1

TRANSFORMS

The first step in data analysis is to learn how to represent and manipulate waveforms in a digital computer. Time and space are ordinarily regarded as continuous, but for purposes of computer analysis we must discretize them. This discretizing is also called digitizing or sampling. Discretizing continuous functions may at first be regarded as an evil that is necessary only because our data are not always known analytic functions. However, after gaining some experience with sampled functions, one realizes that many mathematical concepts are easier with sampled time than with continuous time. For example, in this chapter the concept of the Z transform is introduced and is shown to be equivalent to the Fourier transform. The Z transform is readily understood on a basis of elementary algebra, whereas the Fourier transform requires substantial experience in calculus.

1-1 SAMPLED DATA AND Z TRANSFORMS

Consider the time function graphed in Fig. 1-1.

To analyze such an observed time function in a computer it is necessary to approximate it in some way by a list of numbers. The usual way to do this is to evaluate

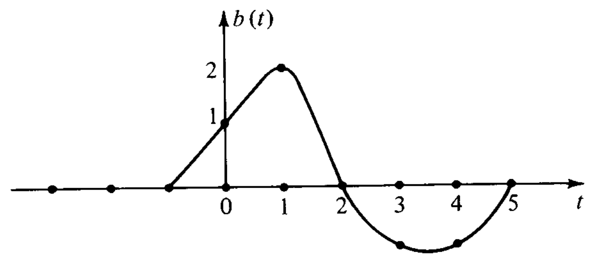

FIGURE 1-1
A continuous time function sampled at uniform time intervals.

or observe $b(t)$ at a uniform spacing of points in time. For this example, such a digital approximation to the continuous function could be denoted by the vector

$$b_t = (\ldots 0, 0, 1, 2, 0, -1, -1, 0, 0, \ldots)$$

Of course if time points were taken more closely together we would have a more accurate approximation. Besides a vector, a function can be represented as a polynomial where the *coefficients* of the polynomial represent the values of $b(t)$ at successive time points. In this example we have

$$B(Z) = 1 + 2Z + 0Z^2 - Z^3 - Z^4 \qquad (1\text{-}1\text{-}1)$$

This polynomial is called a Z transform. What is the meaning of Z in this polynomial? The meaning is not that Z should take on some numerical value; the meaning of Z is that it is the unit delay operator. For example the coefficients of $ZB(Z) = Z + 2Z^2 - Z^4 - Z^5$ are plotted in Fig. 1-2. It is the same waveform as in Fig. 1-1, but it has been delayed.
We see that the time function b_t is delayed n time units when $B(Z)$ is multiplied by Z^n. The delay operator Z is very important in analyzing waves simply because waves take a certain amount of time to get from place to place.

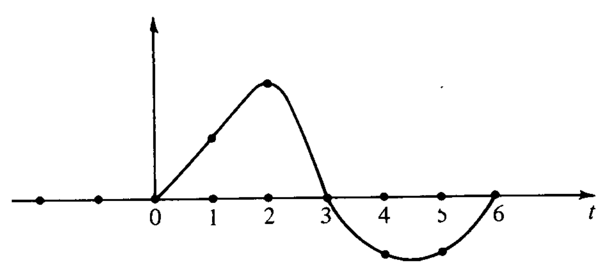

FIGURE 1-2
Coefficients of $Z\,B(Z)$ are a shifted version of the coefficients of $B(Z)$.

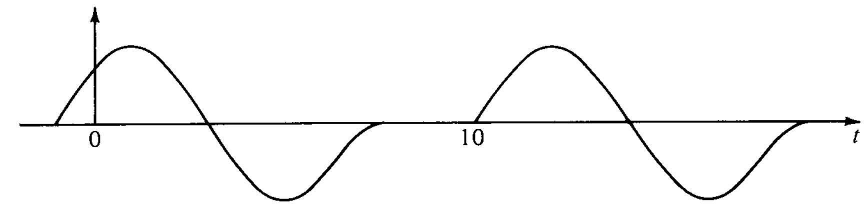

FIGURE 1-3
Response to two explosions.

Another value of the delay operator is that it may be used to build up more complicated time functions from simpler ones. Suppose $b(t)$ represents the acoustic pressure function or the seismogram observed after a distant explosion. Then $b(t)$ is called the *impulse response*. If another explosion occurs at $t = 10$ time units after the first, we expect the pressure function $y(t)$ depicted in Fig. 1-3.

In terms of Z transforms this would be expressed as $Y(Z) = B(Z) + Z^{10}B(Z)$. If the first explosion were followed by an implosion of half strength, we would have $B(Z) - \frac{1}{2}Z^{10}B(Z)$. If pulses overlap one another in time [as would be the case if $B(Z)$ was of degree greater than 10], the waveforms would just add together in the region of overlap. The supposition that they just add together without any interaction is called the *linearity assumption*. This linearity assumption is very often true in practical cases. In seismology we find that—although the earth is a very heterogeneous conglomeration of rocks of different shapes and types—when seismic waves (of usual amplitude) travel through the earth, they do not interfere with one another. They satisfy linear superposition. The plague of nonlinearity arises from large amplitude disturbances. Nonlinearity does not arise from geometrical complications.

Now suppose there was an explosion at $t = 0$, a half-strength implosion at $t = 1$, and another, quarter-strength explosion at $t = 3$. This sequence of events determines a "source" time series, $x_t = (1, -\frac{1}{2}, 0, \frac{1}{4})$. The Z transform of the source is $X(Z) = 1 - \frac{1}{2}Z + \frac{1}{4}Z^3$. The observed y_t for this sequence of explosions and implosions through the seismometer has a Z transform $Y(Z)$ given by

$$\begin{aligned} Y(Z) &= B(Z) - \frac{Z}{2}B(Z) + \frac{Z^3}{4}B(Z) \\ &= \left(1 - \frac{Z}{2} + \frac{Z^3}{4}\right)B(Z) \qquad (1\text{-}1\text{-}2) \\ &= X(Z)B(Z) \end{aligned}$$

The last equation illustrates the underlying basis of linear-system theory that the output $Y(Z)$ can be expressed as the input $X(Z)$ times the impulse response $B(Z)$.

There are many examples of linear systems. A wide class of electronic circuits is comprised of linear systems. Complicated linear systems are formed by taking the output of one system and plugging it into the input of another. Suppose

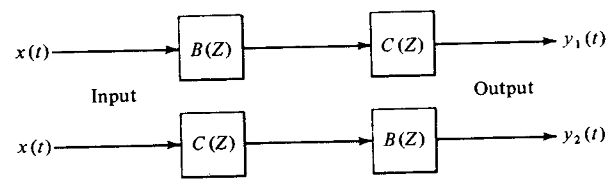

FIGURE 1-4
Two equivalent filtering systems.

we have two linear systems characterized by $B(Z)$ and $C(Z)$, respectively. Then the question arises whether the two combined systems of Fig. 1-4 are equivalent. The use of Z transforms makes it obvious that these two systems are equivalent since products of polynomials commute, i.e.,

$$Y_1(Z) = [X(Z)B(Z)]C(Z) = XBC \qquad (1\text{-}1\text{-}3)$$

$$Y_2(Z) = [X(Z)C(Z)]B(Z) = XCB = XBC \qquad (1\text{-}1\text{-}4)$$

Consider a system with an impulse response $B(Z) = 2 - Z - Z^2$. This polynomial can be factored into $2 - Z - Z^2 = (2 + Z)(1 - Z)$, and so we have the three equivalent systems in Fig. 1-5. Since any polynomial can be factored, any impulse response can be simulated by a cascade of two-term filters (impulse responses whose Z transforms are linear in Z).

