FOURIER TRANSFORMS AND WAVES:
in four long lectures

Jon F. Clærbout
Cecil and Ida Green Professor of Geophysics
Stanford University

Contents

Introduction

0.1 THEMES

These four long lectures on Fourier Transforms and waves follow two general themes,

First, instead of drilling down into analytical details of one-dimensional Fourier analysis, these lectures scan the basic definitions and concepts focusing on the concrete, namely, computation. Then we move on to the basic principles of multidimensional spectra, a field fundamental to wavefield analysis and practical imaging.

The second theme of these lectures is examples. There are 60 illustrations here, the majority of them being computed illustrations of the concepts. The theory here is mostly that required to explain the examples.

All of the illustrations here come from my earlier books

1. Fundamentals of Geophysical Data Processing (**FGDP**)
2. Earth Soundings Analysis, Processing versus Inversion (**PVI**)
3. Basic Earth Imaging (**BEI**)
4. Geophysical Estimation by Example (**GEE**)

These books are all freely available on the web at http://sepwww.stanford.edu/sep/prof/.

0.2 WHY GEOPHYSICS USES FOURIER ANALYSIS

When earth material properties are constant in any of the cartesian variables (t, x, y, z) then it is useful to Fourier transform (FT) that variable.

In seismology, the earth does not change with time (the ocean does!) so for the earth, we can generally gain by Fourier transforming the time axis thereby converting time-dependent differential equations (hard) to algebraic equations (easier) in frequency (temporal frequency).

In seismology, the earth generally changes rather strongly with depth, so we cannot usefully Fourier transform the depth z axis and we are stuck with differential equations in z. On

the other hand, we can model a layered earth where each layer has material properties that are constant in z. Then we get analytic solutions in layers and we need to patch them together.

Thirty years ago, computers were so weak that we always Fourier transformed the x and y coordinates. That meant that their analyses were limited to earth models in which velocity was horizontally layered. Today we still often Fourier transform t, x, y but not z, so we reduce the partial differential equations of physics to ordinary differential equations (ODEs). A big advantage of knowing FT theory is that it enables us to visualize physical behavior without us needing to use a computer.

The Fourier transform variables are called frequencies. For each axis (t, x, y, z) we have a corresponding frequency (ω, k_x, k_y, k_z). The k's are spatial frequencies, ω is the temporal frequency.

The frequency is inverse to the wavelength. Question: A seismic wave from the fast earth goes into the slow ocean. The temporal frequency stays the same. What happens to the spatial frequency (inverse spatial wavelength)?

In a layered earth, the horizonal spatial frequency is a constant function of depth. We will find this to be Snell's law.

In a spherical coordinate system or a cylindrical coordinate system, Fourier transforms are useless but they are closely related to "spherical harmonic functions" and Bessel transformations which play a role similar to FT.

After we develop some techniques for 2-D Fourier transform of surface seismic data, we'll see how to use wave theory to take these observations made on the earth's surface and "downward continue" them, to extrapolate them into the earth. This is a central tool in earth imaging.

Then we introduce the concepts of reflection coefficient and layered media. The main difference between wave theory found in physics books and geophysics books is due to the near omnipresent gravitational stratification of earth materials.

A succeeding chapter considers two-dimensional spectra of any function, how such functions can be modeled, what it means to deconvolve 2-D functions, and an all-purpose method of filling in missing data in a 2-D function based on its spectrum.

The final chapter returns us to one dimension for the "uncertainty principle," for the basic concepts of resolution and statistical fluctuation.

Chapter 1

Convolution and Spectra

When Fourier transforms are applicable, it means the "earth response" now is the same as the earth response later. Switching our point of view from time to space, the applicability of Fourier transformation means that the "**impulse response**" here is the same as the impulse response there. An *impulse* is a column vector full of zeros with somewhere a one, say $(0,0,1,0,0,\cdots)'$ (where the prime $()'$ means transpose the row into a column.) An *impulse response* is a column from the matrix

$$\mathbf{q} \quad = \quad \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \\ q_6 \\ q_7 \end{bmatrix} \quad = \quad \begin{bmatrix} b_0 & 0 & 0 & 0 & 0 & 0 \\ b_1 & b_0 & 0 & 0 & 0 & 0 \\ b_2 & b_1 & b_0 & 0 & 0 & 0 \\ 0 & b_2 & b_1 & b_0 & 0 & 0 \\ 0 & 0 & b_2 & b_1 & b_0 & 0 \\ 0 & 0 & 0 & b_2 & b_1 & b_0 \\ 0 & 0 & 0 & 0 & b_2 & b_1 \\ 0 & 0 & 0 & 0 & 0 & b_2 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix} \quad = \quad \mathbf{Bp} \qquad (1.1)$$

The impulse response is the **q** that comes out when the input **p** is an impulse. In a typical application, the matrix would be about 1000×1000 and not the simple 8×6 example that I show you above. Notice that each column in the matrix contains the same waveform (b_0, b_1, b_2). This waveform is called the "impulse response". The collection of impulse responses in equation (1.1) defines the the *convolution* operation.

Not only do the columns of the matrix contain the same impulse response, but each row likewise contains the same thing, and that thing is the backwards impulse response (b_2, b_1, b_0). Suppose (b_2, b_1, b_0) were numerically equal to $(1, -2, 1)/\Delta t^2$. Then equation (1.1) would be like the differential equation $\frac{d^2}{dt^2} p = q$. Equation (1.1) would be a finite-difference representation of a differential equation. Two important ideas are equivalent; either they are both true or they are both false:

1. The columns of the matrix all hold the same impulse response.

2. The differential equation has constant coefficients.

Let us take a quick peek ahead. The relationship of equation (1.1) with Fourier transforms is that the k-th row in (1.1) is the k-th power of Z in a polynomial multiplication $Q(Z) = B(Z)P(Z)$. The relationship of any polynomial such as $Q(Z)$ to Fourier Transforms results from the relation $Z = e^{i\omega\Delta t}$, as we will see.

1.1 SAMPLED DATA AND Z-TRANSFORMS

Time and space are ordinarily thought of as continuous, but for the purposes of computer analysis we must discretize these axes. This is also called "**sampling**" or "**digitizing**." You might worry that discretization is a practical evil that muddies all later theoretical analysis. Actually, physical concepts have representations that are exact in the world of discrete mathematics.

Consider the idealized and simplified signal in Figure 1.1. To analyze such an observed signal 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 or observe $b(t)$ at a uniform spacing of points in time, call this discretized signal b_t. For Figure 1.1, such a discrete approximation to the continuous function could be denoted by the vector

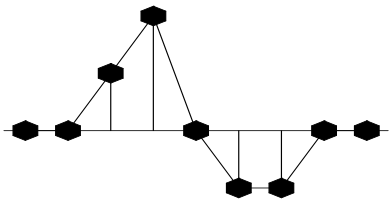

Figure 1.1: A continuous signal sampled at uniform time intervals. cs-triv1 [ER]

$$b_t \quad = \quad (\ldots\ 0, 0, 1, 2, 0, -1, -1, 0, 0, \ldots) \tag{1.2}$$

Naturally, if time points were closer together, the approximation would be more accurate. What we have done, then, is represent a signal by an abstract n-dimensional vector.

Another way to represent a signal is as a polynomial, where the *coefficients* of the polynomial represent the value of b_t at successive times. For example,

$$B(Z) \quad = \quad 1 + 2Z + 0Z^2 - Z^3 - Z^4 \tag{1.3}$$

This polynomial is called a "Z**-transform**." What is the meaning of Z here? Z should not take on some numerical value; it is instead the **unit-delay operator**. For example, the coefficients of $ZB(Z) = Z + 2Z^2 - Z^4 - Z^5$ are plotted in Figure 1.2. Figure 1.2 shows the same waveform as Figure 1.1, but now the waveform has been delayed. So the signal b_t is delayed n time units

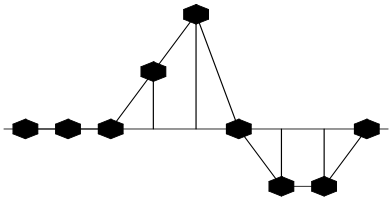

Figure 1.2: The coefficients of $ZB(Z)$ are the shifted version of the coefficients of $B(Z)$. cs-triv2 [ER]

by multiplying $B(Z)$ by Z^n. The delay operator Z is important in analyzing waves simply because waves take a certain amount of time to move from place to place.

Another value of the delay operator is that it may be used to build up more complicated signals 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 occurred at $t = 10$ time units after the first, we would expect the pressure function $y(t)$ depicted in Figure 1.3. In terms of Z-transforms, this pressure function would be expressed as $Y(Z) = B(Z) + Z^{10}B(Z)$.

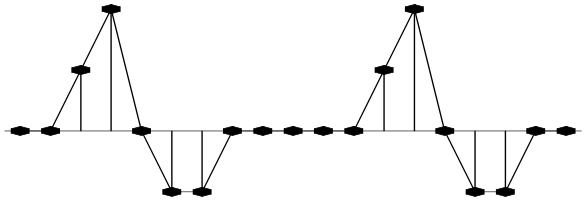

Figure 1.3: Response to two explosions. cs-triv3 [ER]

1.1.1 Linear superposition

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 overlapped one another in time (as would be the case if $B(Z)$ had degree greater than 10), the waveforms would simply add together in the region of overlap. The supposition that they would just add together without any interaction is called the "**linearity**" property. In seismology we find that—although the earth is a heterogeneous conglomeration of rocks of different shapes and types—when seismic waves 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 is a dominating feature in hydrodynamics, where flow velocities are a noticeable fraction of the wave velocity. Nonlinearity is absent from reflection seismology except within a few meters from the source. Nonlinearity does not arise from geometrical complications in the propagation path. An example of two **plane wave**s superposing is shown in Figure 1.4.

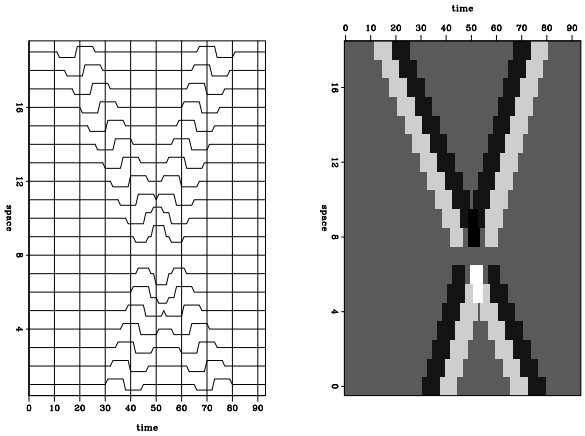

Figure 1.4: Crossing plane waves superposing viewed on the left as "wiggle traces" and on the right as "raster." cs-super [ER]

1.1.2 Convolution with Z-transform

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

$$\begin{aligned} Y(Z) &= B(Z) - \frac{Z}{2}B(Z) + \frac{Z^3}{4}B(Z) \\ &= \left(1 - \frac{Z}{2} + \frac{Z^3}{4}\right) B(Z) \\ &= X(Z)\,B(Z) \end{aligned} \tag{1.4}$$

The last equation shows **polynomial multiplication** as the underlying basis of time-invariant linear-system theory, namely that the output $Y(Z)$ can be expressed as the input $X(Z)$ times the impulse-response **filter** $B(Z)$. When signal values are insignificant except in a "small" region on the time axis, the signals are called "**wavelet**s."

1.1.3 Convolution equation and program

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 $(1, -1)$ would be 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

$$X(Z)\,B(Z) = Y(Z) \tag{1.5}$$

$$(x_0 + x_1 Z + x_2 Z^2 + \cdots)(b_0 + b_1 Z + b_2 Z^2) = y_0 + y_1 Z + y_2 Z^2 + \cdots \tag{1.6}$$

Identifying coefficients of successive powers of Z, we get

$$\begin{aligned} y_0 &= x_0 b_0 \\ y_1 &= x_1 b_0 + x_0 b_1 \\ y_2 &= x_2 b_0 + x_1 b_1 + x_0 b_2 \\ y_3 &= x_3 b_0 + x_2 b_1 + x_1 b_2 \\ y_4 &= x_4 b_0 + x_3 b_1 + x_2 b_2 \\ &= \cdots\cdots\cdots\cdots \end{aligned} \tag{1.7}$$

In matrix form this looks like

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} x_0 & 0 & 0 \\ x_1 & x_0 & 0 \\ x_2 & x_1 & x_0 \\ x_3 & x_2 & x_1 \\ x_4 & x_3 & x_2 \\ 0 & x_4 & x_3 \\ 0 & 0 & x_4 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \tag{1.8}$$

The following equation, called the "convolution equation," carries the spirit of the group shown in (1.7):

$$y_k = \sum_{i=0}^{N_b} x_{k-i} b_i \tag{1.9}$$

To be correct in detail when we associate equation (1.9) with the group (1.7), we should also assert that either the input x_k vanishes before $k = 0$ or N_b must be adjusted so that the sum does not extend before x_0. These end conditions are expressed more conveniently by defining $j = k - i$ in equation (1.9) and eliminating k getting

$$y_{j+i} = \sum_{i=0}^{N_b} x_j b_i \tag{1.10}$$

A convolution program based on equation (1.10) including end effects on both ends, is `convolve()`.

```
#       convolution:    Y(Z) = X(Z) * B(Z)
#
subroutine convolve( nb, bb, nx, xx, yy )
integer nb      # number of coefficients in filter
integer nx      # number of coefficients in input
                # number of coefficients in output will be nx+nb-1
real    bb(nb)  # filter coefficients
real    xx(nx)  # input trace
real    yy(1)   # output trace
integer ib, ix, iy, ny
ny = nx + nb -1
call null( yy, ny)
do ib= 1, nb
        do ix= 1, nx
                yy( ix+ib-1) = yy( ix+ib-1) + xx(ix) * bb(ib)
return; end
```

This program is written in a language called Ratfor, a "rational" dialect of Fortran. It is similar to the Matlab language. You are not responsible for anything in this program, but, if you are interested, more details in the last chapter of PVI[1], the book that I condensed this from.

[1] http://sepwww.stanford.edu/sep/prof/

1.1.4 Negative time

Notice that $X(Z)$ and $Y(Z)$ need not strictly be polynomials; they may contain both positive and negative powers of Z, such as

$$X(Z) = \cdots + \frac{x_{-2}}{Z^2} + \frac{x_{-1}}{Z} + x_0 + x_1 Z + \cdots \tag{1.11}$$

$$Y(Z) = \cdots + \frac{y_{-2}}{Z^2} + \frac{y_{-1}}{Z} + y_0 + y_1 Z + \cdots \tag{1.12}$$

The negative powers of Z in $X(Z)$ and $Y(Z)$ show that the *data* is defined before $t = 0$. The effect of using negative powers of Z in the *filter* is different. Inspection of (1.9) shows that the output y_k that occurs at time k is a linear combination of current and previous inputs; that is, $(x_i, i \le k)$. 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 that 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 that has not yet occurred. However, nonrealizable filters are occasionally useful in computer simulations where all the data is prerecorded.

1.2 FOURIER SUMS

The world is filled with sines and cosines. The coordinates of a point on a spinning wheel are $(x, y) = (\cos(\omega t + \phi), \sin(\omega t + \phi))$, where ω is the angular frequency of revolution and ϕ is the phase angle. The purest tones and the purest colors are sinusoidal. The movement of a pendulum is nearly sinusoidal, the approximation going to perfection in the limit of small amplitude motions. The sum of all the tones in any signal is its "spectrum."

Small amplitude signals are widespread in nature, from the vibrations of atoms to the sound vibrations we create and observe in the earth. Sound typically compresses air by a volume fraction of 10^{-3} to 10^{-6}. In water or solid, the compression is typically 10^{-6} to 10^{-9}. A mathematical reason why sinusoids are so common in nature is that laws of nature are typically expressible as partial differential equations. Whenever the coefficients of the differentials (which are functions of material properties) are constant in time and space, the equations have exponential and sinusoidal solutions that correspond to waves propagating in all directions.

1.2.1 Superposition of sinusoids

Fourier analysis is built from the complex exponential

$$e^{-i\omega t} = \cos\omega t - i \sin\omega t \tag{1.13}$$

A Fourier component of a time signal is a complex number, a sum of real and imaginary parts, say

$$B = \mathrm{Re}\, B + i\, \mathrm{Im}\, B \tag{1.14}$$

which is attached to some frequency. Let j be an integer and ω_j be a set of frequencies. A signal $b(t)$ can be manufactured by adding a collection of complex exponential signals, each complex exponential being scaled by a complex coefficient B_j, namely,

$$b(t) \quad = \quad \sum_j B_j \; e^{-i\omega_j t} \tag{1.15}$$

This manufactures a **complex-valued signal**. How do we arrange for $b(t)$ to be real? We can throw away the imaginary part, which is like adding $b(t)$ to its complex conjugate $\overline{b}(t)$, and then dividing by two:

$$\text{Re } b(t) \quad = \quad \frac{1}{2}\sum_j (B_j \; e^{-i\omega_j t} + \bar{B_j} \; e^{i\omega_j t}) \tag{1.16}$$

In other words, for each positive ω_j with amplitude B_j, we add a negative $-\omega_j$ with amplitude $\bar{B_j}$ (likewise, for every negative ω_j ...). The B_j are called the "frequency function," or the "Fourier transform." Loosely, the B_j are called the "**spectrum**," though in formal mathematics, the word "spectrum" is reserved for the product $\bar{B_j} B_j$. The words "**amplitude spectrum**" universally mean $\sqrt{\bar{B_j} B_j}$.

