightarrow \infty$, does not work for data of random origin.

One might suspect that the paradox disappears if the situation is viewed in the right way. Here is an explanation. In any frequency band, chosen in advance, the amount of power will indeed be measured with increased precision as N is increased. If N is quadrupled the DFT will supply four values in a fixed frequency band that previously contained only one. Even though these four values are no more precise than the previous one, the sum of the four, which represents the new measurement of the power in the band, will have greater precision.

This correct view illuminates the procedure followed for computing power spectra of real data. The computed values of $|F_N(\nu)|^2$ will fall above and below some general trend with ν, and the scatter may be reduced by averaging several adjacent values. If high precision is sought by averaging too many consecutive values, the precision is paid for by loss of resolution in frequency, so a compromise must be arrived at by judgment based on experience with the character of the data. No unique advice can be offered by theory alone; that is why a variety of prescriptions can be found in the literature. In any case, the outcome is to smooth the sequence $|F(\nu)|^2$ by taking running means over a certain number of values, that is, by discrete convolution in the power spectrum domain with some sequence of weights.

Naturally several smoothing sequences have been proposed, but which is optimum? The answer to this depends on the purpose of the analysis and on the character of the data. Although it is true that smoothing increases precision, there are a number of accompanying effects that may be undesirable. For example, if there is a narrow spectral feature that is of interest, then extra smoothing will give an erroneous low value for the central strength, an erroneous large value for the width and may introduce lobe structure on each side. There are cases where absolute strength measurement is important as in chemical spectral analysis performed by Fourier transform spectroscopy, other cases where it is important to separate close spectral features, and others where there is a heavy penalty for false detection of faint features. In the latter case one may suppress lobes that might be counted as real and accept the accompanying loss of resolution as represented by widening of the spectral feature. In another case one might accept lobe structure in order to get a better strength measurement at a frequency peak. Even when such costs and benefits are balanced to the user's satisfaction, the result will not necessarily be optimum for a new batch of data. It is apparent that selection of smoothing sequences goes beyond the realm of mathematical analysis to involve experience and judgment.

In principle, smoothing in the power spectrum domain is achievable by multiplying the autocorrelation function by a tapering factor. The term *lag window* for such a factor applied to the autocorrelation function was introduced by Blackmann and Tukey. (*Spectral window* is the Fourier transform of the lag window.) However, when large amounts of data are involved, it is convenient to compute the autocorrelation by