What do we actually do in a computer when we multiply two Z transforms together? The filter $2 + Z$ would be represented in a computer by the storage in memory of the coefficients (2, 1). Likewise, for $1 - Z$ the numbers (1, −1) are stored. The polynomial multiplication program should take these inputs and produce the sequence (2, −1, −1). Let us see how the computation proceeds in a general case, say

$$X(Z)B(Z) = Y(Z) \qquad (1\text{-}1\text{-}5)$$

$$(x_0 + x_1 Z + x_2 Z^2 + \cdots)(b_0 + b_1 Z + b_2 Z^2) = (y_0 + y_1 Z + y_2 Z^2 + \cdots) \qquad (1\text{-}1\text{-}6)$$

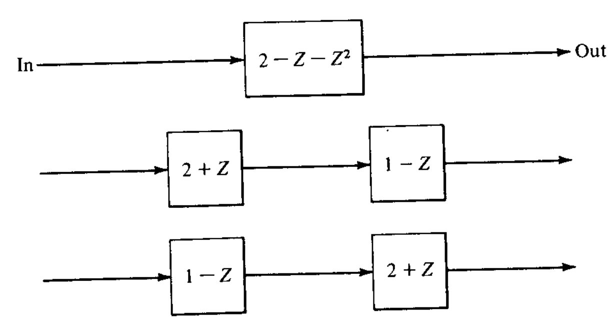

FIGURE 1-5
Three equivalent filtering systems.


```
      DIMENSION X(LX),B(LB),Y(LY)
      LY = LX+LB-1
      DO 10 I=1,LY
10    Y(I) = 0.
      DO 20 I=1,LX
      DO 20 J=1,LB
20    Y(I+J-1) = Y(I+J-1) + X(I)*B(J)
```

FIGURE 1-6
A computer program to do convolution.

Identifying coefficients of successive powers of Z, we get

$$\begin{aligned} y_0 &= x_0 b_0 \\ y_1 &= x_1 b_0 + x_0 b_1 \\ y_2 &= x_2 b_0 + x_1 b_1 + x_0 b_2 \\ y_3 &= x_3 b_0 + x_2 b_1 + x_1 b_2 \\ y_4 &= x_4 b_0 + x_3 b_1 + x_2 b_2 \end{aligned} \qquad (1\text{-}1\text{-}7)$$

$$y_k = \sum_{i=0}^{2} x_{k-i} b_i \qquad (1\text{-}1\text{-}8)$$

Equation (1-1-8) is called a *convolution equation*. Thus, we may say that the product of two polynomials is another polynomial whose coefficients are found by convolution. A simple Fortran computer program which does convolution, including end effects on both ends, is shown in Fig. 1-6. The reader should notice that $X(Z)$ and $Y(Z)$ need not strictly be polynomials; they may contain both positive and negative powers of Z; that is,

$$\begin{aligned} X(Z) &= \cdots \frac{x_{-2}}{Z^2} + \frac{x_{-1}}{Z} + x_0 + x_1 Z + \cdots \\ Y(Z) &= \cdots \frac{y_{-2}}{Z^2} + \frac{y_{-1}}{Z} + y_0 + y_1 Z + \cdots \end{aligned} \qquad (1\text{-}1\text{-}9)$$

The effect of using negative powers of Z in $X(Z)$ and $Y(Z)$ is merely to indicate that data are defined before $t = 0$. The effect of using negative powers of Z in the filter is quite different. Inspection of (1-1-8) shows that the output y_k which occurs at time k is a linear combination of current and previous inputs; that is, $(x_i, i \leq k)$. If the filter $B(Z)$ had included a term like b_{-1}/Z, then the output y_k at time k would be a linear combination of current and previous inputs and x_{k+1}, an input which really has not arrived at time k. Such a filter is called a *nonrealizable* filter because it could not operate in the real world where nothing can respond now to an excitation which has not yet occurred. However, nonrealizable filters are occasionally useful in computer simulations where all of the data are prerecorded.

EXERCISES

1 Let $B(Z) = 1 + Z + Z^2 + Z^3 + Z^4$. Graph the coefficients of $B(Z)$ as a function of the powers of Z. Graph the coefficients of $[B(Z)]^2$.

2 If $x_t = \cos \omega_0 t$, where t takes on integral values $b_t = (b_0, b_1)$ and $Y(Z) = X(Z)B(Z)$, what are A and B in $y_t = A \cos \omega_0 t + B \sin \omega_0 t$?

3 Deduce that, if $x_t = \cos \omega_0 t$ and $b_t = (b_0, b_1, \ldots, b_n)$, then y_t always takes the form $A \cos \omega_0 t + B \sin \omega_0 t$.

1-2 Z-TRANSFORM TO FOURIER TRANSFORM

We have defined the Z transform as

$$B(Z) = \sum_t b_t Z^t \qquad (1\text{-}2\text{-}1)$$

If we make the substitution $Z = e^{i\omega}$ we have a "Fourier sum"

$$B(Z) = B(e^{i\omega}) = \sum_t b_t e^{i\omega t} \qquad (1\text{-}2\text{-}2)$$

This is like a Fourier integral, and we could obviously do a limiting operation to make it into an integral. Another point of view is that the Fourier integral

$$B(\omega) = \int_{-\infty}^{+\infty} b(t) e^{i\omega t}\, dt \qquad (1\text{-}2\text{-}3)$$

reduces to the sum (1-2-2) when $b(t)$ is not a continuous function of time but is defined as

$$b(t) = \sum_k b_k \delta(t - k) \qquad (1\text{-}2\text{-}4)$$

where δ is the Dirac delta function.

In the last section we saw that to multiply two polynomials the coefficients must be convolved. The same process in Fourier transform language is that a product in the frequency domain corresponds to a convolution in the time domain.

Although one thinks of a Fourier transform as an integral which may be difficult or impossible to do, the Z transform is always easy, in fact trivial. To do a Z transform one merely attaches powers of Z to successive data points. When one has $B(Z)$ one can refer to it either as a time function or a frequency function, depending on whether one graphs the polynomial coefficients or if one evaluates and graphs $B(Z = e^{i\omega})$ for various frequencies ω. The reader should observe that as ω goes from zero to 2π, $Z = e^{i\omega} = \cos \omega + i \sin \omega$ migrates once around the unit circle in the counterclockwise direction.