In practice, the collection of frequencies is almost always evenly spaced. Let j be an integer $\omega = j \; \Delta\omega$ so that

$$b(t) \quad = \quad \sum_j B_j \, e^{-i(j \; \Delta\omega)t} \tag{1.17}$$

Representing a signal by a sum of sinusoids is technically known as "inverse Fourier transformation." An example of this is shown in Figure 1.5.

1.2.2 Sampled time and Nyquist frequency

In the world of computers, time is generally mapped into integers too, say $t = n\Delta t$. This is called "discretizing" or "sampling." The highest possible frequency expressible on a **mesh** is $(\cdots, 1, -1, +1, -1, +1, -1, \cdots)$, which is the same as $e^{i\pi n}$. Setting $e^{i\omega_{\max} t} = e^{i\pi n}$, we see that the maximum frequency is

$$\omega_{\max} \quad = \quad \frac{\pi}{\Delta t} \tag{1.18}$$

Time is commonly given in either seconds or sample units, which are the same when $\Delta t = 1$. In applications, frequency is usually expressed in cycles per second, which is the same as **Hertz**, abbreviated **Hz**. In computer work, frequency is usually specified in cycles per sample. In theoretical work, frequency is usually expressed in **radian**s where the relation between radians and cycles is $\omega = 2\pi f$. We use radians because, otherwise, equations are filled with 2π's. When time is given in sample units, the maximum frequency has a name: it is the "**Nyquist frequency**," which is π radians or $1/2$ cycle per sample.

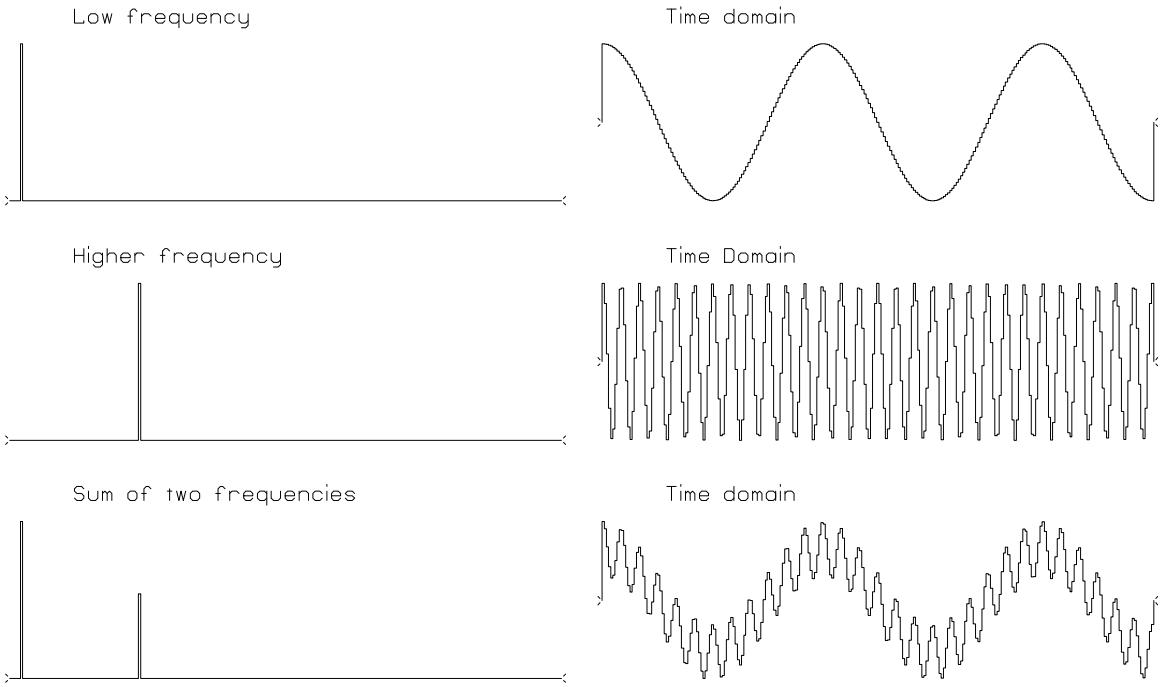

Figure 1.5: Superposition of two sinusoids. cs-cosines [NR]

1.2.3 Fourier sum

In the previous section we superposed uniformly spaced frequencies. Now we will **superpose** delayed impulses. The frequency function of a delayed impulse at time delay t_0 is $e^{i\omega t_0}$. Adding some pulses yields the "**Fourier sum**":

$$B(\omega) \quad = \quad \sum_n b_n \, e^{i\omega t_n} \quad = \quad \sum_n b_n \, e^{i\omega n \Delta t} \tag{1.19}$$

The Fourier sum transforms the signal b_t to the frequency function $B(\omega)$. Time will often be denoted by t, even though its units are sample units instead of physical units. Thus we often see b_t in equations like (1.19) instead of b_n, resulting in an implied $\Delta t = 1$.

1.3 FOURIER AND Z-TRANSFORM

The frequency function of a pulse at time $t_n = n\Delta t$ is $e^{i\omega n \Delta t} = (e^{i\omega \Delta t})^n$. The factor $e^{i\omega \Delta t}$ occurs so often in applied work that it has a name:

$$Z \quad = \quad e^{i\omega \Delta t} \tag{1.20}$$

With this Z, the pulse at time t_n is compactly represented as Z^n. The variable Z makes **Fourier transform**s look like polynomials, the subject of a literature called "Z**-transform**s." The Z-transform is a variant form of the Fourier transform that is particularly useful for time-discretized (sampled) functions.

From the definition (1.20), we have $Z^2 = e^{i\omega 2\Delta t}$, $Z^3 = e^{i\omega 3\Delta t}$, etc. Using these equivalencies, equation (1.19) becomes

$$B(\omega) \quad = \quad B(\omega(Z)) \quad = \quad \sum_n b_n \, Z^n \tag{1.21}$$

1.3.1 Unit circle

In this chapter, ω is a real variable, so $Z = e^{i\omega\Delta t} = \cos\omega\Delta t + i\sin\omega\Delta t$ is a complex variable. It has unit magnitude because $\sin^2 + \cos^2 = 1$. As ω ranges on the real axis, Z ranges on the unit circle $|Z| = 1$.

1.3.2 Differentiator

Calculus defines the differential of a time function like this

$$\frac{d}{dt} \, f(t) \quad = \quad \lim_{\Delta t \to 0} \frac{f(t) \, - \, f(t - \Delta t)}{\Delta t}$$

Computationally, we think of a differential as a finite difference, namely, a function is delayed a bit and then subtracted from its original self. Expressed as a Z-transform, the finite difference operator is $(1 - Z)$ with an implicit $\Delta t = 1$. In the language of filters, the time derivative is the filter $(+1, -1)$.

The filter $(1 - Z)$ is often simply called a "**differentiator**." It is displayed in Figure 1.6. Notice its amplitude spectrum increases with frequency. Theoretically, the amplitude spectrum

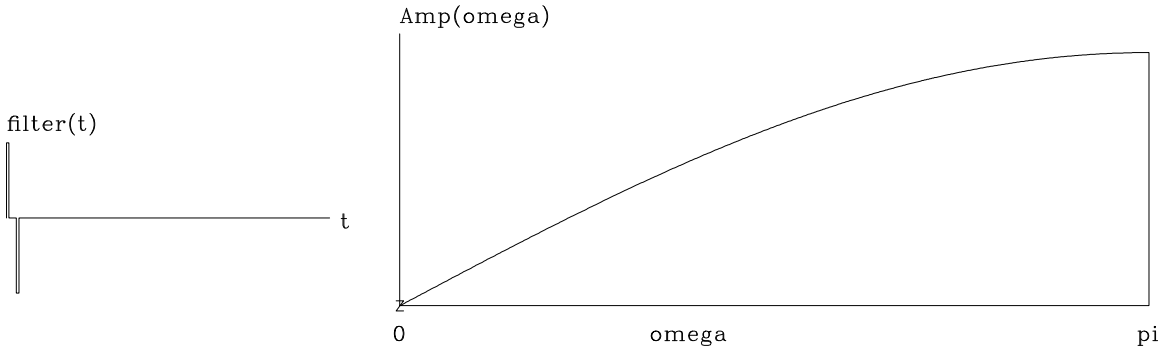

Figure 1.6: A discrete representation of the first-derivative operator. The filter $(1, -1)$ is plotted on the left, and on the right is an amplitude response, i.e., $|1 - Z|$ versus ω. cs-ddt [NR]

of a time derivative operator increases linearly with frequency. Here is why. Begin from Fourier representation of a time function (1.15).

$$b(t) \quad = \quad \sum_j B_j \, e^{-i\omega_j t} \tag{1.22}$$

$$\frac{d}{dt} \, b(t) \quad = \quad \sum_j -i\,\omega_j \, B_j \, e^{-i\omega_j t} \tag{1.23}$$

and notice that where the original function has Fourier coefficients B_j, the time derivative has Fourier coefficients $-i\omega B_j$.

In Figure 1.6 we notice the spectrum begins looking like a linear function of ω, but at higher frequencies, it curves. This is because at high frequencies, a finite difference is different from a differential.

1.3.3 Gaussian examples

The filter $(1+Z)/2$ is a running average of two adjacent time points. Applying this filter N times yields the filter $(1+Z)^N/2^N$. The coefficients of the filter $(1+Z)^N$ are generally known as **Pascal's triangle**. For large N the coefficients tend to a mathematical limit known as a **Gaussian** function, $\exp(-\alpha(t-t_0)^2)$, where α and t_0 are constants that we will not determine here. We will not prove it here, but this Gaussian-shaped signal has a Fourier transform that also has a Gaussian shape, $\exp(-\beta\omega^2)$. The Gaussian shape is often called a "bell shape." Figure 1.7 shows an example for $N \approx 15$. Note that, except for the rounded ends, the bell shape seems a good fit to a triangle function. Curiously, the filter $(.75+.25Z)^N$ also tends

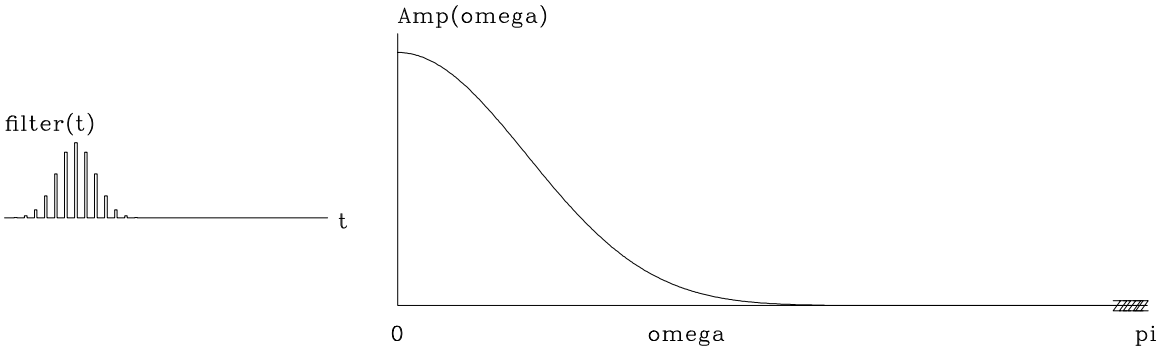

Figure 1.7: A Gaussian approximated by many powers of $(1+Z)$. cs-gauss [NR]

to the same Gaussian but with a different t_0. A mathematical theorem says that almost any polynomial raised to the N-th power yields a Gaussian.

In seismology we generally fail to observe the **zero frequency**. Thus the idealized seismic pulse cannot be a Gaussian. An analytic waveform of longstanding popularity in seismology is the second derivative of a Gaussian, also known as a "**Ricker wavelet**." Starting from the Gaussian and multiplying by $(1-Z)^2 = 1-2Z+Z^2$ produces this old, favorite wavelet, shown in Figure 1.8.

1.4 FT AS AN INVERTIBLE MATRIX

Happily, Fourier sums are exactly invertible: given the output, the input can be quickly found. Because signals can be transformed to the frequency domain, manipulated there, and then

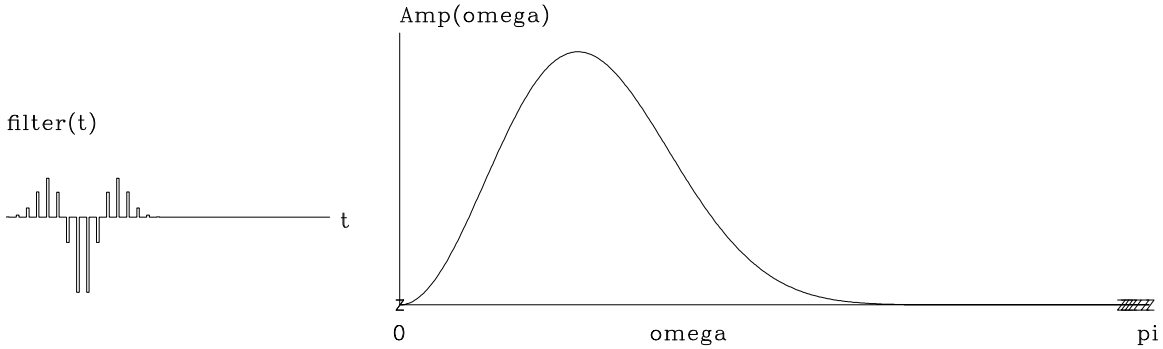

Figure 1.8: Ricker wavelet. cs-ricker [NR]

returned to the time domain, convolution and correlation can be done faster. Time derivatives can also be computed with more accuracy in the frequency domain than in the time domain. Signals can be shifted a fraction of the time sample, and they can be shifted back again exactly. In this chapter we will see how many operations we associate with the time domain can often be done better in the frequency domain. We will also examine some two-dimensional Fourier transforms.

A **Fourier sum** may be written

$$B(\omega) \quad = \quad \sum_t b_t \, e^{i\omega t} \quad = \quad \sum_t b_t \, Z^t \tag{1.24}$$

where the complex value Z is related to the real frequency ω by $Z = e^{i\omega}$. This Fourier sum is a way of building a continuous function of ω from discrete signal values b_t in the time domain. Here we specify both time and frequency domains by a set of points. Begin with an example of a signal that is nonzero at four successive instants, (b_0, b_1, b_2, b_3). The transform is

$$B(\omega) \quad = \quad b_0 + b_1 Z + b_2 Z^2 + b_3 Z^3 \tag{1.25}$$

The evaluation of this polynomial can be organized as a matrix times a vector, such as

$$\begin{bmatrix} B_0 \\ B_1 \\ B_2 \\ B_3 \end{bmatrix} \quad = \quad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & W & W^2 & W^3 \\ 1 & W^2 & W^4 & W^6 \\ 1 & W^3 & W^6 & W^9 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \tag{1.26}$$

Observe that the top row of the matrix evaluates the polynomial at $Z = 1$, a point where also $\omega = 0$. The second row evaluates $B_1 = B(Z = W = e^{i\omega_0})$, where ω_0 is some base frequency. The third row evaluates the Fourier transform for $2\omega_0$, and the bottom row for $3\omega_0$. The matrix could have more than four rows for more frequencies and more columns for more time points. I have made the matrix square in order to show you next how we can find the inverse matrix. The size of the matrix in (1.26) is $N = 4$. If we choose the base frequency ω_0 and hence W

correctly, the inverse matrix will be

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = 1/N \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1/W & 1/W^2 & 1/W^3 \\ 1 & 1/W^2 & 1/W^4 & 1/W^6 \\ 1 & 1/W^3 & 1/W^6 & 1/W^9 \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ B_2 \\ B_3 \end{bmatrix} \tag{1.27}$$

Multiplying the matrix of (1.27) with that of (1.26), we first see that the diagonals are +1 as desired. To have the off diagonals vanish, we need various sums, such as $1+W+W^2+W^3$ and $1+W^2+W^4+W^6$, to vanish. Every element (W^6, for example, or $1/W^9$) is a unit vector in the complex plane. In order for the sums of the unit vectors to vanish, we must ensure that the vectors pull symmetrically away from the origin. A uniform distribution of directions meets this requirement. In other words, W should be the N-th root of unity, i.e.,

$$W = \sqrt[N]{1} = e^{2\pi i/N} \tag{1.28}$$

The lowest frequency is zero, corresponding to the top row of (1.26). The next-to-the-lowest frequency we find by setting W in (1.28) to $Z = e^{i\omega_0}$. So $\omega_0 = 2\pi/N$; and for (1.27) to be inverse to (1.26), the frequencies required are

$$\omega_k = \frac{(0,1,2,\ldots,N-1)\,2\pi}{N} \tag{1.29}$$

1.4.1 The Nyquist frequency

The highest frequency in equation (1.29), $\omega = 2\pi(N-1)/N$, is almost 2π. This frequency is twice as high as the Nyquist frequency $\omega = \pi$. The **Nyquist frequency** is normally thought of as the "highest possible" frequency, because $e^{i\pi t}$, for integer t, plots as $(\cdots,1,-1,1,-1,1,-1,\cdots)$. The double Nyquist frequency function, $e^{i2\pi t}$, for integer t, plots as $(\cdots,1,1,1,1,1,\cdots)$. So this frequency above the highest frequency is really zero frequency! We need to recall that $B(\omega) = B(\omega - 2\pi)$. Thus, all the frequencies near the upper end of the range equation (1.29) are really small negative frequencies. Negative frequencies on the interval $(-\pi, 0)$ were moved to interval $(\pi, 2\pi)$ by the matrix form of Fourier summation.