If taking a Z transform amounts to attaching powers of Z to successive points of a time function, then the inverse Z transform must be merely identifying coefficients of various powers of Z with different points in time. How can this simple "identification of coefficients" be the same as the apparently more complicated operation of inverse Fourier integrals? The inverse Fourier integral is

$$b(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} B(\omega) e^{-i\omega t}\, d\omega \qquad (1\text{-}2\text{-}5)$$

First notice that the integration of Z^n about the unit circle or $e^{in\omega}$ over $-\pi \le \omega < +\pi$ gives zero unless $n = 0$ because cosine and sine are oscillatory; that is,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{in\omega}\, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\cos n\omega + i \sin n\omega)\, d\omega = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n = \text{non-zero integer} \end{cases} \qquad (1\text{-}2\text{-}6)$$

In terms of our discretized time functions, the inverse Fourier integral (1-2-5) is

$$b_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\cdots + b_{-1} e^{-i\omega} + b_0 + b_1 e^{+i\omega} + \cdots) e^{-i\omega t}\, d\omega \qquad (1\text{-}2\text{-}7)$$

Of all the terms in the integrand (1-2-7) we see by (1-2-6) that only the term with b_t will contribute to the integral; all the rest oscillate and cancel. In other words, it is only the coefficient of Z to the zero power which contributes to the integral, reducing (1-2-7) to

$$b_t = \frac{1}{2\pi} b_t \int_{-\pi}^{+\pi} d\omega = b_t \qquad (1\text{-}2\text{-}8)$$

This shows how inverse Fourier transformation is just like identifying coefficients of powers of Z.

In this book and many others, it is common to assume that the time span between data samples $\Delta t = 1$ is unity. To adapt given equations to other values of Δt, one only need replace ω by $\omega\, \Delta t$; that is,

$$\omega_{\text{book}} = \omega_{\text{book}} \Delta t_{\text{book}} = \omega_{\text{true}} \Delta t_{\text{true}} \qquad (1\text{-}2\text{-}9)$$

With Z transforms we have the spectrum given on a range of 2π for ω_{book}. In the limit Δt_{true} goes to zero, ω_{true} has the same infinite limits as the Fourier integral.

When a continuous function is approximated by a sampled function, it is necessary to take the sample spacing Δt_{true} small enough. The basic result of elementary texts is that, if there is no appreciable energy in a Fourier transform for frequencies higher than some frequency $\omega_{\max}$, then there is no appreciable loss of information if the sample spacing is $\Delta t = \pi/\omega_{\max}$. In other words, a cosine wave must be sampled at least two points per wavelength. Figure 1-7*a* shows how insufficient sampling of a sine wave often causes it to appear as a sine wave of lower frequency.

Next we wish to examine odd/even symmetries to see how they are affected in Fourier transformation. The even part e_t of a time function b_t is defined as

$$e_t = \frac{b_t + b_{-t}}{2} \qquad (1\text{-}2\text{-}10)$$

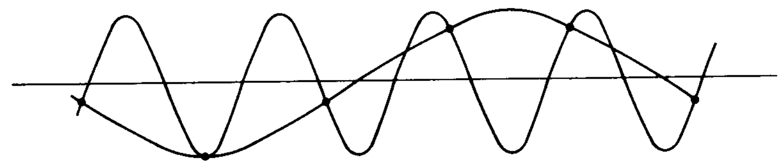

FIGURE 1-7*a*
If a high-frequency sinusoid is sampled insufficiently often, it becomes indistinguishable from a lower-frequency sinusoid. For this reason $\omega_{max} = \pi/\Delta t$ is said to be the folding frequency, as higher frequencies are folded down to look like lower frequencies. In practice, quasi-sinusoidal waves are always sampled more frequently than twice per wavelength. Good theoretical reasons for sampling eight or more points per wavelength are developed on pp. 44 to 47.

The odd part is

$$o_t = \frac{b_t - b_{-t}}{2} \qquad (1\text{-}2\text{-}11)$$

A function is the sum of its even and odd parts. By adding (1-2-10) and (1-2-11), we get

$$b_t = e_t + o_t \qquad (1\text{-}2\text{-}12)$$

Consider a simple, real, even time function such as $(b_{-1}, b_0, b_1) = (1, 0, 1)$. Its transform $Z + 1/Z = 2\cos\omega$ is an even function of ω since $\cos\omega = \cos(-\omega)$. Consider the real, odd time function $(b_{-1}, b_0, b_1) = (-1, 0, 1)$. Its transform $Z - 1/Z = 2(\sin\omega)/i$ is imaginary and odd, since $\sin\omega = -\sin(-\omega)$. Likewise, the transform of the imaginary even function $(i, 0, i)$ is the imaginary even function $i\cos\omega$ and the transform of the imaginary odd function $(-i, 0, i)$ is real and odd. Let r and i refer to real and imaginary, *e* and *o* refer to even and odd, and lower-case and upper-case refer to time and frequency functions. A summary of the symmetries of Fourier transformation is shown in Fig. 1-7*b*.

More elaborate time functions can be made up by adding together the two point functions we have considered. Since sums of even functions are even, and so on, the table of Fig. 1-7*b* applies to all time functions. Note that an arbitrary time function takes the form $b_t = (\text{re} + \text{ro}) + i(\text{ie} + \text{io})_t$. On transformation of b_t, each of the four individual parts transforms according to the table.

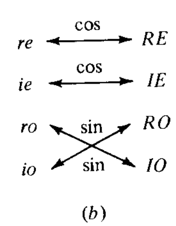

FIGURE 1-7*b*
Mnemonic table illustrating how even/odd and real/imaginary properties are affected by Fourier transformation.

EXERCISES

1 Normally a function is specified entirely in the time domain or entirely in the frequency domain. When one is known, the other is determined by transformation. Now let us give half the information in the time domain by specifying that $b_t = 0$ for $t < 0$, and half in the frequency domain by giving the real part RE + RO in the frequency domain. How can you determine the rest of the function?

1-3 THE FAST FOURIER TRANSFORM

When we write the expression

$$B(Z) = b_0 + b_1 Z + \cdots \qquad (1\text{-}3\text{-}1)$$

we have both a time function and its Fourier transform. If we plot the coefficients $(b_0, b_1, \ldots)$, we plot the time function. If we evaluate and plot (1-3-1) at numerous real ω, then we have plotted the transform. (Note that for real ω, Z is of unit magnitude; i.e., on the unit circle.) Since ω is a continuous variable and everything in a computer is finite, how do we select a finite number of values ω_k for plotting? The usual choice is to take evenly spaced frequencies. The lowest frequency can be zero. [Note $Z(\omega = 0) = e^{i0} = 1$.] A frequency as high as $\omega = 2\pi$ [note $Z(\omega = 2\pi) = e^{i2\pi} = 1$ also] need not be considered, since (1-3-1) gives the same value for it as for zero frequency. Choosing uniformly spaced frequencies between these limits we have

$$\omega_k = \frac{(0, 1, 2, \ldots, M-1)2\pi}{M} \qquad (1\text{-}3\text{-}2)$$

where M is some integer. Now let us abbreviate $B(Z(\omega_k))$ as B_k.