A picture of the Fourier transform matrix of equation (1.26) is shown in Figure 1.9. Notice the Nyquist frequency is the center row and center column of each matrix.

1.4.2 Convolution in one domain...

Z-transforms taught us that a multiplication of polynomials is a convolution of their coefficients. Interpreting $Z = e^{i\omega}$ as having numerical values for real numerical values of ω, we have the idea that

> Convolution in the time domain is multiplication in the frequency domain.

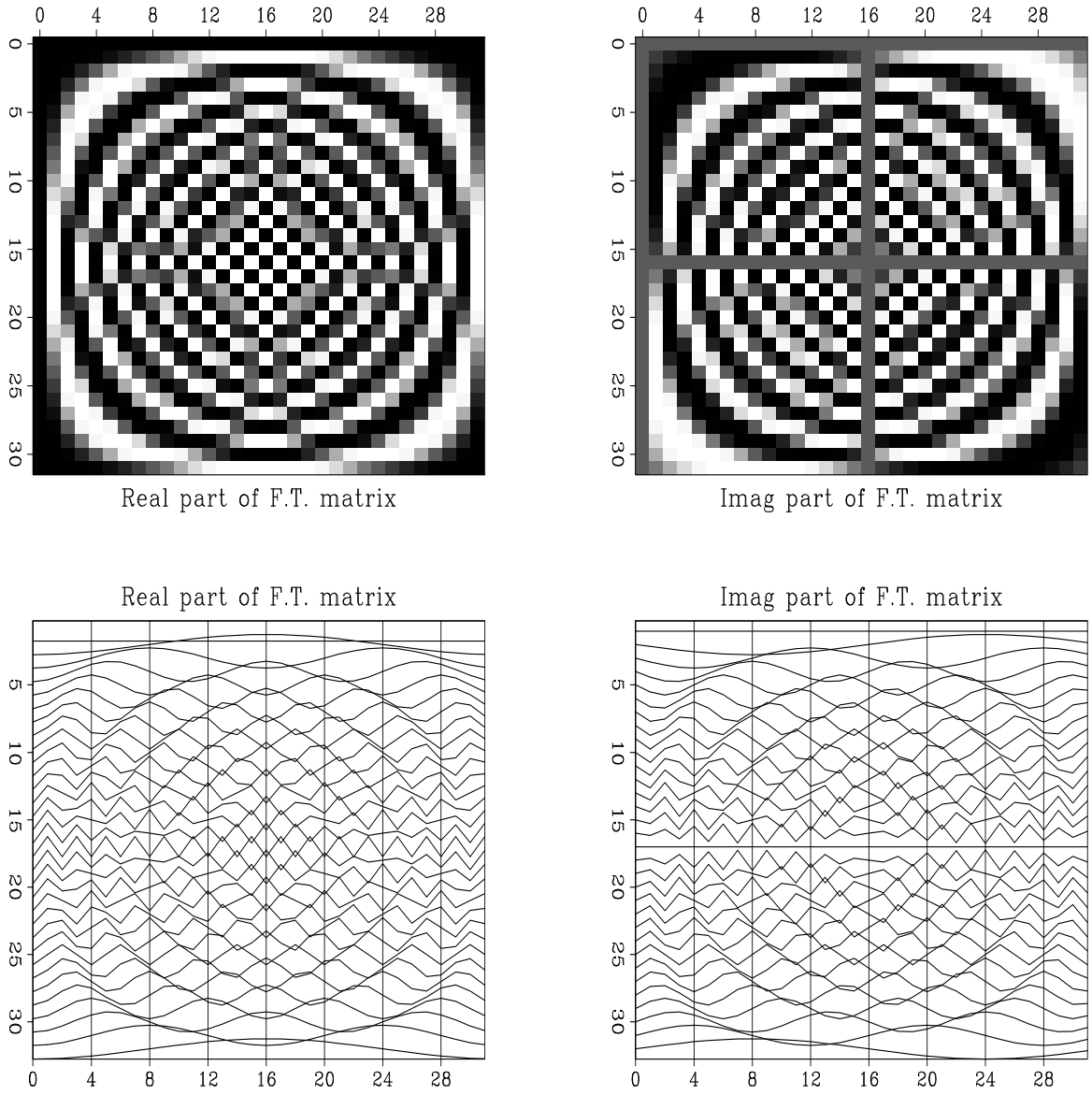

Figure 1.9: Two different graphical means of showing the real and imaginary parts of the Fourier transform matrix of size 32×32. cs-matrix [ER]

Expressing Fourier transformation as a matrix, we see that except for a choice of sign, Fourier transform is essentially the same process as as inverse fourier transform. This creates an underlying symmetry between the time domain and the frequency domain. We therefore deduce the symmetrical principle that

Convolution in the frequency domain is multiplication in the time domain.

1.4.3 Inverse Z-transform

Fourier analysis is widely used in mathematics, physics, and engineering as a **Fourier integral** transformation pair:

$$B(\omega) = \int_{-\infty}^{+\infty} b(t)\, e^{i\omega t}\, dt \tag{1.30}$$

$$\bar{b}(t) = \int_{-\infty}^{+\infty} B(\omega)\, e^{-i\omega t}\, d\omega \tag{1.31}$$

These integrals correspond to the sums we are working with here except for some minor details. Books in electrical engineering redefine $e^{i\omega t}$ as $e^{-i\omega t}$. That is like switching ω to $-\omega$. Instead, we have chosen the **sign convention** of physics, which is better for wave-propagation studies (as explained in IEI). The infinite limits on the integrals result from expressing the **Nyquist frequency** in **radian**s/second as $\pi/\Delta t$. Thus, as Δt tends to zero, the **Fourier sum** tends to the integral. It can be shown that if a scaling divisor of 2π is introduced into either (1.30) or (1.31), then $b(t)$ will equal $\bar{b}(t)$.

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 As ω moves from zero to positive frequencies, where is Z and which way does it rotate around the unit circle, clockwise or counterclockwise?

3 Identify locations on the unit circle of the following frequencies: (1) the zero frequency, (2) the Nyquist frequency, (3) negative frequencies, and (4) a frequency sampled at 10 points per wavelength.

4 Sketch the amplitude spectrum of Figure 1.8 from 0 to 4π.

1.5 SYMMETRIES

Next we examine odd/even **symmetries** to see how they are affected in Fourier transform. The even part e_t of a signal b_t is defined as

$$e_t = \frac{b_t + b_{-t}}{2} \tag{1.32}$$

The odd part is

$$o_t = \frac{b_t - b_{-t}}{2} \tag{1.33}$$

By adding (1.32) and (1.33), we see that a function is the sum of its even and odd parts:

$$b_t = e_t + o_t \tag{1.34}$$

Consider a simple, real, even signal such as $(b_{-1}, b_0, b_1) = (1, 0, 1)$. Its transform $Z + 1/Z = e^{i\omega} + e^{-i\omega} = 2\cos\omega$ is an even function of ω, since $\cos\omega = \cos(-\omega)$.

Consider the real, odd signal $(b_{-1}, b_0, b_1) = (-1, 0, 1)$. Its transform $Z - 1/Z = 2i\sin\omega$ 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 $i\,2\cos\omega$. Finally, 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 to even and odd, and lower-case and upper-case letters to time and frequency functions. A summary of the symmetries of Fourier transform is shown in Figure 1.10.

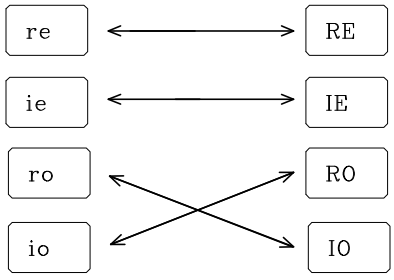

Figure 1.10: Odd functions swap real and imaginary. Even functions do not get mixed up with complex numbers. cs-reRE [NR]

More elaborate signals can be made by adding together the three-point functions we have considered. Since sums of even functions are even, and so on, the diagram in Figure 1.10 applies to all signals. An arbitrary signal is made from these four parts only, i.e., the function has the form $b_t = (\mathrm{re} + \mathrm{ro})_t + i(\mathrm{ie} + \mathrm{io})_t$. On transformation of b_t, each of the four individual parts transforms according to the table.

Most "industry standard" methods of Fourier transform set the zero frequency as the first element in the vector array holding the transformed signal, as implied by equation (1.26). This is a little inconvenient, as we saw a few pages back. The Nyquist frequency is then the first point past the middle of the even-length array, and the negative frequencies lie beyond. Figure 1.11 shows an example of an **even function** as it is customarily stored.

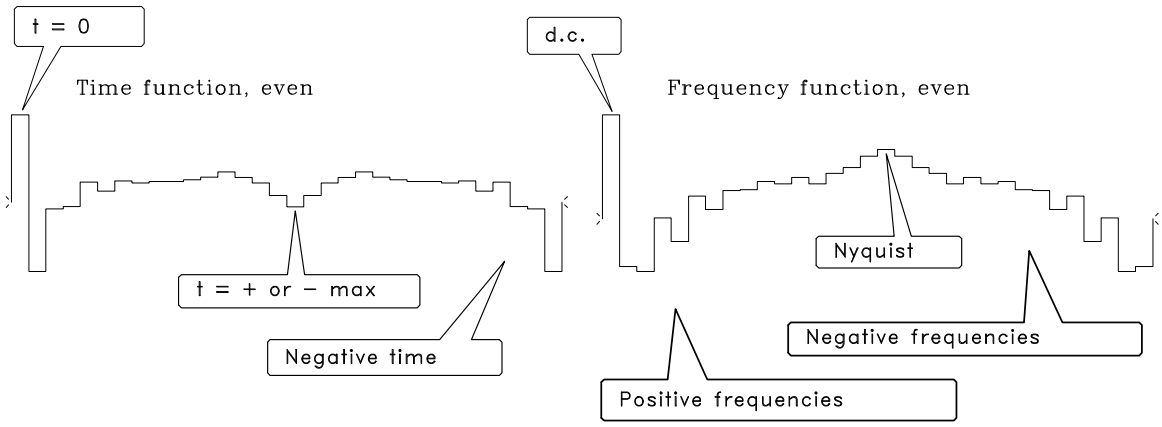

Figure 1.11: Even functions as customarily stored by "industry standard" FT programs. cs-even [NR]

EXERCISES:

1 You can learn about Fourier transformation by studying math equations in a book. You can also learn by experimenting with a program where you can "draw" a function in the time domain (or frequency domain) and instantly see the result in the other domain. The best way to learn about FT is to combine the above two methods. Spend at least an hour at the web site:

`http://sepwww.stanford.edu/oldsep/hale/FftLab.html`

which contains the interactive Java FT laboratory written by Dave Hale. This program applies equations (1.26) and (1.27).

1.5.1 Laying out a mesh

In theoretical work and in programs, the definition $Z = e^{i\omega\Delta t}$ is often simplified to $\Delta t = 1$, leaving us with $Z = e^{i\omega}$. How do we know whether ω is given in radians per second or radians per sample? We may not invoke a cosine or an exponential unless the argument has no physical dimensions. So where we see ω without Δt, we know it is in units of radians per sample.

In practical work, frequency is typically given in cycles or **Hertz**, f, rather than radians, ω (where $\omega = 2\pi f$). Here we will now switch to f. We will design a computer **mesh** on a physical object (such as a waveform or a function of space). We often take the mesh to begin at $t = 0$, and continue till the end $t_{\max}$ of the object, so the time range $t_{\text{range}} = t_{\max}$. Then we decide how many points we want to use. This will be the N used in the discrete Fourier-transform program. Dividing the range by the number gives a mesh interval Δt.

Now let us see what this choice implies in the frequency domain. We customarily take the maximum frequency to be the Nyquist, either $f_{\max} = .5/\Delta t$ Hz or $\omega_{\max} = \pi/\Delta t$ radians/sec. The frequency range f_{range} goes from $-.5/\Delta t$ to $.5/\Delta t$. In summary:

- $\Delta t \quad = \quad t_{\text{range}}/N$ is time **resolution**.
- $f_{\text{range}} \quad = \quad 1/\Delta t \quad = \quad N/t_{\text{range}}$ is frequency range.
- $\Delta f \quad = \quad f_{\text{range}}/N \quad = \quad 1/t_{\text{range}}$ is frequency **resolution**.

In principle, we can always increase N to refine the calculation. For an experiment of a fixed size t_{range}, notice that increasing N sharpens the time resolution (makes Δt smaller) but does not sharpen the frequency resolution Δf, which remains fixed. Increasing N increases the frequency *range*, but not the frequency *resolution.*

What if we want to increase the frequency resolution? Then we need to choose t_{range} larger than required to cover our object of interest. Thus we either record data over a larger range, or we assert that such measurements would be zero. Three equations summarize the facts:

$$\Delta t \; f_{\text{range}} \quad = \quad 1 \tag{1.35}$$

$$\Delta f \; t_{\text{range}} \quad = \quad 1 \tag{1.36}$$

$$\Delta f \; \Delta t \quad = \quad \frac{1}{N} \tag{1.37}$$

> Increasing *range* in the time domain increases *resolution* in the frequency domain and vice versa. Increasing **resolution** in one domain does not increase **resolution** in the other.

1.6 CORRELATION AND SPECTRA

The spectrum of a signal is a positive function of frequency that says how much of each tone is present. The Fourier transform of a spectrum yields an interesting function called an "**autocorrelation**," which measures the similarity of a signal to itself shifted.

1.6.1 Spectra in terms of Z-transforms

Let us look at spectra in terms of Z-transforms. Let a **spectrum** be denoted $S(\omega)$, where

$$S(\omega) \quad = \quad |B(\omega)|^2 \quad = \quad \overline{B(\omega)}B(\omega) \tag{1.38}$$

Expressing this in terms of a three-point Z-transform, we have

$$S(\omega) \quad = \quad (\bar{b_0} + \bar{b_1}e^{-i\omega} + \bar{b_2}e^{-i2\omega})(b_0 + b_1 e^{i\omega} + b_2 e^{i2\omega}) \tag{1.39}$$

$$S(Z) \quad = \quad \left(\bar{b_0} + \frac{\bar{b_1}}{Z} + \frac{\bar{b_2}}{Z^2}\right)(b_0 + b_1 Z + b_2 Z^2) \tag{1.40}$$

$$S(Z) \quad = \quad \overline{B}\left(\frac{1}{Z}\right) B(Z) \tag{1.41}$$

It is interesting to multiply out the polynomial $\bar{B}(1/Z)$ with $B(Z)$ in order to examine the coefficients of $S(Z)$:

$$\begin{aligned} S(Z) &= \frac{\bar{b}_2 b_0}{Z^2} + \frac{(\bar{b}_1 b_0 + \bar{b}_2 b_1)}{Z} + (\bar{b}_0 b_0 + \bar{b}_1 b_1 + \bar{b}_2 b_2) + (\bar{b}_0 b_1 + \bar{b}_1 b_2) Z + \bar{b}_0 b_2 Z^2 \\ S(Z) &= \frac{s_{-2}}{Z^2} + \frac{s_{-1}}{Z} + s_0 + s_1 Z + s_2 Z^2 \end{aligned} \tag{1.42}$$

The coefficient s_k of Z^k is given by

$$s_k = \sum_i \bar{b}_i b_{i+k} \tag{1.43}$$

Equation (1.43) is the **autocorrelation** formula. The autocorrelation value s_k at lag 10 is s_{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 (1.43), we see that the autocorrelation coefficients are real, and $s_k = s_{-k}$.

Specializing to a real time series gives

$$S(Z) = s_0 + s_1 \left(Z + \frac{1}{Z} \right) + s_2 \left(Z^2 + \frac{1}{Z^2} \right) \tag{1.44}$$

$$S(Z(\omega)) = s_0 + s_1 (e^{i\omega} + e^{-i\omega}) + s_2 (e^{i2\omega} + e^{-i2\omega}) \tag{1.45}$$

$$S(\omega) = s_0 + 2s_1 \cos\omega + 2s_2 \cos 2\omega \tag{1.46}$$

$$S(\omega) = \sum_k s_k \cos k\omega \tag{1.47}$$

$$S(\omega) = \text{cosine transform of } \; s_k \tag{1.48}$$

This proves a classic theorem that for real-valued signals can be simply stated as follows:

> For any real signal, the cosine transform of the **autocorrelation** equals the magnitude squared of the Fourier transform.

1.6.2 Two ways to compute a spectrum

There are two computationally distinct methods by which we can compute a spectrum: (1) compute all the s_k coefficients from (1.43) and then form the cosine sum (1.47) for each ω; and alternately, (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. When there are more than about twenty lags, method (2) is cheaper, because the fast Fourier transform discussed in chapter 2 can be used.

1.6.3 The comb function

Consider a constant function of time. In the frequency domain, it is an impulse at zero frequency. The **comb** function is defined to be zero at alternate time points. Multiply this constant

function by the comb function. The resulting signal contains equal amounts of two frequencies; half is zero frequency, and half is Nyquist frequency. We see this in the second row in Figure 1.12, where the Nyquist energy is in the middle of the frequency axis. In the third row, 3 out of 4 points are zeroed by another comb. We now see something like a new Nyquist frequency at half the Nyquist frequency visible on the second row.