For the special case of an N-point time function where $N = 4$, (1-3-1) may be expressed by the matrix multiplication

$$\begin{bmatrix} B_0 \\ B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & W & W^2 & W^3 \\ 1 & W^2 & W^4 & W^6 \\ 1 & W^3 & W^6 & W^9 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad (1\text{-}3\text{-}3)$$

where

$$W = e^{2\pi i/N} \qquad (1\text{-}3\text{-}4)$$

It is not essential to choose $N = M$ as we have done in (1-3-3), but it is a convenience. There is no loss of generality because one may always append zeros to a time function before inserting it into (1-3-3). A convenience of the choice $N = M$ is that the matrix in (1-3-3) will then be square and there will be an exact inverse. In fact, the inverse to (1-3-3) may be easily shown to be

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = 1/N \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1/W & 1/W^2 & 1/W^3 \\ 1 & 1/W^2 & 1/W^4 & 1/W^6 \\ 1 & 1/W^3 & 1/W^6 & 1/W^9 \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ B_2 \\ B_3 \end{bmatrix} \qquad (1\text{-}3\text{-}5)$$

Since $1/W$ is the complex conjugate of W, the matrices of (1-3-3) and (1-3-5) are just complex conjugates of one another. In fact, one observes no fundamental mathematical difference between time functions and frequency functions. This "duality" would be even more complete if we had used a scale factor of $N^{-1/2}$ in each of (1-3-3) and (1-3-5) rather than 1 in (1-3-3) and N^{-1} in (1-3-5). Note also that time functions and frequency functions could be interchanged in the mnemonic table describing symmetries. In fact, our earlier observation that the product of two frequency functions amounts to a convolution of the corresponding two time functions has a dual statement that the product of two time functions corresponds to the convolution of the corresponding two frequency functions. We will not "prove" this duality as it is standard fare in both mathematics and systems theory books. However we will occasionally call upon the reader to realize that in any theorem the meanings of "time" and "frequency" may be interchanged.

In making a plot of the transform B_k for ($k = 0, 1, \ldots, M-1$) the frequency axis ranges as $0 \le \omega_k < 2\pi$. It is often more natural to display the interval $-\pi \le \omega < \pi$. Since the transform is periodic with period 2π, values of B_k on the interval $\pi \le \omega < 2\pi$ may simply be moved to the interval $-\pi \le \omega < 0$ for display.

Thus, for $N = 8$ one might plot successively

$$B_4 \quad B_5 \quad B_6 \quad B_7 \quad B_0 \quad B_1 \quad B_2 \quad B_3$$

corresponding to values of ω equal to

$$-\pi, -\frac{3\pi}{4}, -\frac{\pi}{2}, -\frac{\pi}{4}, 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$$

One advantage of this display interval is that for continuous time series which are sampled sufficiently densely in time the transform values B_k get small on both ends. If the time series is real, the real part of B_k has even symmetry about B_0; the imaginary part has odd symmetry about B_0. Then, one need not bother to display half the values. Choice of an odd value of N would enable us to put $\omega = 0$ exactly in the middle of the interval, but the reader will soon see why we stick to an even number of data points.

The matrix times vector operation in (1-3-3) requires N^2 multiplications and additions. The rest of this section describes a trick method, called the fast Fourier transform, of accomplishing the matrix multiplication in $N \log_2 N$ multiplications and additions. Since, for example, $\log_2 1024$ is 10, this is a tremendous saving in effort.

A basic building block in the fast Fourier transform is called doubling. Given a series $(x_0, x_1, \ldots, x_{N-1})$ and its sampled Fourier transform $(X_0, X_1, \ldots, X_{N-1})$ and another series $(y_0, y_1, \ldots, y_{N-1})$ and its transform $(Y_0, Y_1, \ldots, Y_{N-1})$, one finds the transform of the interlaced double-length series

$$z_t = (x_0, y_0, x_1, y_1, \ldots, x_{N-1}, y_{N-1})$$

The process of doubling is used many times during the process of computing a fast Fourier transform. As the word *doubling* might suggest, it will be convenient to suppose that N is an integer formed by raising 2 to some integer power. Suppose

$N = 8 = 2^3$. We begin by dividing our eight-point series $x_0, x_1, \ldots, x_7$ into eight different series of one point each. The Fourier transform of each of the one-point series is just the point. Next, we use doubling four times to get the transforms of the four different two point series (x_0, x_4), (x_1, x_5), (x_2, x_6), and (x_3, x_7). We use doubling twice more to get the transforms of the two different four point series (x_0, x_2, x_4, x_6) and (x_1, x_3, x_5, x_7). Finally, we use doubling once more to get the transform of the original eight-point series $(x_0, x_1, x_2, \ldots, x_7)$.

It remains to look into the details of the doubling process.

Let

$$V = e^{i2\pi/2N} = W^{1/2} \qquad (1\text{-}3\text{-}6)$$

$$V^N = e^{i\pi} = -1 \qquad (1\text{-}3\text{-}7)$$

The transforms of two N-point series are by definition

$$X_k = \sum_{j=0}^{N-1} x_j V^{2jk} \qquad (k = 0, 1, \ldots, N-1)$$

$$Y_k = \sum_{j=0}^{N-1} y_j V^{2jk} \qquad (k = 0, 1, \ldots, N-1)$$

The transform of the interlaced series $z_j = (x_0, y_0, x_1, y_1, \ldots, x_{N-1}, y_{N-1})$ is by definition

$$Z_k = \sum_{l=0}^{2N-1} z_l V^{lk} \qquad (k = 0, 1, \ldots, 2N-1)$$

To make Z_k from X_k and Y_k we require two separate formulas: one for $k = 0, 1, \ldots, N-1$, and the other for $k = N, N+1, \ldots, 2N-1$.

First

$$Z_k = \sum_{l=0}^{2N-1} z_l V^{lk} \qquad (k = 0, 1, \ldots, N-1)$$

We split the sum into two parts, noting that x_j multiplies even powers of V and y_j multiplies odd powers.

$$\begin{aligned} Z_k &= \sum_{j=0}^{N-1} x_j V^{2jk} + V^k \sum_{j=0}^{N=1} y_j V^{2jk} \\ &= X_k + V^k Y_k \end{aligned} \qquad (1\text{-}3\text{-}8)$$

We obtain the last half of the Z_k by

$$\begin{aligned} Z_k &= \sum_{l=0}^{2N-1} z_l V^{lk} \qquad (k = N, N+1, \ldots, 2N-1) \\ &= \sum_{l=0}^{2N-1} z_l V^{l(m+N)} \qquad (k - N = m = 0, 1, \ldots, N-1) \\ &= \sum_{l=0}^{2N-1} z_l V^{lm} (V^N)^l \end{aligned}$$

```
      SUBROUTINE FORK(LX,CX,SIGNI)                    2/15/69
C FAST FOURIER
C                        LX
C  CX(K) = SQRT(1/LX)  SUM  (CX(J)*EXP(2*PI*SIGNI*I*(J-1)*(K-1)/LX))
C                        J=1             FOR K=1,2,...,(LX=2**INTEGER)
      COMPLEX CX(LX),CARG,CEXP,CW,CTEMP
      J=1
      SC=SQRT(1./LX)
      DO 30 I=1,LX
      IF(I.GT.J) GO TO 10
      CTEMP=CX(J)*SC
      CX(J)=CX(I)*SC
      CX(I)=CTEMP
10    M=LX/2
20    IF(J.LE.M) GO TO 30
      J=J-M
      M=M/2
      IF(M.GE.1) GO TO 20
30    J=J+M
      L=1
40    ISTEP=2*L
      DO 50 M=1,L
      CARG=(0.,1.)*(3.14159265*SIGNI*(M-1))/L
      CW=CEXP(CARG)
      DO 50 I=M,LX,ISTEP
      CTEMP=CW*CX(I+L)
      CX(I+L)=CX(I)-CTEMP
50    CX(I)=CX(I)+CTEMP
      L=ISTEP
      IF(L.LT.LX) GO TO 40
      RETURN
      END
```