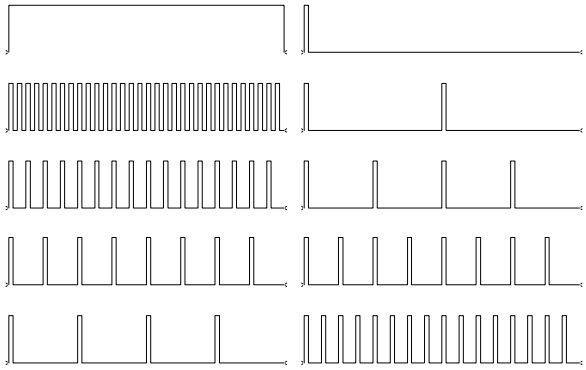

Figure 1.12: A zero-frequency function and its cosine transform. Successive rows show increasingly sparse sampling of the zero-frequency function. cs-comb [NR]

1.6.4 Undersampled field data

Figure 1.13 shows a recording of an **airgun** along with its spectrum. The original data is

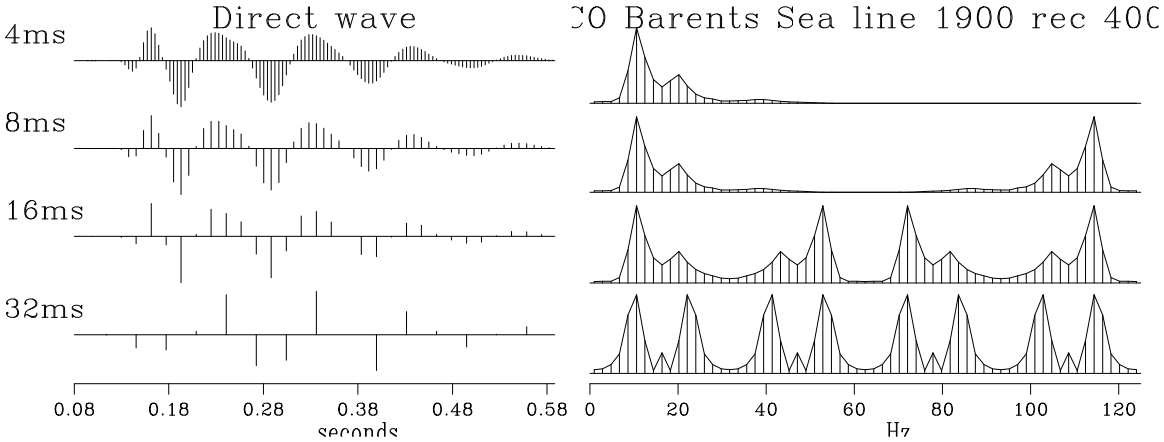

Figure 1.13: Raw data is shown on the top left, of about a half-second duration. Right shows amplitude spectra (magnitude of FT). In successive rows the data is sampled less densely. cs-undersample [ER]

sampled at an interval of 4 milliseconds, which is 250 times per second. Thus, the **Nyquist frequency** $1/(2\Delta t)$ is 125 Hz. Negative frequencies are not shown, since the amplitude spectrum at negative frequency is identical with that at positive frequency. Think of extending the top row of spectra in Figure 1.13 to range from minus 125 Hz to plus 125 Hz. Imagine the even function of frequency centered at zero frequency—we will soon see it. In the second row of the plot, I decimated the data to 8 ms. This drops the Nyquist frequency to 62.5 Hz. Energy that was at -10 Hz appears at $125-10$ Hz in the *second* row spectrum. The appearance of

what were formerly small negative frequencies near the Nyquist frequency is called "**fold**ing" of the spectrum. In the next row the data is sampled at 16 ms intervals, and in the last row at 32 ms intervals. The 8 ms sampling seems OK, whereas the 32 ms sampling looks poor. Study how the spectrum changes from one row to the next.

The spectrum suffers no visible harm in the drop from 4 ms to 8 ms. The 8 ms data could be used to construct the original 4 ms data by transforming the 8 ms data to the frequency domain, replacing values at frequencies above 125/2 Hz by zero, and then inverse transforming to the time domain.

(Airguns usually have a higher frequency content than we see here. Some high-frequency energy was removed by the recording geometry, and I also removed some when preparing the data.)

1.6.5 Common signals

Figure 1.14 shows some common signals and their **autocorrelation**s. Figure 1.15 shows the cosine transforms of the autocorrelations. Cosine transform takes us from time to frequency and it also takes us from frequency to time. Thus, transform pairs in Figure 1.15 are sometimes more comprehensible if you interchange time and frequency. The various signals are given names in the figures, and a description of each follows:

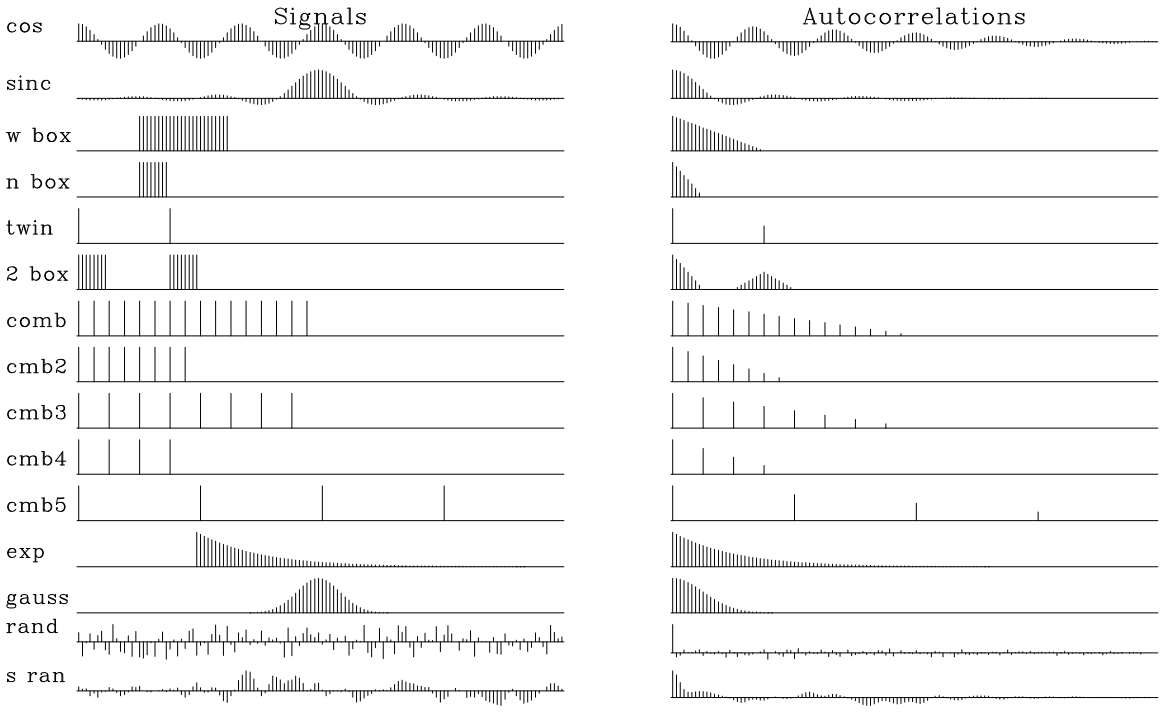

Figure 1.14: Common signals and one side of their autocorrelations. cs-autocor [ER]

cos The theoretical spectrum of a sinusoid is an impulse, but the sinusoid was truncated (mul-

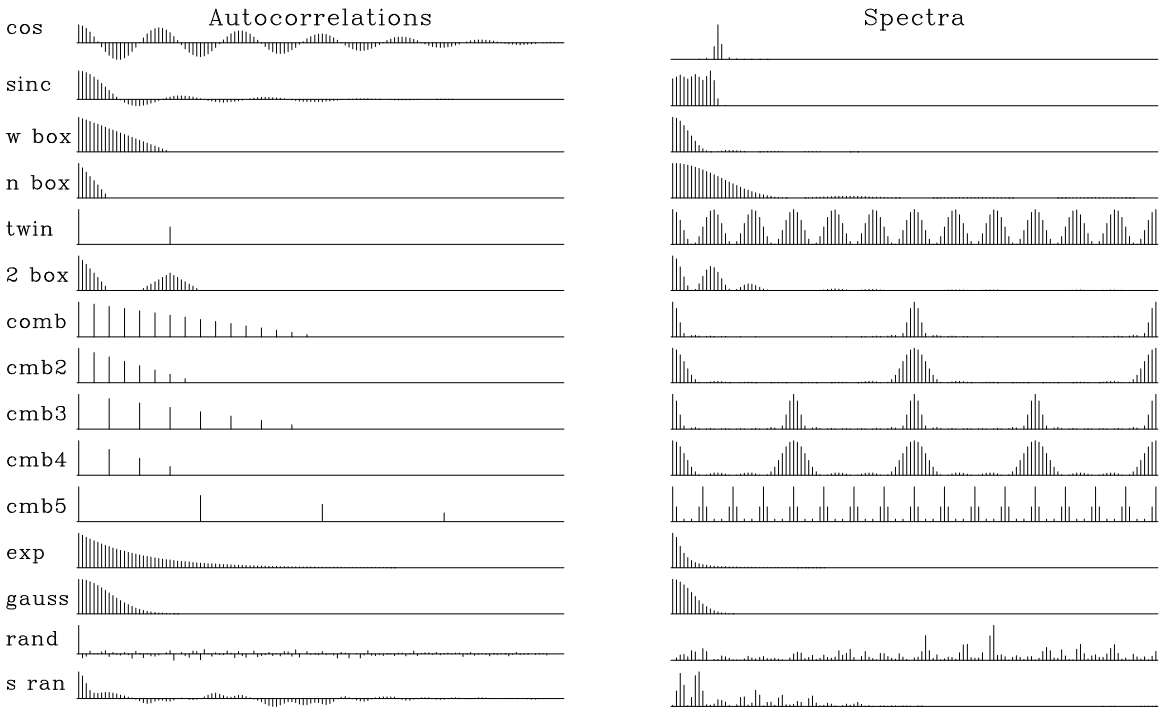

Figure 1.15: Autocorrelations and their cosine transforms, i.e., the (energy) spectra of the common signals. cs-spectra [ER]

tiplied by a rectangle function). The autocorrelation is a sinusoid under a triangle, and its spectrum is a broadened impulse (which can be shown to be a narrow sinc-squared function).

sinc The **sinc** function is $\sin(\omega_0 t)/(\omega_0 t)$. Its autocorrelation is another sinc function, and its spectrum is a rectangle function. Here the rectangle is corrupted slightly by "**Gibbs sidelobes**," which result from the time truncation of the original sinc.

wide box A wide **rectangle function** has a wide triangle function for an autocorrelation and a narrow sinc-squared spectrum.

narrow box A narrow rectangle has a wide sinc-squared spectrum.

twin Two pulses.

2 boxes Two separated narrow boxes have the spectrum of one of them, but this spectrum is modulated (multiplied) by a sinusoidal function of frequency, where the modulation frequency measures the time separation of the narrow boxes. (An oscillation seen in the frequency domain is sometimes called a "**quefrency**.")

comb Fine-toothed-**comb** functions are like rectangle functions with a lower Nyquist frequency. Coarse-toothed-comb functions have a spectrum which is a fine-toothed comb.

exponential The autocorrelation of a transient **exponential** function is a **double-sided exponential** function. The spectrum (energy) is a Cauchy function, $1/(\omega^2+\omega_0^2)$. The curious thing about the **Cauchy function** is that the amplitude spectrum diminishes inversely with frequency to the *first* power; hence, over an infinite frequency axis, the function has infinite integral. The sharp edge at the onset of the transient exponential has much high-frequency energy.

Gauss The autocorrelation of a **Gaussian** function is another Gaussian, and the spectrum is also a Gaussian.

random **Random** numbers have an autocorrelation that is an impulse surrounded by some short grass. The spectrum is positive random numbers.

smoothed random Smoothed random numbers are much the same as random numbers, but their spectral bandwidth is limited.

1.6.6 Spectral transfer function

Filters are often used to change the spectra of given data. With input $X(Z)$, filters $B(Z)$, and output $Y(Z)$, we have $Y(Z)=B(Z)X(Z)$ and the Fourier conjugate $\overline{Y}(1/Z)=\overline{B}(1/Z)\overline{X}(1/Z)$. Multiplying these two relations together, we get

$$\overline{Y}Y \quad = \quad (\overline{B}B)(\overline{X}X) \tag{1.49}$$

which says that the spectrum of the input times the spectrum of the filter equals the spectrum of the output. Filters are often characterized by the shape of their spectra; this shape is the same as the **spectral ratio** of the output over the input:

$$\overline{B}B \quad = \quad \frac{\overline{Y}Y}{\overline{X}X} \tag{1.50}$$

EXERCISES:

1 Suppose a wavelet is made up of complex numbers. Is the autocorrelation relation $s_k = s_{-k}$ true? Is s_k real or complex? Is $S(\omega)$ real or complex?

Chapter 2

Two-dimensional Fourier transform

2.1 SLOW PROGRAM FOR INVERTIBLE FT

Typically, signals are real valued. But the programs in this chapter are for complex-valued signals. In order to use these programs, copy the real-valued signal into a complex array, where the signal goes into the real part of the complex numbers; the imaginary parts are then automatically set to zero.

2.1.1 The slow FT code

The `slowft()` routine exhibits features found in many physics and engineering programs. For example, the time-domain signal (which I call "`tt()`"), has `nt` values subscripted, from `tt(1)` to `tt(nt)`. The first value of this signal `tt(1)` is located in real physical time at `t0`. The time interval between values is `dt`. The value of `tt(it)` is at time `t0+(it-1)*dt`. I do not use "`if`" as a pointer on the frequency axis because `if` is a keyword in most programming languages. Instead, I count along the frequency axis with a variable named `ie`.

```
subroutine slowft( adj, add, t0,dt,tt,nt,  f0,df,  ff,nf)
integer it,ie,     adj, add,          nt,             nf
complex cexp, cmplx,              tt(nt),         ff(nf)
real pi2, freq, time, scale, t0,dt,        f0,df
call adjnull(      adj, add,       tt,2*nt,        ff,2*nf)
                        pi2   = 2. * 3.14159265;
                        scale = 1./sqrt( 1.*nt)
                        df    = (1./dt) / nf
                        f0    =   - .5/dt

do ie = 1, nf {   freq= f0 + df*(ie-1)
do it = 1, nt {   time= t0 + dt*(it-1)

        if( adj == 0 )
                ff(ie)= ff(ie) + tt(it) * cexp(cmplx(0., pi2*freq*time)) * scale
        else
```

```
                tt(it)= tt(it) + ff(ie) * cexp(cmplx(0.,-pi2*freq*time)) * scale
        }}
return; end
```

The total frequency band is 2π radians per sample unit or $1/\Delta t$ Hz. Dividing the total interval by the number of points `nf` gives Δf. We could choose the frequencies to run from 0 to 2π radians/sample. That would work well for many applications, but it would be a nuisance for applications such as differentiation in the frequency domain, which require multiplication by $-i\omega$ including the **negative frequencies** as well as the positive. So it seems more natural to begin at the most negative frequency and step forward to the most positive frequency.

2.2 TWO-DIMENSIONAL FT

Let us review some basic facts about **two-dimensional Fourier transform**. A two-dimensional function is represented in a computer as numerical values in a matrix, whereas a one-dimensional Fourier transform in a computer is an operation on a vector. A 2-D Fourier transform can be computed by a sequence of 1-D Fourier transforms. We can first transform each column vector of the matrix and then each row vector of the matrix. Alternately, we can first do the rows and later do the columns. This is diagrammed as follows:

$$\begin{array}{ccc} p(t,\ x) & \longleftrightarrow & P(t,\ k_x) \\ \updownarrow & & \updownarrow \\ P(\omega,\ x) & \longleftrightarrow & P(\omega,\ k_x) \end{array}$$

The diagram has the notational problem that we cannot maintain the usual convention of using a lower-case letter for the domain of physical space and an upper-case letter for the Fourier domain, because that convention cannot include the mixed objects $P(t,k_x)$ and $P(\omega,x)$. Rather than invent some new notation, it seems best to let the reader rely on the context: the arguments of the function must help name the function.

An example of **two-dimensional Fourier transform**s on typical deep-ocean data is shown in Figure 2.1. In the deep ocean, sediments are fine-grained and deposit slowly in flat, regular, horizontal beds. The lack of permeable rocks such as sandstone severely reduces the potential for petroleum production from the deep ocean. The fine-grained shales overlay irregular, igneous, **basement rock**s. In the plot of $P(t,k_x)$, the lateral continuity of the sediments is shown by the strong spectrum at low k_x. The igneous rocks show a k_x spectrum extending to such large k_x that the deep data may be somewhat **spatially aliased** (sampled too coarsely). The plot of $P(\omega,x)$ shows that the data contains no low-frequency energy. The dip of the sea floor shows up in (ω,k_x)-space as the energy crossing the origin at an angle.