FIGURE 1-8
A program to do fast Fourier transform. Modified from Brenner. Calling this program twice returns the original data. SIGNI should be + 1. on one call and − 1. on the other. LX must be a power of 2.

$$= \sum_{l=0}^{2N-1} z_l V^{lm}(-1)^l$$

$$= \sum_{j=0}^{N-1} x_j V^{2jm} - V^m \sum_{j=0}^{N-1} y_j V^{2jm}$$

$$= X_m - V^m Y_m$$

$$Z_k = X_{k-N} - V^{k-N} Y_{k-N} \qquad (k = N, N+1, \ldots, 2N-1) \qquad (1\text{-}3\text{-}9)$$

The first machine computation with this algorithm known to the author was done by Vern Herbert, who used it extensively in the interpretation of reflection seismic data. He programmed it on an IBM 1401 computer at Chevron Standard Ltd., Calgary, Canada in 1962. Herbert never published the method. It was rediscovered and widely publicized by Cooley and Tukey in 1965. Thus it has come to be known as the Cooley and Tukey algorithm. (A good reference to literature on the subject is Ref. [9].)

EXERCISES

1 Verify that for an arbitrary $N \times N$ case the matrix of (1-3-5) is indeed the inverse of the matrix of (1-3-3).

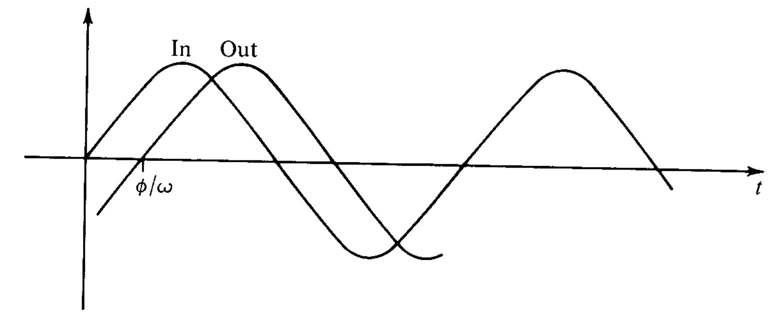

FIGURE 1-9.
A sinusoid sin ωt goes into a filter and a delayed sinusoid sin $(\omega t - \phi)$ comes out.

1-4 PHASE DELAY AND GROUP DELAY

Some filters make drastic changes to signals propagating through. Other filters do their best to make little or no change. In the latter category are transducers and recorders. In such cases, the principal form of signal change may be merely delay. One way to characterize the delay of a filter is to put in a sinusoid and compare its phase to that of the output. See Fig. 1-9.

If the input is sin ωt and the output is sin $(\omega t - \phi)$ then the so-called phase delay t_p is given by solving

$$\sin(\omega t - \phi) = \sin \omega(t - t_p)$$
$$\omega t - \phi = \omega t - \omega t_p \qquad (1\text{-}4\text{-}1)$$
$$t_p = \frac{\phi}{\omega}$$

A more interesting kind of delay is called group delay. It is analogous to group velocity in wave propagation theory. Indeed, in the modeling of wave propagation on a computer the propagation of a wave from say point A to point B may be simulated with a filter.

When the waveshape observed at A differs from that at point B but the energy envelope at A resembles with delay that at B, then we have a situation where the idea of group velocity, meaning the energy envelope velocity, may be very useful. The sum of two cosine waves of slightly differing frequencies will beat together. Refer to Fig. 1-10.

When such a waveform goes through a filter, each frequency may suffer a different delay and the result will be that the envelope or beat will have a delay which differs from the phase delay of either frequency. The envelope delay, or group delay, may not even resemble the average of the phase delays of the two frequencies. We may understand this as follows: The input waveform x_t is

$$x_t = \cos \omega_1 t + \cos \omega_2 t \qquad (1\text{-}4\text{-}2)$$

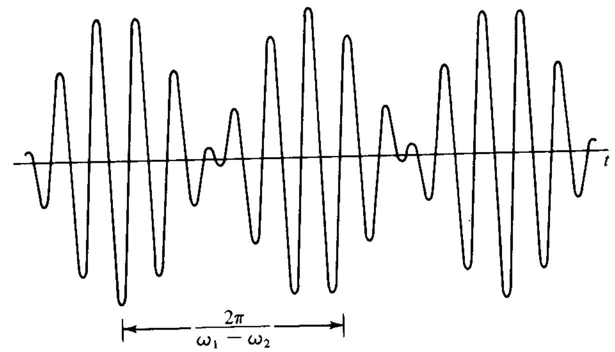

FIGURE 1-10
A graph of $\cos \omega_1 t + \cos \omega_2 t$ looks like an amplitude-modulated cosine of the average frequency.

By using a trigonometric identity

$$x_t = 2 \cos \left(\frac{\omega_1 + \omega_2}{2} t\right) \cos \left(\frac{\omega_1 - \omega_2}{2} t\right) \qquad (1\text{-}4\text{-}3)$$

we see that the sum of two cosines looks like a cosine of the average frequency multiplied by a cosine of half the difference frequency. Since the frequencies are taken close together, the difference frequency factor represents a slowly variable amplitude on the average frequency. Now let us take the output of the filter y_t to be

$$y_t = \cos(\omega_1 t - \phi_1) + \cos(\omega_2 t - \phi_2) \qquad (1\text{-}4\text{-}4)$$

In taking the output of the filter to be of the form of (1-4-4), we have assumed that neither frequency was attenuated. To allow differential attenuation of the two frequency components would greatly complicate the discussion. Utilizing the same trigonometric identity on (1-4-4), we get

$$y_t = 2 \cos \left(\frac{\omega_1 + \omega_2}{2} t - \frac{\phi_1 + \phi_2}{2}\right) \cos \left(\frac{\omega_1 - \omega_2}{2} t - \frac{\phi_1 - \phi_2}{2}\right) \qquad (1\text{-}4\text{-}5)$$

Rewriting the beat factor in terms of a time delay t_g, we have

$$\cos \left[\frac{\omega_1 - \omega_2}{2} (t - t_g)\right] = \cos \left(\frac{\omega_1 - \omega_2}{2} t - \frac{\phi_1 - \phi_2}{2}\right)$$

or

$$(\omega_1 - \omega_2) t_g = \phi_1 - \phi_2$$

or the group delay is given by

$$t_g = \frac{\phi_1 - \phi_2}{\omega_1 - \omega_2} = \frac{\Delta\phi}{\Delta\omega} \qquad (1\text{-}4\text{-}6)$$

In practice one never has two pure cosines but a band of frequencies. The group delay is then a frequency-dependent function given by $t_g = d\phi/d\omega$. The phase angle ϕ may be computed as the arctangent of the ratio of imaginary to real parts of the Fourier transform, namely $\phi(\omega) = \arctan\,[\text{Im}\, B(\omega)/\text{Re}\, B(\omega)]$. It is sometimes convenient to recall the definition of complex logarithm. Say,

$$B = re^{i\phi}$$
$$\ln B = \ln |r| + \ln e^{i\phi}$$
$$= \ln |r| + i\phi$$

So

$$\phi = \text{Im} \ln B$$
$$t_g = \frac{d\phi}{d\omega} = \text{Im} \frac{d}{d\omega} \ln B(\omega) \qquad (1\text{-}4\text{-}7)$$
$$= \text{Im} \frac{1}{B} \frac{dB}{d\omega}$$