Altogether, the **two-dimensional Fourier transform** of a collection of seismograms involves only twice as much computation as the one-dimensional Fourier transform of each

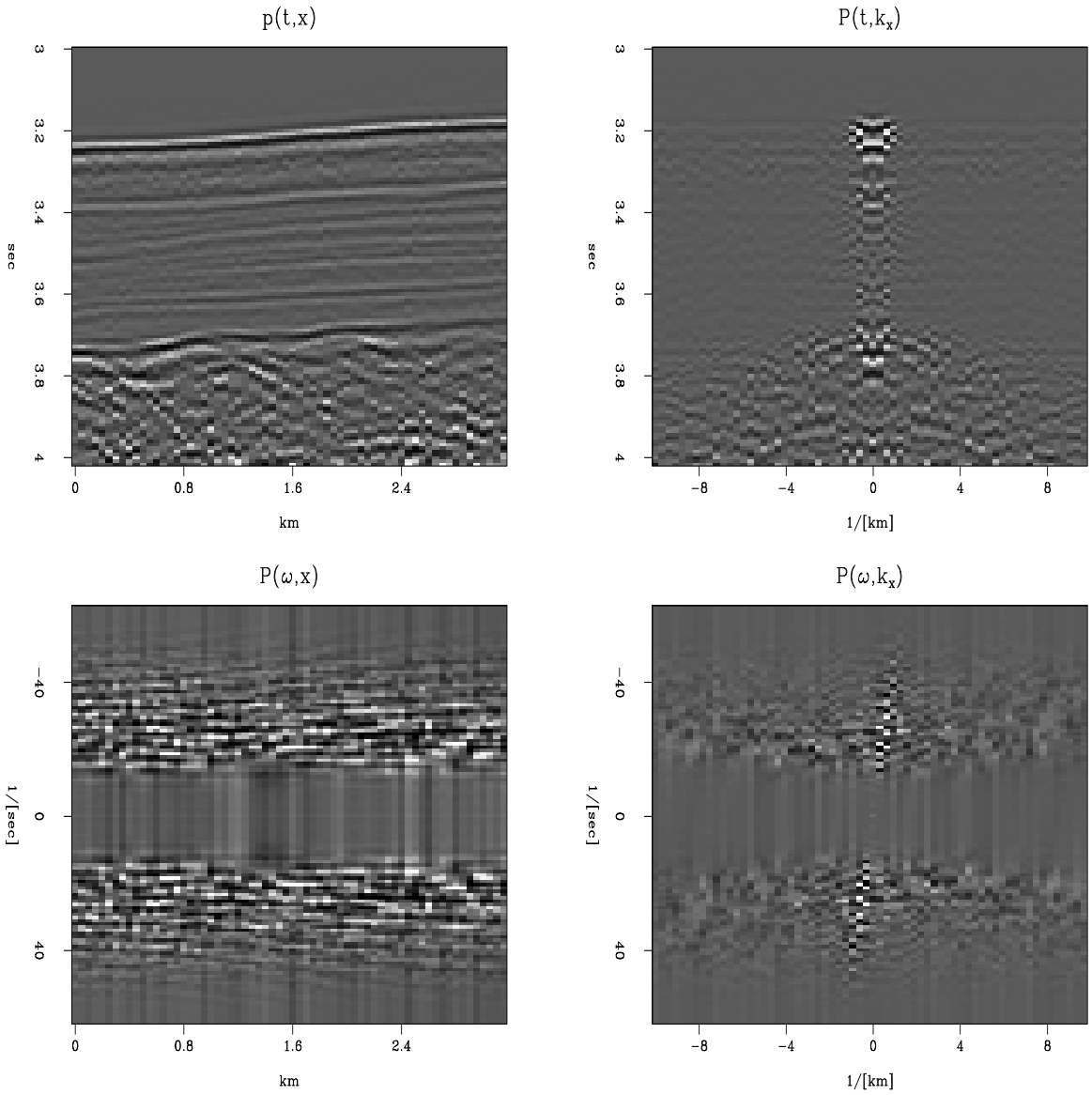

Figure 2.1: A deep-marine dataset $p(t,x)$ from Alaska (U.S. Geological Survey) and the *real* part of various Fourier transforms of it. Because of the long traveltime through the water, the time axis does not begin at $t = 0$. dft-plane4 [ER]

seismogram. This is lucky. Let us write some equations to establish that the asserted procedure does indeed do a 2-D Fourier transform. Say first that any function of x and t may be expressed as a superposition of sinusoidal functions:

$$p(t,x) \;=\; \int\int e^{-i\omega t + ik_x x}\; P(\omega,k_x)\; d\omega\; dk_x \tag{2.1}$$

The double integration can be nested to show that the temporal transforms are done first (inside):

$$\begin{aligned} p(t,x) &= \int e^{i\,k_x x} \left[\int e^{-i\omega t}\; P(\omega,k_x)\; d\omega\right] dk_x \\ &= \int e^{i\,k_x x}\; P(t,k_x)\; dk_x \end{aligned}$$

The quantity in brackets is a Fourier transform over ω done for each and every k_x. Alternately, the nesting could be done with the k_x-integral on the inside. That would imply rows first instead of columns (or vice versa). It is the separability of $\exp(-i\,\omega t + i\,k_x x)$ into a product of exponentials that makes the computation this easy and cheap.

2.2.1 Signs in Fourier transforms

In Fourier transforming t-, x-, and z-coordinates, we must choose a sign convention for each coordinate. Of the two alternative **sign convention**s, electrical engineers have chosen one and physicists another. While both have good reasons for their choices, our circumstances more closely resemble those of physicists, so we will use their convention. For the *inverse* Fourier transform, our choice is

$$p(t,x,z) \quad = \quad \int\int\int e^{-i\omega t + ik_x x + ik_z z}\; P(\omega,k_x,k_z)\; d\omega\, dk_x\, dk_z \tag{2.2}$$

For the *forward* Fourier transform, the space variables carry a *negative* sign, and time carries a *positive* sign.

Let us see the reasons why electrical engineers have made the opposite choice, and why we go with the physicists. Essentially, engineers transform only the time axis, whereas physicists transform both time and space axes. Both are simplifying their lives by their choice of sign convention, but physicists complicate their time axis in order to simplify their many space axes. The engineering choice minimizes the number of minus signs associated with the time axis, because for engineers, d/dt is associated with $i\omega$ instead of, as is the case for us and for physicists, with $-i\omega$. We confirm this with equation (2.2). Physicists and geophysicists deal with many more independent variables than time. Besides the obvious three space axes are their mutual combinations, such as midpoint and offset.

You might ask, why not make *all* the signs positive in equation (2.2)? The reason is that in that case waves would not move in a positive direction along the space axes. This would be especially unnatural when the space axis was a radius. Atoms, like geophysical sources,

always radiate from a point to infinity, not the other way around. Thus, in equation (2.2) the sign of the spatial frequencies must be opposite that of the temporal frequency.

The only good reason I know to choose the engineering convention is that we might compute with an array processor built and microcoded by engineers. Conflict of sign convention is not a problem for the programs that transform complex-valued time functions to complex-valued frequency functions, because there the sign convention is under the user's control. But sign conflict does make a difference when we use any program that converts real-time functions to complex frequency functions. The way to live in both worlds is to imagine that the frequencies produced by such a program do not range from 0 to $+\pi$ as the program description says, but from 0 to $-\pi$. Alternately, we could always take the complex conjugate of the transform, which would swap the sign of the ω-axis.

2.2.2 Examples of 2-D FT

An example of a **two-dimensional Fourier transform** of a pulse is shown in Figure 2.2.

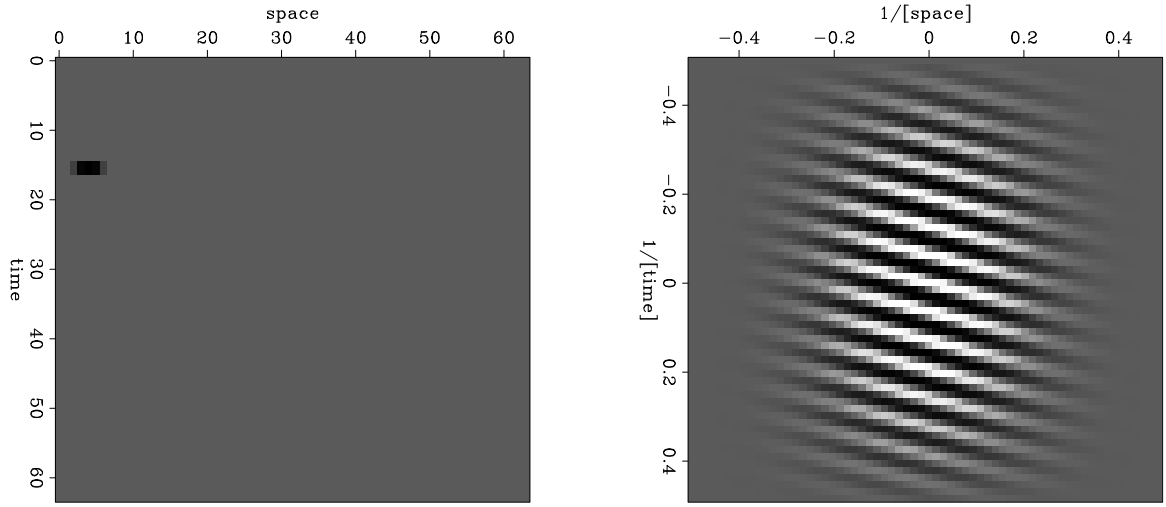

Figure 2.2: A broadened pulse (left) and the real part of its FT (right). dft-ft2dofpulse [ER]

Notice the location of the pulse. It is closer to the time axis than the frequency axis. This will affect the real part of the FT in a certain way (see exercises). Notice the broadening of the pulse. It was an impulse smoothed over time (vertically) by convolution with (1,1) and over space (horizontally) with (1,4,6,4,1). This will affect the real part of the FT in another way.

Another example of a two-dimensional Fourier transform is given in Figure 2.3. This example simulates an impulsive air wave originating at a point on the x-axis. We see a wave propagating in each direction from the location of the source of the wave. In Fourier space there are also two lines, one for each wave. Notice that there are other lines which do not go through the origin; these lines are called "**spatial alias**es." Each actually goes through the origin of another square plane that is not shown, but which we can imagine alongside the one shown. These other planes are periodic replicas of the one shown.

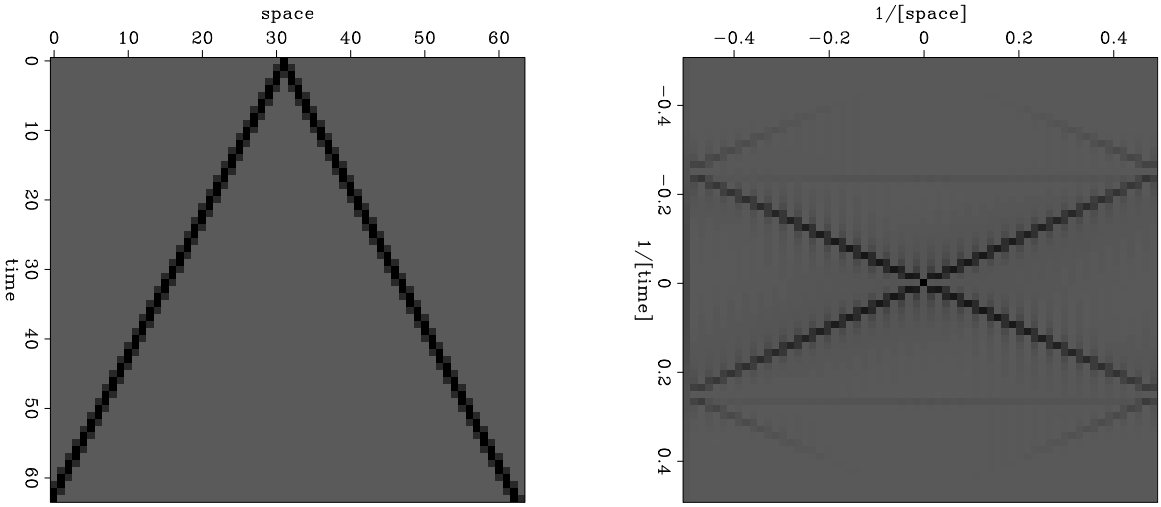

Figure 2.3: A simulated air wave (left) and the amplitude of its FT (right). dft-airwave [ER]

EXERCISES:

1 Most time functions are real. Their imaginary part is zero. Show that this means that $F(\omega,k)$ can be determined from $F(-\omega,-k)$.

2 What would change in Figure 2.2 if the pulse were moved (a) earlier on the t-axis, and (b) further on the x-axis? What would change in Figure 2.2 if instead the time axis were smoothed with (1,4,6,4,1) and the space axis with (1,1)?

3 What would Figure 2.3 look like on an earth with half the earth velocity?

4 Numerically (or theoretically) compute the two-dimensional spectrum of a plane wave $[\delta(t-px)]$, where the plane wave has a randomly fluctuating amplitude: say, rand(x) is a random number between ± 1, and the randomly modulated plane wave is $[(1+.2\,\text{rand}(x))\,\delta(t-px)]$.

5 Explain the horizontal "layering" in Figure 2.1 in the plot of $P(\omega,x)$. What determines the "layer" separation? What determines the "layer" slope?

Chapter 3

Kirchhoff migration

Migration is the basic image making process in reflection seismology. A simplified conceptual basis is (1) the superposition principle, (2) Pythagoras theorem, and (3) a subtle and amazing analogy called the **exploding reflector concept**. After we see this analogy we'll look at the most basic **migration** method, the Kirchhoff method.

3.1 THE EXPLODING REFLECTOR CONCEPT

Figure 3.1 shows two wave-propagation situations. The first is realistic field sounding. A shot s and a receiving geophone g attached together go to all places on the earth surface and record for us the echo function of t. The second is a thought experiment in which the reflectors in the earth suddenly explode. Waves from the hypothetical explosion propagate up to the earth's surface where they are observed by a hypothetical string of geophones.

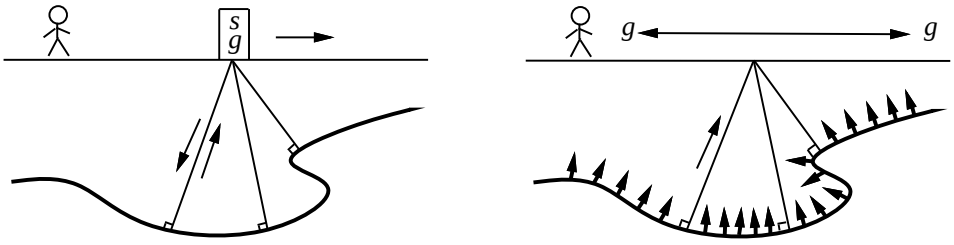

Figure 3.1: Echoes collected with a source-receiver pair moved to all points on the earth's surface (left) and the "exploding-reflectors" conceptual model (right). krch-expref [NR]

Notice in the figure that the ray paths in the field-recording case seem to be the same as

those in the **exploding-reflector** case. It is a great conceptual advantage to imagine that the two wavefields, the observed and the hypothetical, are indeed the same. If they are the same, the many thousands of experiments that have really been done can be ignored, and attention can be focused on the one hypothetical experiment. One obvious difference between the two cases is that in the field geometry waves must first go down and then return upward along the same path, whereas in the hypothetical experiment they just go up. Travel time in field experiments could be divided by two. In practice, the data of the field experiments (two-way time) is analyzed assuming the sound velocity to be half its true value.

3.2 THE KIRCHHOFF IMAGING METHOD

There are two basic tasks to be done: (1) make data from a model, and (2) make models from data. The latter is often called **imaging**.

Imagine the earth had just one point reflector at (x_0, z_0). This reflector explodes at $t = 0$. The data at $z = 0$ is a function of location x and travel time t would be an impulsive signal along the hyperbolic trajectory $t^2 = ((x - x_0)^2 + z_0^2)/v^2$. Imagine the earth had more point reflectors in it. The data would then be a superposition of more hyperbolic arrivals. A dipping bed could be represented as points along a line in (x, z). The data from that dipping bed must be a superposition of many hyperbolic arrivals.

Now let us take the opposite point of view: we have data and we want to compute a model. This is called "**migration**". Conceptually, the simplest approach to migration is also based on the idea of an impulse response. Suppose we recorded data that was zero everywhere except at one point (x_0, t_0). Then the earth model should be a spherical mirror centered at (x_0, z_0) because this model produces the required data, namely, no received signal except when the sender-receiver pair are in the center of the semicircle.

This observation plus the superposition principle suggests an algorithm for making earth images: For each location (x, t) on the data mesh $d(x, t)$ add in a semicircular mirror of strength d into the model $m(x, z)$. You need to add in a semicircle for every value of (t, x). Notice again we use the same equation $t^2 = (x^2 + z^2)/v^2$. This equation is the "conic section". A slice at constant t is a circle in (x, z). A slice at constant z is a hyperbola in (x, t).

Examples are shown in Figure 3.2. Points making up a line reflector diffract to a line reflection, and how points making up a line reflection migrate to a line reflector.

Besides the semicircle superposition migration method, there is another migration method that produces a similar result conceptually, but it has more desirable results numerically. This is the "**adjoint modeling**" idea. In it, we sum the data over a hyperbolic trajectory to find a value in model space that is located at the apex of the hyperbola. This is also called the "**pull**" method of migration.

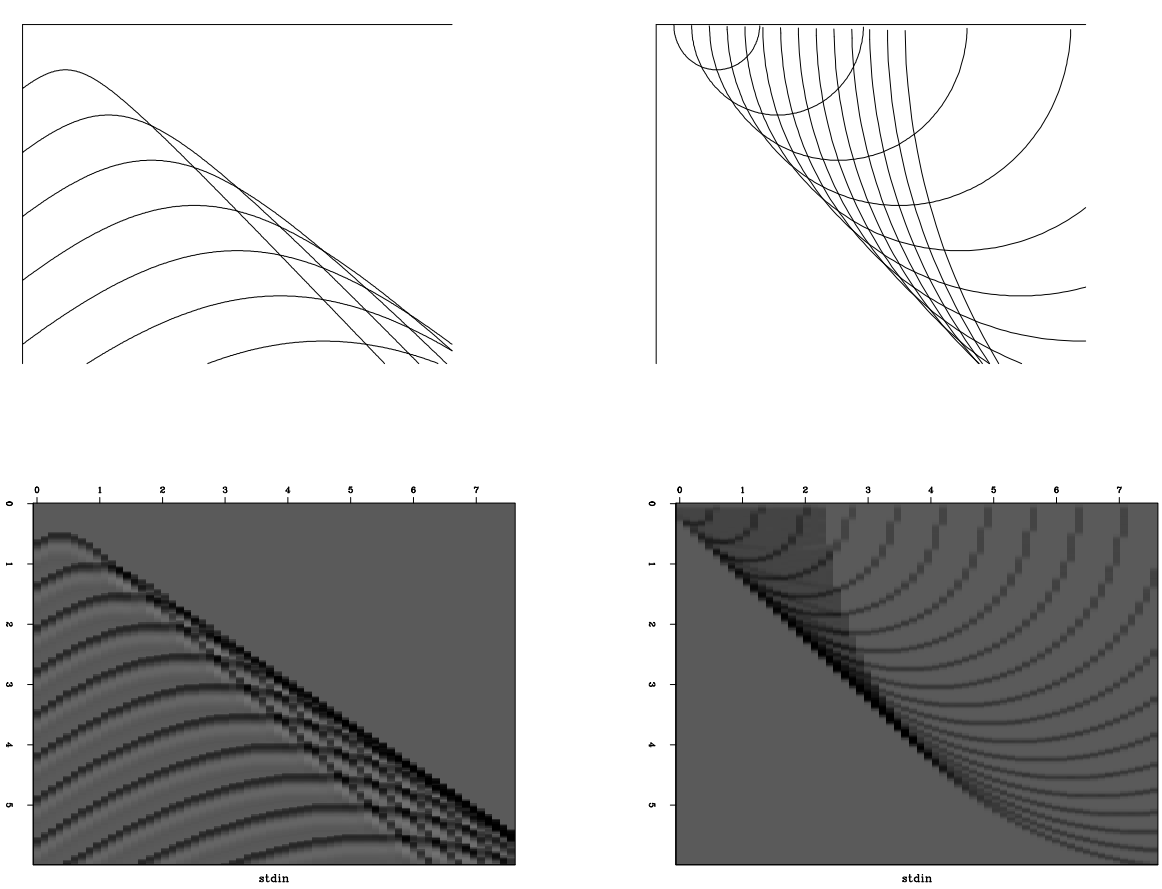

Figure 3.2: Left is a superposition of many hyperbolas. The top of each hyperbola lies along a straight line. That line is like a reflector, but instead of using a continuous line, it is a sequence of points. Constructive interference gives an apparent reflection off to the side. Right shows a superposition of semicircles. The bottom of each semicircle lies along a line that could be the line of an observed plane wave. Instead the plane wave is broken into point arrivals, each being interpreted as coming from a semicircular mirror. Adding the mirrors yields a more steeply dipping reflector. krch-dip [NR]

3.2.1 Tutorial Kirchhoff code

Subroutine `kirchslow()` below is the best tutorial **Kirchhoff migration**-modeling program I could devise. Think of data as a function of traveltime t and the horizontal axis x. Think of model (or image) as a function of traveltime t and the horizontal axis x. The program copies information from data space `data(it,iy)` to model space `modl(iz,ix)` or vice versa.

Data space and model space each have two axes. Of the four axes, three are independent (stated by loops) and the fourth is derived by the circle-hyperbola relation $t^2 = \tau^2 + x^2/v^2$. Subroutine `kirchslow()` for `adj=0` copies information from model space to data space, i.e. from the hyperbola top to its flanks. For `adj=1`, data summed over the hyperbola flanks is put at the hyperbola top.

```
# Kirchhoff migration and diffraction.  (tutorial, slow)
#
subroutine kirchslow(   adj, add,  velhalf, t0,dt,dx, modl,nt,nx,  data)
integer ix,iy,it,iz,nz, adj, add,                          nt,nx
real x0,y0,dy,z0,dz,t,x,y,z,hs,    velhalf, t0,dt,dx, modl(nt,nx), data(nt,nx)
call adjnull(           adj, add,                     modl,nt*nx,  data,nt*nx)
x0=0.;  y0=0;  dy=dx;  z0=t0;  dz=dt; nz=nt
do ix= 1, nx {  x = x0 + dx * (ix-1)
do iy= 1, nx {  y = y0 + dy * (iy-1)
do iz= 1, nz {  z = z0 + dz * (iz-1)            # z = travel-time depth
        hs=      (x-y) / velhalf
        t = sqrt( z * z  +  hs * hs )
        it = 1.5 + (t-t0) / dt
        if( it <= nt )
                if( adj == 0 )
                        data(it,iy) = data(it,iy) + modl(iz,ix)
                else
                        modl(iz,ix) = modl(iz,ix) + data(it,iy)

        }}}
return; end
```

Figure 3.3 shows an example. The model includes dipping beds, syncline, anticline, fault, unconformity, and buried focus. The result is as expected with a "**bow tie**" at the buried focus. On a video screen, I can see hyperbolic events originating from the unconformity and the fault. At the right edge are a few faint **edge artifacts**. We could have reduced or eliminated these edge **artifacts** if we had extended the model to the sides with some empty space.

3.2.2 Kirchhoff artifacts

Given one of our pair of Kirchoff operations, we can manufacture data from models. With the other, we try to reconstruct the original model. This is called imaging. It does not work perfectly. Observed departures from perfection are called "artifacts".

Reconstructing the earth model with the adjoint option yields the result in Figure 3.4. The reconstruction generally succeeds but is imperfect in a number of interesting ways. Near the

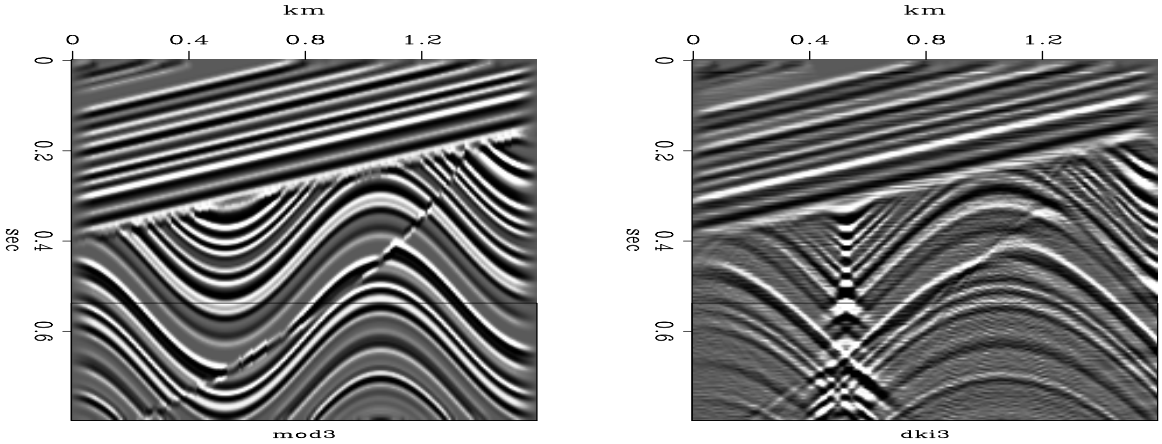

Figure 3.3: Left is the model. Right is diffraction to synthetic data. We notice the "syncline" (depression) turns into a "bow tie" whereas the anticline (bulge up) broadens. krch-kfgood [NR]

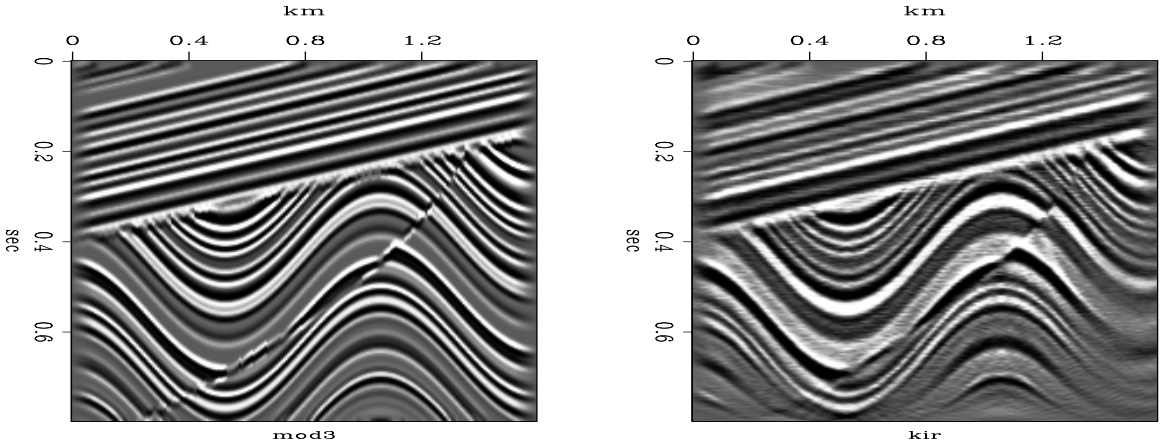

Figure 3.4: Left is the original model. Right is the reconstruction. krch-skmig [NR]

bottom and right side, the reconstruction fades away, especially where the dips are steeper. Bottom fading results because in modeling the data we abandoned arrivals after a certain maximum time. Thus energy needed to reconstruct dipping beds near the bottom was abandoned. Likewise along the side we abandoned rays shooting off the frame.

Unlike the example shown here, real data is notorious for producing **semicircular artifacts**. The simplest explanation is that real data has impulsive noises that are unrelated to primary reflections. Here we have controlled everything fairly well so no such semicircles are obvious, but on a video screen I can see some semicircles.

3.3 SAMPLING AND ALIASING

Spatial aliasing is a vexing issue of numerical analysis. The Kirchhoff codes shown here do not work as expected unless the space mesh size is suitably more refined than the time mesh.

*Spatial **alias**ing* means insufficient sampling of the data along the space axis. This difficulty is so universal, that all migration methods must consider it.

Data should be sampled at more than two points per wavelength. Otherwise the wave arrival direction becomes ambiguous. Figure 3.5 shows synthetic data that is sampled with insufficient density along the x-axis. You can see that the problem becomes more acute at high frequencies and steep dips.

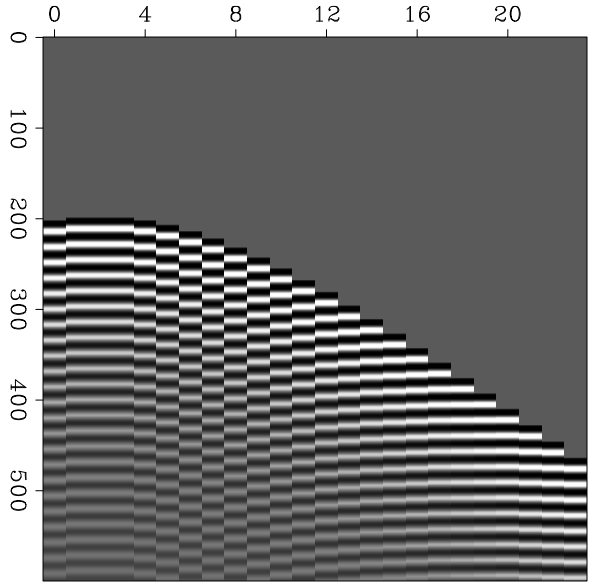

Figure 3.5: Insufficient spatial sampling of synthetic data. To better perceive the ambiguity of arrival angle, view the figures at a grazing angle from the side. krch-alias [NR]

There is no generally-accepted, automatic method for migrating spatially aliased data. In such cases, human beings may do better than machines, because of their skill in recognizing true slopes. When the data is adequately sampled however, computer migrations give better results than manual methods.

Chapter 4

Downward continuation of waves

4.1 DIPPING WAVES

We begin with equations to describe a dipping plane wave in a medium of constant velocity. Figure 4.1 shows a ray moving down into the earth at an angle θ from the vertical. Perpendicular to the ray is a **wavefront**. By elementary geometry the angle between the

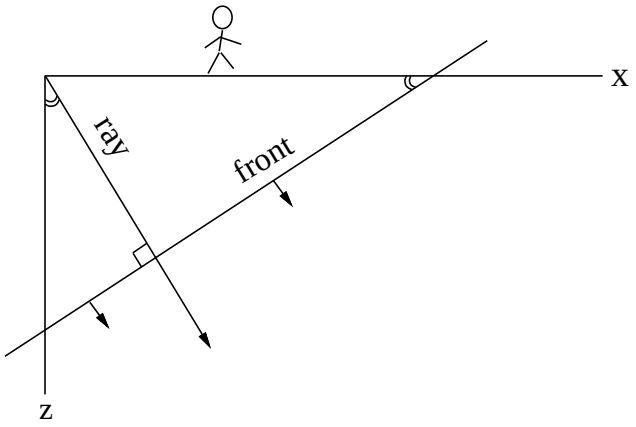

Figure 4.1: Downgoing ray and wavefront. dwnc-front [NR]

wave**front** and the earth's surface is also θ. The **ray** increases its length at a speed v. The speed that is observable on the earth's surface is the intercept of the wavefront with the earth's surface. This speed, namely $v/\sin\theta$, is faster than v. Likewise, the speed of the intercept of the wavefront and the vertical axis is $v/\cos\theta$. A mathematical expression for a straight line like that shown to be the wavefront in Figure 4.1 is

$$z \;=\; z_0 \;-\; x \;\tan\theta \tag{4.1}$$

In this expression z_0 is the intercept between the wavefront and the vertical axis. To make the intercept move downward, replace it by the appropriate velocity times time:

$$z \;=\; \frac{v\,t}{\cos\theta} \;-\; x \;\tan\theta \tag{4.2}$$

Solving for time gives

$$t(x,z) \;=\; \frac{z}{v}\,\cos\theta \;+\; \frac{x}{v}\,\sin\theta \tag{4.3}$$

Equation (4.3) tells the time that the wavefront will pass any particular location (x,z). The expression for a shifted waveform of arbitrary shape is $f(t-t_0)$. Using (4.3) to define the time shift t_0 gives an expression for a wavefield that is some waveform moving on a **ray**.

$$\text{moving wavefield} \;=\; f\left(\, t \;-\; \frac{x}{v}\,\sin\theta \;-\; \frac{z}{v}\,\cos\theta \right) \tag{4.4}$$

4.1.1 Snell waves

In reflection seismic surveys the velocity contrast between shallowest and deepest reflectors ordinarily exceeds a factor of two. Thus depth variation of velocity is almost always included in the analysis of field data. Seismological theory needs to consider waves that are just like plane waves except that they bend to accommodate the velocity stratification $v(z)$. Figure 4.2 shows this in an idealized geometry: waves radiated from the horizontal flight of a supersonic airplane. The airplane passes location x at time $t_0(x)$ flying horizontally at a constant speed.