A convenient approximation when B is sampled in a computer is

$$t_g \approx \frac{2}{\Delta\omega} \text{Im} \frac{B_{k+1} - B_k}{B_{k+1} + B_k} \qquad (1\text{-}4\text{-}8)$$

An important aspect of wave propagation theory is the distinction of phase velocity from group velocity. These are similar to phase delay and group delay. For example, if waves propagate along a two-dimensional surface, the phase function may be given by

$$\phi(x, y) = k_x(x - x_0) + k_y(y - y_0) \qquad (1\text{-}4\text{-}9)$$

Here (x_0, y_0) is the location of the filter input and (x, y) is where the phase is observed (like the filter output). The symbols k_x and k_y denote the "spatial frequencies," that is, k_x is 2π divided by the wavelength measured along the x axis. Methods of theoretical physics provide a relationship between ω and k_x and k_y. Often it can be explicitly given in the form

$$\omega = \omega(k_x, k_y) \qquad (1\text{-}4\text{-}10)$$

Since velocity is distance divided by time we can define the phase velocity along the x direction as

$$(V\ \text{phase})_x = \frac{x - x_0}{\text{phase delay}}$$
$$= \frac{x - x_0}{\phi/\omega}$$
$$= \frac{\omega}{k_x}$$

For the x component of group velocity

$$(V \text{ group})_x = \frac{x - x_0}{\text{group delay}} = \frac{x - x_0}{d\phi/d\omega} = (x - x_0)\frac{d\omega}{d\phi} \qquad (1\text{-}4\text{-}11)$$

Say $y = y_0$, then (1-4-9) reduces to

$$\phi = k_x(x - x_0)$$

which gives

$$\frac{\partial k_x}{\partial \phi} = \frac{1}{x - x_0}$$

and together with (1-4-11) gives

$$(V \text{ group})_x = (x - x_0)\frac{\partial \omega}{\partial k_x}\frac{\partial k_x}{\partial \phi} = \frac{\partial \omega}{\partial k_x} \qquad (1\text{-}4\text{-}12)$$

Thus the vector group velocity is $(\partial\omega/\partial k_x, \partial\omega/\partial k_y)$. It sometimes happens that physical theory is so complicated that an explicit relationship like (1-4-10) cannot be found and one gets instead a complicated implicit relation, say $0 = F(\omega, k_x, k_y)$. In such a case it is useful to recall the relationship from the theory of partial derivatives:

$$\frac{\partial \omega}{\partial k_x} = -\frac{\partial F/\partial k_x}{\partial F/\partial \omega}$$

In observational geophysics the velocity one deals with is nearly always the group velocity. It is the velocity with which bundles of energy move. In the example shown in Fig. 1-11 there is an excessive amount of "noise" (not unusual in observational geophysics); however, it can be seen that the disturbance first displays the long-period oscillations and then the shorter-period oscillations. The group velocity is found by dividing the distance by the time of arrival. One could observe phase velocities by having two observation stations near each other and measuring the time delay of some particular zero crossing. The reason for having the stations near one another is that the waveforms are steadily changing, and if the stations are too far apart, it may not be possible to tell which zero crossings are to be compared.

1-5 CORRELATION AND SPECTRA

The spectrum of a time function is the magnitude squared of the Fourier transform of the function. In the case of a real function, the Fourier transform has an even

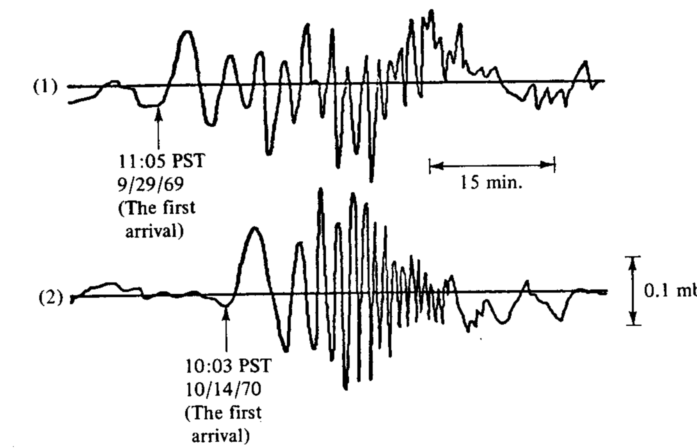

FIGURE 1-11
An example of a wave packet in which different frequencies may be seen propagating at different speeds. This example is of two air-pressure waves thought to result from nuclear explosions in Asia; they were recorded in California on one of the author's microbarographs.

real part RE and an imaginary odd part IO. Taking the squared magnitude, one has $(\text{RE} + i\text{IO})(\text{RE} - i\text{IO}) = (\text{RE})^2 + (\text{IO})^2$. The square of an even function is obviously even and the square of an odd function is also even. Thus, the spectrum of a real time function is even so that its values at plus frequencies are the same as its values at minus frequencies. In other words, there is no special meaning to be attached to negative frequencies.

Although most time functions which arise in applications are real time functions, a discussion of correlation and spectra is not mathematically complete without considering complex-valued time functions. Furthermore, complex-valued time functions can be extremely useful in many physical problems in which rotation occurs. For example consider two vector-component wind-speed indicators: one pointing north, recording n_t, and the other pointing west, recording w_t. Now if one makes up a complex-valued time series $v_t = n_t + iw_t$, the magnitude and phase angle of the complex numbers have obvious physical interpretation. The $(\text{RE}+i\text{IO})$ part of the transform relates to n_t and the $(\text{RO} + i\text{IE})$ part relates to w_t. The spectrum, however, is $(\text{RE} + \text{RO})^2 + (\text{IE} + \text{IO})^2$, which is neither even nor odd, and the fact that $V(+\omega) \neq V(-\omega)$ must have some interpretation. Indeed it does, and the meaning is that $+\omega$ corresponds to rotation in one sense (counterclockwise) and $(-\omega)$ to rotation in the other direction. To see this, suppose $n_t = \cos(\omega_0 t + \phi)$ and $w_t = -\sin(\omega_0 t + \phi)$. Then $v_t = e^{-i(\omega_0 t + \phi)}$. The transform is

$$V(\omega) = \int_{-\infty}^{+\infty} e^{-i(\omega_0 t + \phi)} e^{i\omega t}\, dt \qquad (1\text{-}5\text{-}1)$$

$$= \delta(\omega - \omega_0)e^{-i\phi} \qquad (1\text{-}5\text{-}2)$$

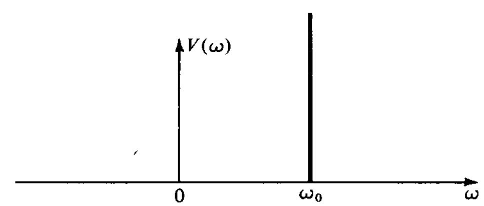

FIGURE 1-12
Spectrum of the complex time series $e^{-i(\omega_0 t+\phi)}$.