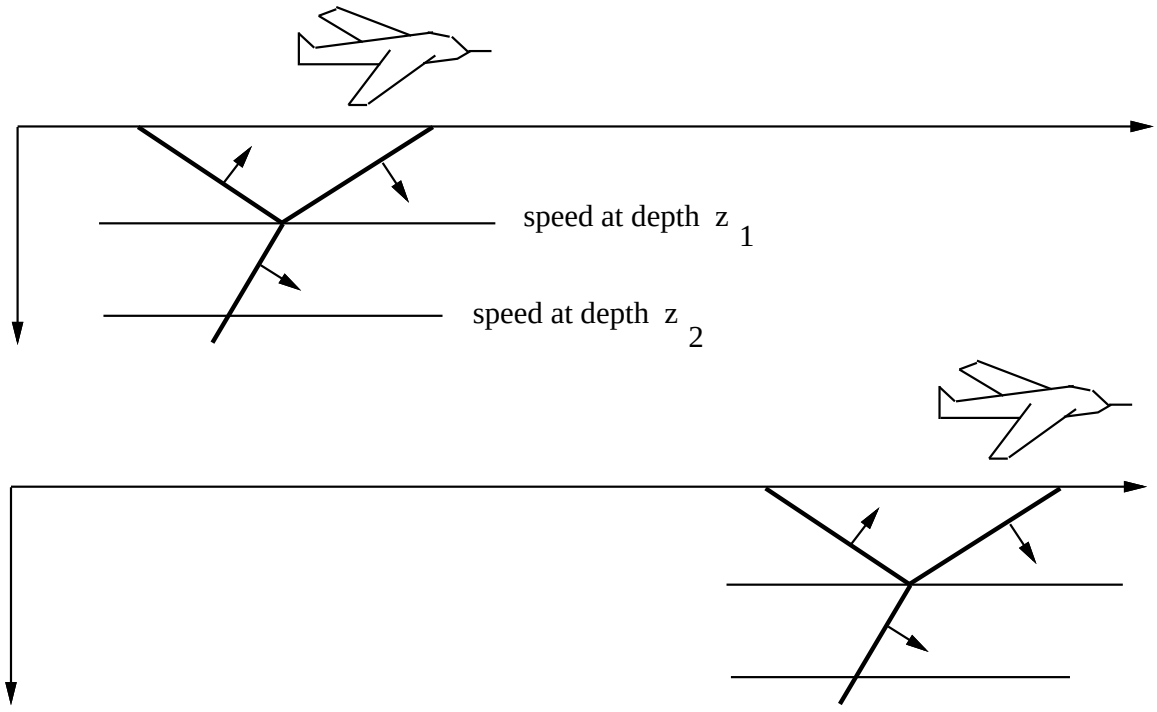

Figure 4.2: Fast airplane radiating a sound wave into the earth. From the figure you can deduce that the horizontal speed of the wavefront is the same at depth z_1 as it is at depth z_2. This leads (in isotropic media) to Snell's law. dwnc-airplane [NR]

Imagine an earth of horizontal plane layers. In this model there is nothing to distinguish any point on the x-axis from any other point on the x-axis. But the seismic velocity varies from layer to layer. There may be reflections, head waves, shear waves, converted waves,

anisotropy, and multiple reflections. Whatever the picture is, it moves along with the airplane. A picture of the wavefronts near the airplane moves along with the airplane. The top of the picture and the bottom of the picture both move laterally at the same speed even if the earth velocity increases with depth. If the top and bottom didn't go at the same speed, the picture would become distorted, contradicting the presumed symmetry of translation. This horizontal speed, or rather its inverse $\partial t_0/\partial x$, has several names. In practical work it is called the ***stepout***. In theoretical work it is called the ***ray parameter***. It is very important to note that $\partial t_0/\partial x$ does not change with depth, even though the seismic velocity does change with depth. In a constant-velocity medium, the angle of a wave does not change with depth. In a stratified medium, $\partial t_0/\partial x$ does not change with depth.

Figure 4.3 illustrates the differential geometry of the wave. The diagram shows that

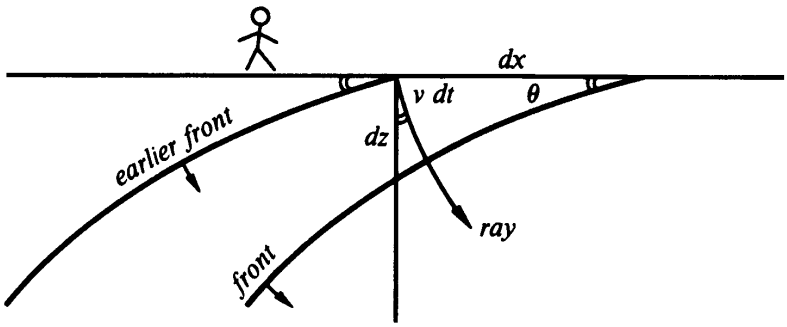

Figure 4.3: Downgoing fronts and rays in stratified medium $v(z)$. The wavefronts are horizontal translations of one another. dwnc-frontz [NR]

$$\frac{\partial t_0}{\partial x} = \frac{\sin\theta}{v} \tag{4.5}$$

$$\frac{\partial t_0}{\partial z} = \frac{\cos\theta}{v} \tag{4.6}$$

These two equations define two (inverse) speeds. The first is a horizontal speed, measured along the earth's surface, called the *horizontal* ***phase velocity***. The second is a vertical speed, measurable in a borehole, called the *vertical phase velocity.* Notice that both these speeds *exceed* the velocity v of wave propagation in the medium. Projection of wave *fronts* onto coordinate axes gives speeds larger than v, whereas projection of *rays* onto coordinate axes gives speeds smaller than v. The inverse of the phase velocities is called the ***stepout*** or the ***slowness***.

Snell's law relates the angle of a wave in one layer with the angle in another. The constancy of equation (4.5) in depth is really just the statement of Snell's law. Indeed, we have just derived Snell's law. All waves in seismology propagate in a velocity-stratified medium. So they cannot be called plane waves. But we need a name for waves that are near to plane waves. A ***Snell wave*** will be defined to be the generalization of a plane wave to a stratified medium $v(z)$. A plane wave that happens to enter a medium of depth-variable velocity $v(z)$ gets changed into a Snell wave. While a plane wave has an angle of propagation, a Snell wave has instead a ***Snell parameter*** $p = \partial t_0/\partial x$.

It is noteworthy that Snell's parameter $p = \partial t_0/\partial x$ is directly observable at the surface, whereas neither v nor θ is directly observable. Since $p = \partial t_0/\partial x$ is not only observable, but constant in depth, it is customary to use it to eliminate θ from equations (4.5) and (4.6):

$$\frac{\partial t_0}{\partial x} = \frac{\sin\theta}{v} = p \tag{4.7}$$

$$\frac{\partial t_0}{\partial z} = \frac{\cos\theta}{v} = \sqrt{\frac{1}{v(z)^2} - p^2} \tag{4.8}$$

4.1.2 Evanescent waves

Suppose the velocity increases to infinity at infinite depth. Then equation (4.8) tells us that something strange happens when we reach the depth for which p^2 equals $1/v(z)^2$. That is the depth at which the ray turns horizontal. Below this critical depth the seismic wavefield damps exponentially with increasing depth. Such waves are called **evanescent**. For a physical example of an evanescent wave, forget the airplane and think about a moving bicycle. For a bicyclist, the slowness p is so large that it dominates $1/v(z)^2$ for all earth materials. The bicyclist does not radiate a wave, but produces a ground deformation that decreases exponentially into the earth. To radiate a wave, a source must move faster than the material velocity.

4.2 DOWNWARD CONTINUATION

Given a vertically upcoming plane wave at the earth's surface, say $u(t,x,z=0) = u(t)\text{const}(x)$, and an assumption that the earth's velocity is vertically stratified, i.e. $v = v(z)$, we can presume that the upcoming wave down in the earth is simply time-shifted from what we see on the surface. (This assumes no multiple reflections.) Time shifting can be represented as a linear operator in the time domain by representing it as convolution with an impulse function. In the frequency domain, time shifting is simply multiplying by a complex exponential. This is expressed as

$$u(t,z) = u(t,z=0) * \delta(t + z/v) \tag{4.9}$$

$$U(\omega,z) = U(\omega,z=0)\, e^{-i\omega z/v} \tag{4.10}$$

4.2.1 Continuation of a dipping plane wave.

Next consider a plane wave **dip**ping at some angle θ. It is natural to imagine continuing such a wave back along a ray. Instead, we will continue the wave straight down. This requires the assumption that the plane wave is a perfect one, namely that the same waveform is observed at all x. Imagine two sensors in a vertical well bore. They should record the same signal except for a time shift that depends on the angle of the wave. Notice that the arrival time difference between sensors at two different depths is greatest for vertically propagating waves,

and the time difference drops to zero for horizontally propagating waves. So the time shift Δt is $v^{-1}\cos\theta\,\Delta z$ where θ is the angle between the wavefront and the earth's surface (or the angle between the well bore and the ray). Thus an equation to **downward continue** the wave is

$$U(\omega,\theta,z+\Delta z) = U(\omega,\theta,z)\,\exp(-i\omega\,\Delta t) \tag{4.11}$$

$$U(\omega,\theta,z+\Delta z) = U(\omega,\theta,z)\,\exp\left(-i\omega\,\frac{\Delta z\cos\theta}{v}\right) \tag{4.12}$$

Equation (4.12) is a downward continuation formula for any angle θ. We can generalize the method to media where the velocity is a function of depth. Evidently we can apply equation (4.12) for each layer of thickness Δz, and allow the velocity vary with z. This is a well known approximation that handles the timing correctly but keeps the amplitude constant (since $|e^{i\phi}|=1$) when in real life, the amplitude should vary because of reflection and transmission coefficients. Suffice it to say that in practical earth imaging, this approximation is almost universally satisfactory.

In a stratified earth, it is customary to eliminate the angle θ which is depth variable, and change it to the Snell's parameter p which is constant for all depths. Thus the downward continuation equation for any Snell's parameter is

$$U(\omega,p,z+\Delta z) = U(\omega,p,z)\,\exp\left(-\frac{i\omega\Delta z}{v(z)}\,\sqrt{1-p^2v(z)^2}\right) \tag{4.13}$$

It is natural to wonder where in real life we would encounter a **Snell wave** that we could downward continue with equation (4.13). The answer is that any wave from real life can be regarded as a sum of waves propagating in all angles. Thus a field data set should first be decomposed into Snell waves of all values of p, and then equation (4.13) can be used to downward continue each p, and finally the components for each p could be added. This process akin to Fourier analysis. We now turn to Fourier analysis as a method of downward continuation which is the same idea but the task of decomposing data into Snell waves becomes the task of decomposing data into sinusoids along the x-axis.

4.2.2 Downward continuation with Fourier transform

One of the main ideas in Fourier analysis is that an impulse function (a delta function) can be constructed by the superposition of sinusoids (or complex exponentials). In the study of time series this construction is used for the *impulse response* of a filter. In the study of functions of space, it is used to make a physical point source that can manufacture the downgoing waves that initialize the reflection seismic experiment. Likewise observed upcoming waves can be Fourier transformed over t and x.

Recall that a plane wave carrying an arbitrary waveform is given by equation (4.4). Specializing the arbitrary function to be the real part of the function $\exp[-i\omega(t-t_0)]$ gives

$$\text{moving cosine wave} = \cos\left[\,\omega\left(\frac{x}{v}\,\sin\theta\,+\,\frac{z}{v}\,\cos\theta\,-\,t\,\right)\,\right] \tag{4.14}$$

Using Fourier integrals on time functions we encounter the ***Fourier kernel*** $\exp(-i\omega t)$. To use Fourier integrals on the space-axis x the spatial angular frequency must be defined. Since we will ultimately encounter many space axes (three for shot, three for geophone, also the midpoint and offset), the convention will be to use a subscript on the letter k to denote the axis being Fourier transformed. So k_x is the angular spatial frequency on the x-axis and $\exp(ik_x x)$ is its Fourier kernel. For each axis and Fourier kernel there is the question of the sign before the i. The sign convention used here is the one used in most physics books, namely, the one that agrees with equation (4.14). With this convention, a wave moves in the *positive* direction along the space axes. Thus the Fourier kernel for (x,z,t)-space will be taken to be

$$\text{Fourier kernel} \;=\; e^{i\,k_x x}\; e^{i\,k_z z}\; e^{-i\omega t} \;=\; \exp[i(k_x x \;+\; k_z z \;-\; \omega t)] \tag{4.15}$$

Now for the whistles, bells, and trumpets. Equating (4.14) to the real part of (4.15), physical angles and velocity are related to Fourier components. The Fourier kernel has the form of a plane wave. These relations should be memorized!

Angles and Fourier Components	
$\sin\theta = \dfrac{v\;k_x}{\omega}$	$\cos\theta = \dfrac{v\;k_z}{\omega}$

(4.16)

A point in (ω,k_x,k_z)-space is a plane wave. The one-dimensional Fourier kernel extracts frequencies. The multi-dimensional Fourier kernel extracts (monochromatic) plane waves.

Equally important is what comes next. Insert the angle definitions into the familiar relation $\sin^2\theta + \cos^2\theta = 1$. This gives a most important relationship:

$$k_x^2 \;+\; k_z^2 \;=\; \frac{\omega^2}{v^2} \tag{4.17}$$

The importance of (4.17) is that it enables us to make the distinction between an arbitrary function and a chaotic function that actually is a wavefield. Imagine any function $u(t,x,z)$. Fourier transform it to $U(\omega,k_x,k_z)$. Look in the (ω,k_x,k_z)-volume for any nonvanishing values of U. You will have a wavefield if and only if all nonvanishing U have coordinates that satisfy (4.17). Even better, in practice the (t,x)-dependence at $z=0$ is usually known, but the z-dependence is not. Then the z-dependence is found by assuming U is a wavefield, so the z-dependence is inferred from (4.17).

Equation (4.17) also achieves fame as the "dispersion relation of the scalar **wave equation**," a topic developed more fully in IEI.

4.2.3 Linking Snell waves to Fourier transforms

To link **Snell wave**s to Fourier transforms we merge equations (4.5) and (4.6) with equations (4.16). To do this, move with the wave by setting the phase to be constant, say zero, so

$0 = i(k_x x + k_z z - \omega t)$ and form the derivative of t by x and by z.

$$\frac{k_x}{\omega} = \frac{\partial t_0}{\partial x} = \frac{\sin\theta}{v} = p \tag{4.18}$$

$$\frac{k_z}{\omega} = \frac{\partial t_0}{\partial z} = \frac{\cos\theta}{v} = \frac{\sqrt{1-p^2v^2}}{v} \tag{4.19}$$

The basic **downward continuation** equation for upcoming waves in Fourier space follows from equation (4.13) by eliminating p by using equation (4.18). For analysis of real seismic data we introduce a minus sign because equation (4.19) refers to downgoing waves and observed data is made from up-coming waves.

$$U(\omega, k_x, z+\Delta z) = U(\omega, k_x, z) \exp\left(-\frac{i\omega\Delta z}{v} \sqrt{1 - \frac{v^2 k_x^2}{\omega^2}} \right) \tag{4.20}$$

In Fourier space we delay signals by multiplying by $e^{i\omega\Delta t}$, analogously, equation (4.20) says we downward continue signals into the earth by multiplying by $e^{ik_z\Delta z}$. Multiplication in the Fourier domain means convolution in time which can be depicted by the engineering diagram in Figure 4.4.

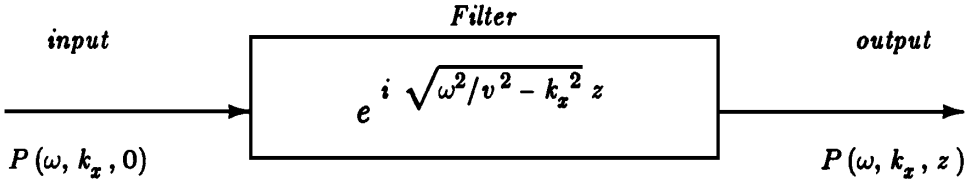

Figure 4.4: Downward continuation of a downgoing wavefield. dwnc-inout [NR]

Downward continuation is a product relationship in both the ω-domain and the k_x-domain. Thus it is a convolution in both time and x. What does the filter look like in the time and space domain? It turns out like a cone, that is, it is roughly an impulse function of $x^2 + z^2 - v^2t^2$. More precisely, it is the Huygens secondary wave source that was exemplified by ocean waves entering a gap through a storm barrier. Adding up the response of multiple gaps in the barrier would be convolution over x.

Migration is more than downward continuation of wave fields. You start with a surface data set $d(t,x)$ and downward continue it to all depth z getting $d(t,x,z)$. Then you have the earth image by selecting the wavefield at time $t = 0$ by the **exploding reflector** concept.

In practice, the downward continuation is done in the frequency domain. Recall to find the zero frequency component $F(\omega = 0)$ we sum over all time t. Interchanging time domain and frequency domain, we realize that to find the exploding reflector signal at zero time $f(t = 0)$ we can simply sum over all frequencies ω.

A nuisance of using Fourier transforms in migration and modeling is that spaces become periodic. This is demonstrated in Figure 4.5. Anywhere an event exits the frame at a side,

top, or bottom boundary, the event immediately emerges on the opposite side. In practice, the unwelcome effect of **periodicity** is generally ameliorated by **padding** zeros around the data and the model.

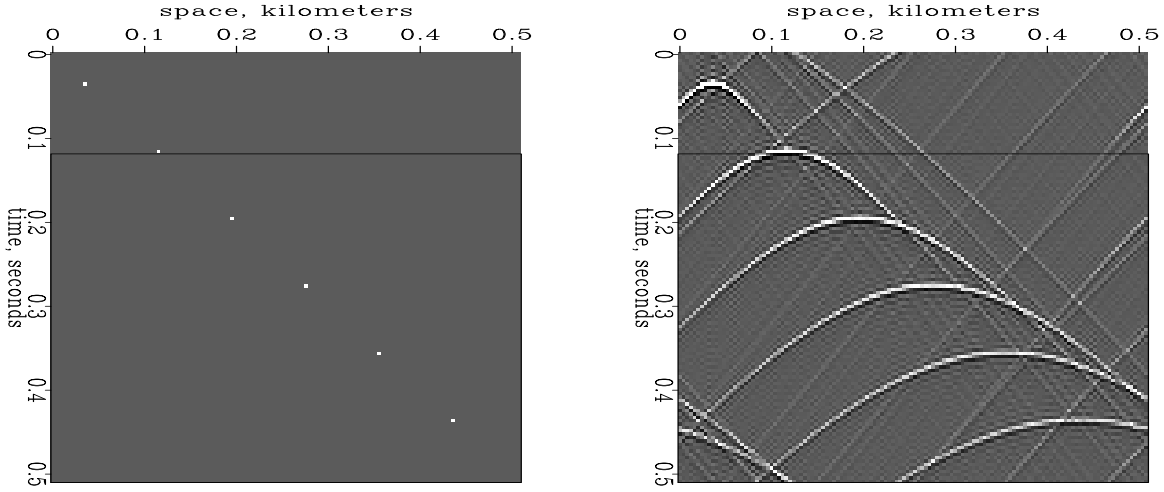

Figure 4.5: A reflectivity model on the left and synthetic data using a Fourier method on the right. dwnc-diag [ER]

4.3 DOWNWARD CONTINUATION AND IMAGING

By now you are probably wondering what good are downward-continued seismic waves. The answer is that once we know the wavefields inside the earth we can do many seismic data analysis tasks. The obvious one is migration. We can do it better than Kirchhoff. Given the upcoming wave at the earth surface Wave$(t,x,z=0)$, pushing it down to all z in the earth gives us Wave(t,x,z). The exploding reflector idea is that the source is on the reflectors at $t=0$. Thus we construct images with

$$\text{Image}\ (x,z) \quad = \quad \text{Wave}\ (t=0,x,z) \tag{4.21}$$

There is a curious symmetry between the input and output of the imaging process:

observations		model
$z=0$ all t	migration $\longrightarrow$ $\longleftarrow$ diffraction	$t=0$ all z

Diffraction is sometimes regarded as the natural process that creates and enlarges hyperboloids. *Migration* is the computer process that does the reverse.