The spectrum $\delta^2(\omega - \omega_0)$ is shown in Fig. 1-12.
Conversely, if w_t were $\sin(\omega_0 t + \phi)$, then the spectrum would have been a delta function at $-\omega_0$, meaning that the wind velocity vector is rotating the other way. Other examples of complex time series in geophysics are

1 Yielding of the elastic earth to the gravitational attraction of the moon causes local ground tilt. The north-south tilt could determine an x_t time series, and the east-west tilt could determine a y_t time series. Then $x_t + iy_t$ would tend to have one rotational sense in the northern hemisphere and the opposite sense in the southern hemisphere.
2 Vertical and horizontal seismograph motions could make up a complex time series.
3 Nutation of the earth's figure axis about the angular momentum axis (Chandler Wobble).
4 Rotational polarizations of an electromagnetic wave.

Let us look at the spectrum in terms of Z transforms. Let the spectrum be $R(\omega)$, where

$$R(\omega) = |B(\omega)|^2 = \overline{B(\omega)}B(\omega) \qquad (1\text{-}5\text{-}3)$$

Let us express this in terms of a three-point Z transform:

$$R(\omega) = (\bar b_0 + \bar b_1 e^{-i\omega} + \bar b_2 e^{-i2\omega})(b_0 + b_1 e^{i\omega} + b_2 e^{i2\omega}) \qquad (1\text{-}5\text{-}4)$$

$$R(\omega) = \left(\bar b_0 + \frac{\bar b_1}{Z} + \frac{\bar b_2}{Z^2}\right)(b_0 + b_1 Z + b_2 Z^2) \qquad (1\text{-}5\text{-}5)$$

$$R(\omega) = \bar B\left(\frac{1}{Z}\right) B(Z) \qquad (1\text{-}5\text{-}6)$$

It is of interest to multiply out the polynomials $\bar B(1/Z)$ with $B(Z)$ in order to examine the coefficients of $R(Z)$.

$$R(Z) = \frac{\bar b_2 b_0}{Z^2} + \frac{(\bar b_1 b_0 + \bar b_2 b_1)}{Z} + (\bar b_0 b_0 + \bar b_1 b_1 + \bar b_2 b_2) + (\bar b_0 b_1 + \bar b_1 b_2)Z + \bar b_0 b_2 Z^2 \qquad (1\text{-}5\text{-}7)$$

$$R(Z) = \frac{r_{-2}}{Z^2} + \frac{r_{-1}}{Z} + r_0 + r_1 Z + r_2 Z^2 \qquad (1\text{-}5\text{-}8)$$

The coefficient r_k of Z^k is given by

$$r_k = \sum_i \bar b_i b_{i+k} \qquad (1\text{-}5\text{-}9)$$

Equation (1-5-9) is known as the *autocorrelation* formula. The autocorrelation value r_k at lag 10 is r_{10}. It is a measure of the similarity of b_i with itself shifted 10 units in time. In the most frequently occurring case, b_i is real; then by inspection of (1-5-7) or (1-5-9) one sees that the autocorrelation coefficients are real and $r_k = r_{-k}$. With the specialization to real time series, then, we have

$$R(Z) = r_0 + r_1\left(Z + \frac{1}{Z}\right) + r_2\left(Z^2 + \frac{1}{Z^2}\right) \qquad (1\text{-}5\text{-}10)$$

$$R(Z) = r_0 + r_1(e^{i\omega} + e^{-i\omega}) + r_2(e^{i2\omega} + e^{-i2\omega}) \qquad (1\text{-}5\text{-}11)$$

$$R(Z) = r_0 + 2r_1 \cos\omega + 2r_2 \cos 2\omega \qquad (1\text{-}5\text{-}12)$$

$$R(Z) = \sum_k r_k \cos k\,\omega \qquad (1\text{-}5\text{-}13)$$

$$R(Z) = \text{cosine transform of } r_k \qquad (1\text{-}5\text{-}14)$$

We have just shown what is a fairly difficult theorem in continuous time textbooks, namely that the cosine transform of the autocorrelation equals the magnitude squared of the Fourier transform. There are two computationally distinct methods to compute a spectrum: (1) Compute the r_k coefficients from (1-5-9) once, then form the cosine sum (1-5-13); or (2) evaluate $B(Z)$ for some value of Z on the unit circle, and multiply the resulting number by its complex conjugate. Repeat for many values of Z on the unit circle. The second method is the cheapest because the fast Fourier transform may be used.

The concept of autocorrelation and spectrum is easily generalized to cross-correlation and cross spectrum. Consider two Z transforms $A(Z)$ and $B(Z)$. Then the cross spectrum $C(Z)$ is defined by

$$C(Z) = \bar A\left(\frac{1}{Z}\right)B(Z) \qquad (1\text{-}5\text{-}15)$$

If some particular coefficient c_k in $C(Z)$ is greater than any of the others, then it may be said that the waveform a_t most resembles the waveform b_t if one is delayed k time units with respect to the other.

EXERCISES

1 Suppose a wavelet is made up of complex numbers. Is the autocorrelation relation $r_k = r_{-k}$ true? Is r_k real or complex? Is $R(\omega)$ real or complex?
2 Let x_t be some real time function. Let $y_t = x_{t+3}$ be another real time function. Sketch the phase as a function of frequency of the cross spectrum $X(1/Z)Y(Z)$ as computed by a computer which put all arctangents in the principal quadrants $-\pi/2 < \arctan < \pi/2$. Label axis scales.
3 If concepts of time and frequency are interchanged, what does the meaning of spectrum become?

1-6 HILBERT TRANSFORM

A filter which converts sines into cosines is called a 90° phase shift filter or a quadrature filter. More specifically if the input is $\cos(\omega t + \phi_1)$, then the output should be $\cos(\omega t + \phi_1 + \pi/2)$. Such a filter can be useful in constructing the envelope of a time function. Let $X(Z)$ denote the Z transform of a real data series, $Q(Z)$ denote a quadrature filter, and let $Y(Z) = Q(Z)X(Z)$ be the output of the quadrature filter. Then the envelope time function may be defined by $e_t = (x_t^2 + y_t^2)^{1/2}$. Alternatively, one could construct a complex time function $u_t = x_t + iy_t$. In terms of Z transforms we have

$$U(Z) = [1 + iQ(Z)]X(Z)$$

Now $u_t\bar{u}_t$ represents the squared envelope function. Likewise the phase ϕ_t as a function of time may be defined as $\phi_t = \arctan(y_t/x_t)$. The instantaneous frequency is $d\phi/dt$. This may be approximated in the following way.

$$\phi_t = \operatorname{Im} \ln u_t$$

$$\frac{d\phi}{dt} = \operatorname{Im} \frac{1}{u}\frac{du}{dt}$$

$$\approx \operatorname{Im} \frac{2}{\Delta t} \frac{u_t - u_{t-1}}{u_t + u_{t-1}}$$

Now that we have some idea what a 90° phase shift filter can be used for, let us find out the numerical values of q_t. The time derivative operation has the desired 90° phase-shifting property we seek. The trouble with a differentiator is that higher frequencies are amplified with respect to lower frequencies. Specifically

$$f(t) = \int F(\omega) e^{-i\omega t}\, d\omega$$

$$\frac{df}{dt} = \int -i\omega F(\omega) e^{-i\omega t}\, d\omega$$

Thus we see that time differentiation corresponds to the weight factor $-i\omega$ in the frequency domain. The weight $-i\omega$ has the proper phase but the wrong amplitude. The desired weight factor is $Q(\omega) = -i\omega/|\omega|$. It is the step function shown in Fig. 1-13.

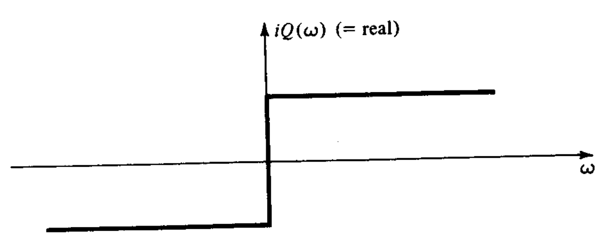

FIGURE 1-13
Frequency response of 90° phase-shifting filter.

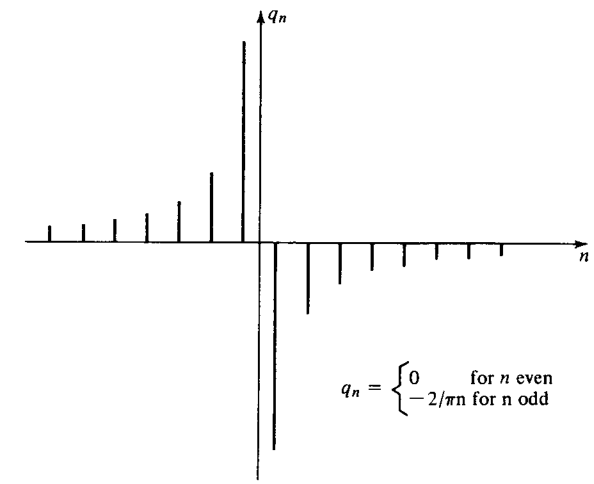

FIGURE 1-14
Quadrature filter.

Let us transform $Q(\omega)$ into the time domain

$$q_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q(\omega) e^{-i\omega n}\, d\omega$$

$$= \frac{i}{2\pi}\int_{-\pi}^{0} e^{-i\omega n}\, d\omega - \frac{i}{2\pi}\int_{0}^{\pi} e^{-i\omega n}\, d\omega$$

$$= \frac{i}{2\pi}\left(\frac{e^{-i\omega n}}{-in}\Big|_{-\pi}^{0} - \frac{e^{-i\omega n}}{-in}\Big|_{0}^{\pi}\right)$$

$$= \frac{1}{2\pi n}(-1 + e^{+in\pi} + e^{-in\pi} - 1)$$

$$= \begin{cases} 0 & \text{for } n \text{ even} \\ -2/\pi n & \text{for } n \text{ odd} \end{cases}$$

The result is shown in Fig. 1-14.

Since the filter does not vanish for negative n, this is obviously a nonrealizable filter (one which requires future inputs to create its present output). If the discussion were in continuous time rather than sampled time, the filter would be of the form $1/t$, a function which has a singularity at $t = 0$ and whose integral over $+t$ is divergent. Convolution with the filter coefficients q_n is therefore very awkward because the infinite sequence drops off very slowly. Convolution with the filter q is called *Hilbert transformation.*

Let us return to the filter $1 + iQ(Z)$ mentioned earlier. As shown in Fig. 1-15, this filter is simply a step function in the frequency domain. A cheap way to achieve the 90° phase shift operation is to do it in the frequency domain. One begins with $x_t + i \cdot 0$ and transforms it to the frequency domain. Then multiply by the step of Fig. 1-15. Finally, inverse transformation gives $x_t + iy_t$. The progress of even, odd, real, and imaginary parts is detailed in Fig. 1-16.

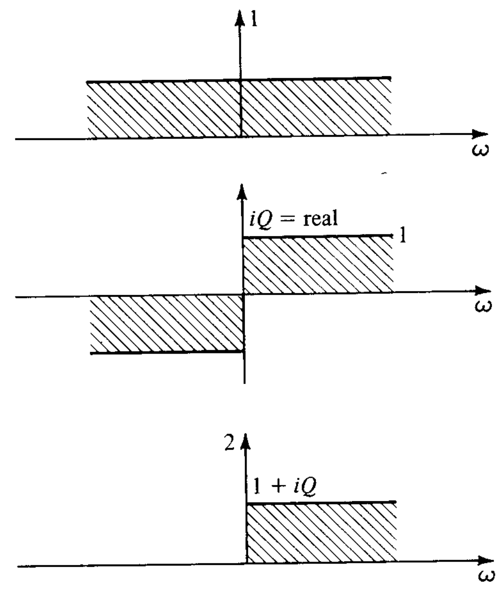

FIGURE 1-15
The filter $1 + iQ(Z)$ is real and one-sided in the frequency domain but complex and two-sided in the time domain.

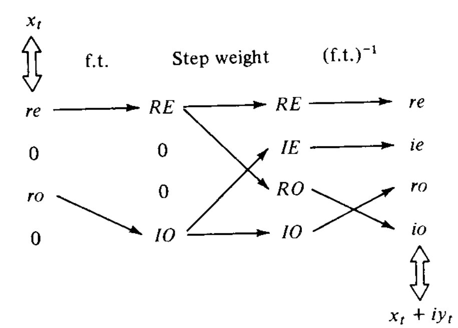

FIGURE 1-16
Hilbert transform or quadrature filtering by step weight in the frequency domain.

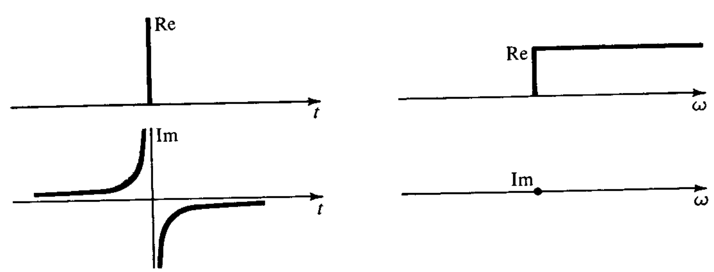

FIGURE 1-17
Impulse plus i times a 90° phase-shift filter becomes a real step in the frequency domain.

The function $1 + iQ$ plays a special role in theoretical time series analysis which, in later chapters, will be shown to be related to the principle of causality. For future reference we summarize the properties of this function in Fig. 1-17.

EXERCISES

1 By means of partial fractions convolve the waveform

$$(2/\pi)(\ldots,\ -\tfrac{1}{5}, 0, -\tfrac{1}{3}, 0, -1, 0, 1, 0, \tfrac{1}{3}, 0, \tfrac{1}{5}, \ldots)$$

with itself. What is the interpretation of the fact that the result is $(\ldots, 0, 0, -1, 0, 0, \ldots)$? (HINT: $\pi^2/8 = 1 + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \cdots$).

2 In terms of the fast Fourier transform matrix the quadrature filter $Q(\omega)$ may be represented by the column vector

$$-i(0, 1, 1, 1, \ldots, 0, -1, -1, -1, \ldots, -1)^T$$

Multiply this into the inverse transform matrix to show that the transform is proportional to $(\cos \pi k/N)/(\sin \pi k/N)$. What is the scale factor? Sketch it for $k \ll N$ indicating the limit $N \to \infty$. [HINT: $1 + x + x^2 + \cdots x^N = (1 - x^{N+1})/(1 - x)$.]