Compared to the **Kirchhoff** method, **downward continuation** methods offer several advantages: (1) we handle $v(z)$ media exactly without calculating any ray paths; (2) we get the correct amplitude versus angle; (3) we automatically include various phase shifts and wave phenomena; (4) interpolation and aliasing artifacts are much reduced.

Chapter 5

Waves in layered media

The gravitational stratification of the earth arouses our interest in **layered media**. There are many kinds of waves that can exist. Here we study waves without assuming knowledge of physics and differential equations. We use only assumptions about the general principles of delay, continuity, and **energy conservation**. The results are directly applicable to sound waves, water waves, light in thin films, normal incident elastic waves of both pressure and shear type, electromagnetic waves, transmission lines, and electrical ladder networks. This chapter seems to concern itself only with vertically incident waves. Actually, with certain precautions, the theory here applies to the set of all waves with a fixed **Snell parameter** $p = \sin\theta/v$.

A lengthier version of this chapter appears in chapter 8 in my book **Fundamental of Geophysical Data Processing** (**FGDP**) (available on the web). The lengthier web version also presents the solution to the more practical problem, to deduce the reflection coefficients given the observed waves.

5.1 REFLECTION AND TRANSMISSION COEFFICIENTS

Consider two halfspaces (deep ocean on top of earth, for example). If a wave of unit amplitude is incident onto the boundary, there is a **transmitted wave** of amplitude t and a **reflected wave** of amplitude c as depicted in Figure 5.1. A very simple relationship exists between t and c. The wave amplitudes have a physical meaning of something like pressure, material displacement, traction, or tangential electric or magnetic fields. These physical variables must be the same value on either side of the boundary. This means the transmitted wave must equal the sum of the incident plus reflected waves.

$$t = 1 + c \tag{5.1}$$

$$t' = 1 + c' \tag{5.2}$$

The **reflection coefficient** c may be positive or negative so the **transmission coefficient** t may be greater than unity. It may seem surprising that t can be greater than unity. This does not

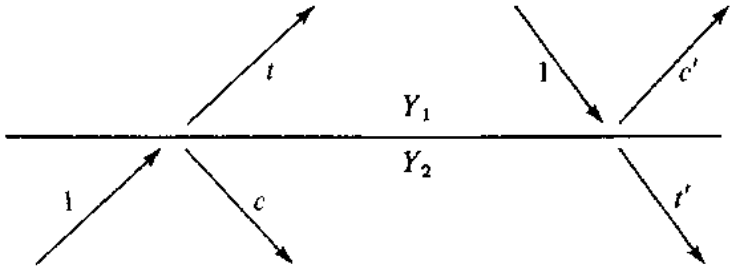

Figure 5.1: Waves incident, reflected c, and transmitted t at an interface. fgdp8-8-1 [NR]

violate any physical laws. At the seashore we see waves approaching the shore and they get larger as they arrive. Energy is not determined by wave height alone. Energy is equal to the squared wave amplitude multiplied by a proportionality factor Y depending upon the medium in which the wave is measured. If we denote the factor of the top medium by Y_1 and the bottom by Y_2, then the statement that the energy before incidence equals the energy after incidence is

$$Y_2 1^2 = Y_2 c^2 + Y_1 t^2 \tag{5.3}$$

solving for c leads us to

$$\begin{aligned} 0 &= -Y_2 + Y_2 c^2 + Y_1 (1+c)^2 \\ 0 &= Y_2 (c-1) + Y_1 (1+c) \\ 0 &= (Y_2 + Y_1) c + (Y_1 - Y_2) \\ c &= \frac{Y_2 - Y_1}{Y_2 + Y_1} \end{aligned} \tag{5.4}$$

In acoustics the up- and downgoing wave variables may be normalized to either pressure or velocity. When they measure velocity, the scale factor multiplying velocity squared is called the **impedance** I. When they measure pressure, the scale factor is called the **admittance** Y.

The wave c' which reflects when energy is incident from the other side is obtained from (5.4) if Y_1 and Y_2 are interchanged. Thus

$$c' = -c \tag{5.5}$$

A perfectly reflecting interface is one which does not allow energy through. This comes about not only when $t = 0$ or $c = -1$, but also when $t = 2$ or $c = +1$. To see this, note that on the left in Figure 5.1

$$\begin{aligned} \frac{\text{Energy transmitted}}{\text{Energy incident}} &= \frac{Y_1 t^2}{Y_2 1^2} = \frac{Y_1}{Y_2}\left(1 + \frac{Y_2 - Y_1}{Y_1 + Y_2}\right)^2 \\ &= \frac{Y_1}{Y_2}\left(\frac{2\,Y_2}{Y_1 + Y_2}\right)^2 = \frac{4\,Y_1\,Y_2}{(Y_1 + Y_2)^2} \end{aligned} \tag{5.6}$$

Equation (5.6) says that 100 percent of the incident energy is transmitted when $Y_1 = Y_2$, but the percentage of transmission is very small when Y_1 and Y_2 are very different.

Ordinarily there are two kinds of variables used to describe waves, and both of these can be continuous at a material discontinuity. One is a scalar like pressure, tension, voltage, potential, stress, or temperature. The other is a vector which we use the vertical component. Examples of the latter are velocity, stretch, electric current, dislacement, and heat flow. Occasionally a wave variable is a **tensor**. When a **boundary condition** is the vanishing of one of the motion components, then the boundary is often said to be rigid. When it is the pressure or potential which vanishes, then the boundary is often said to be free. Rigid and free boundaries reflect waves with unit magnitude reflection coefficients.

A goal here is to establish fundamental mathematical properties of waves in layers while minimizing specialization to any particular physical type of waves. Each physical problem has its own differential equations. These equations are Fourier transformed over x and y leading to coupled ordinary differential equations in depth z. This analytical process is explained in more detail in FGDP. The next step is an **eigenvector** analysis that relates the physical variables to our abstract up- and down-going waves U and D. In order to better understand boundary conditions we will examine one concrete example, acoustic waves.

In acoustics we have **pressure** P and vertical component of **parcel velocity** W (not to be confused with wave velocity v). The acoustic pressure P is the sum of U and D. Vertical velocity W obviously changes sign when the z axis changes sign (interchange of up and down) and that accounts for the minus sign in the definition of W. (The eigenvector analysis provides us with the scaling factor Y.)

$$P = D + U \tag{5.7}$$

$$W = (D - U)Y \tag{5.8}$$

These definitions are easily inverted.

$$D = \frac{P + W/Y}{2} \tag{5.9}$$

$$U = \frac{P - W/Y}{2} \tag{5.10}$$

For sound waves in the ocean, the sea surface is nearly a perfect reflector because of the great contrast between air and water. If this interface is idealized to a perfect reflector, then it is a free surface. Since the pressure vanishes on a free surface, we have $D = -U$ at the surface so the reflection coefficient is -1 as shown in Figure 5.2. In principle we should measure velocity W at the water surface. In practice, we generally measure pressure P a few meters below the free surface.

The pressure normally vanishes at the sea surface, but if we wish to initiate an impulsive disturbance, the pressure may momentarily take on some other value, say 1. This 1 denotes a constant function of frequency which is an impulsive function of time at $t = 0$. Ensuing waves are depicted in Figure 5.3 where the upcoming wave $-R(Z)$ is a consequence of both the downgoing 1 and the downgoing $+R(Z)$. The vertical component of velocity W of the sea surface due to the source and to the resulting acoustic wave is $D - U = 1 + 2R(Z)$.

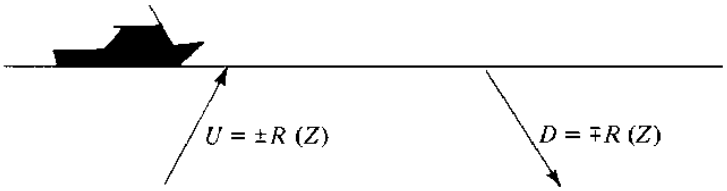

Figure 5.2: A waveform $R(Z)$ reflecting at the surface of the sea. Pressure equal to $U+D$ vanishes at the surface. The vertical velocity of the surface is proportional to $D-U$. Theoretically, waves are observed by measuring W at the surface. fgdp8-8-2 [NR]

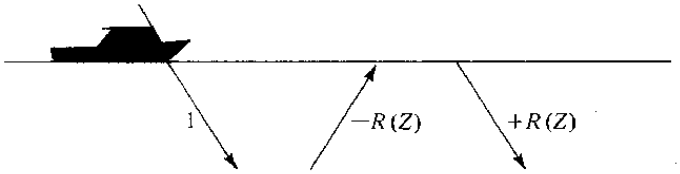

Figure 5.3: An initial downgoing disturbance 1 results in a later upgoing reflected wave $-R(Z)$ which reflects back down as $R(Z)$. The pressure at the surface is $D+U=1+R-R=1$. fgdp8-8-3 [NR]

EXERCISES:

1 In a certain application continuity is expressed by saying that $D-U$ is the same on either side of the interface. This implies that $t=1-c$. Derive an equation like (5.4) for the reflection coefficient in terms of the admittance Y.

2 What are reflection and transmission coefficients in terms of the impedance I? (Clear fractions from your result.)

3 From the principle of energy conservation we showed that $c'=-c$. It may also be deduced from time reversal. To do this, copy Figure 5.1 with arrows reversed. Scale and linearly superpose various figures in an attempt to create a situation where a figure like the right-hand side of Figure 5.1 has $-c'$ for the reflected wave. (HINT: Draw arrows at normal incidence.)

5.2 THE LAYER MATRIX

Consider a layer with waves **reverberating** in it as shown in Figure 5.4. Let the delay operator Z refer to the time delay of a travel path across the layer and back again. The wave seen above the layer is an **infinite sequence** that is expressible as a simple denominator (because $1+X+X^2+X^3+\cdots=1/(1-X)$).

$$t_2t_1[1+c_1c_2'Z+(c_1c_2')^2Z^2+\cdots] \quad = \quad \frac{t_2t_1}{1-c_1c_2'Z}$$

It is no accident that the infinite series may be summed. We will soon see that for n layers the waves, which are of infinite duration, may be expressed as simple polynomials of degree

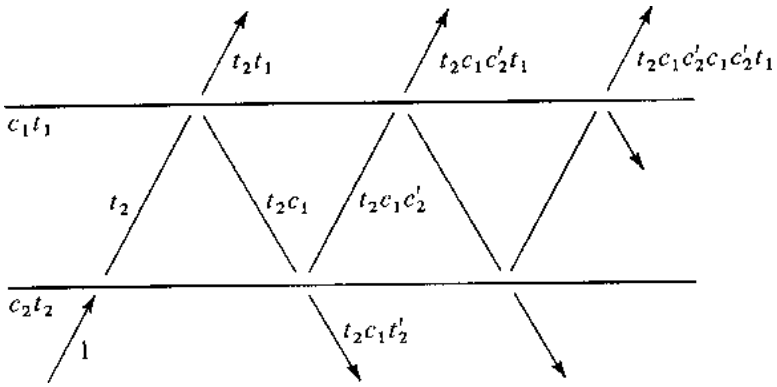

Figure 5.4: Some rays corresponding to resonance in a layer. fgdp8-8-4 [NR]

n. We consider many layers and the general problem of determining waves given reflection coefficients.

Equation (5.11) relates to Figure 5.5 and shows how from the waves U and D' we extrapolate into the future to get U' and D.

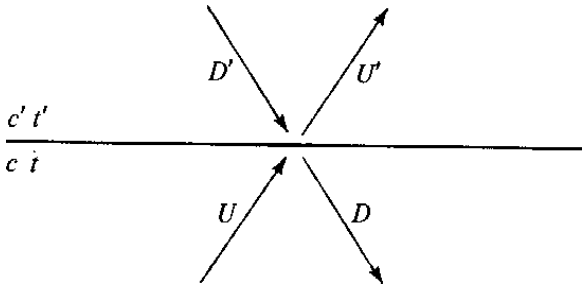

Figure 5.5: Waves incident and reflected from an interface. fgdp8-8-5 [NR]

$$\begin{aligned} U' &= tU + c'D' \\ D &= cU + t'D' \end{aligned} \tag{5.11}$$

Let us rearrange (5.11) to get U' and D' on the right and U and D on the left to get an equation which extrapolates from the primed medium to the unprimed medium. We get

$$\begin{aligned} -tU \qquad &= -U' + c'D' \\ -cU + D &= \qquad t'D' \end{aligned}$$

Arrange to matrix form

$$\begin{bmatrix} -t & 0 \\ -c & 1 \end{bmatrix} \begin{bmatrix} U \\ D \end{bmatrix} = \begin{bmatrix} -1 & c' \\ 0 & t' \end{bmatrix} \begin{bmatrix} U \\ D \end{bmatrix}'$$

Premultiply by the inverse of the left-hand matrix

$$\begin{aligned} \begin{bmatrix} U \\ D \end{bmatrix} &= \frac{1}{-t} \begin{bmatrix} 1 & 0 \\ c & -t \end{bmatrix} \begin{bmatrix} -1 & c' \\ 0 & t' \end{bmatrix} \begin{bmatrix} U \\ D \end{bmatrix}' \\ &= \frac{1}{-t} \begin{bmatrix} -1 & c' \\ -c & cc' - tt' \end{bmatrix} \begin{bmatrix} U \\ D \end{bmatrix}' \end{aligned}$$

getting a result we want, an equation to extrapolate the infinitesimal distance from the bottom of one layer to the top of the next.

$$\begin{bmatrix} U \\ D \end{bmatrix} = \frac{1}{t}\begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix}\begin{bmatrix} U \\ D \end{bmatrix}' \tag{5.12}$$

Now we consider a layered medium where each layer has the same travel-time thickness [**Goupillaud** medium]. Presumably, any continuously variable medium can be approximated this way. Examine the definitions in the layered medium in Figure 5.6. For this arrangement

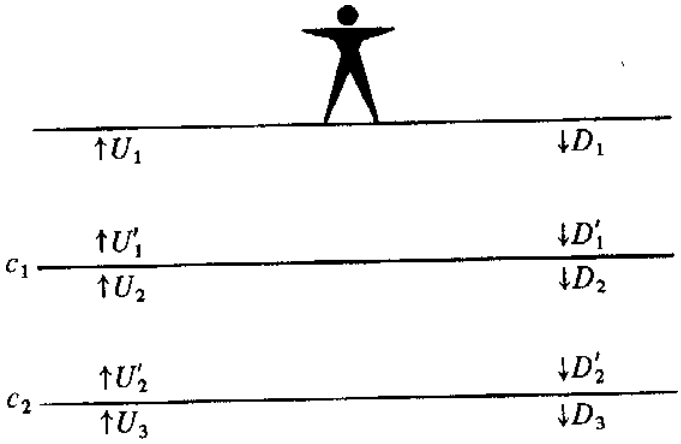

Figure 5.6: Goupillaud-type layered medium (layers have equal travel time). fgdp8-8-6 [NR]

of layers, (5.12) may be written

$$\begin{bmatrix} U \\ D \end{bmatrix}_{k+1} = \frac{1}{t_k}\begin{bmatrix} 1 & c_k \\ c_k & 1 \end{bmatrix}\begin{bmatrix} U \\ D \end{bmatrix}_k'$$

Let $Z = e^{i\omega T}$ where T, the two-way travel time, equals the data sampling interval. Clearly, multiplication by $\sqrt{Z}$ is equivalent to delaying a function by $T/2$, the travel time across a layer. This gives in the kth layer a relation between primed and unprimed waves.

$$\begin{bmatrix} U \\ D \end{bmatrix}_k' = \begin{bmatrix} 1/\sqrt{Z} & 0 \\ 0 & \sqrt{Z} \end{bmatrix}\begin{bmatrix} U \\ D \end{bmatrix}_k \tag{5.13}$$

Inserting (5.13) into (5.12) we get a ***layer matrix*** that takes us from the top of one layer to the top of the next deeper one.

$$\begin{aligned} \begin{bmatrix} U \\ D \end{bmatrix}_{k+1} &= \frac{1}{t_k}\begin{bmatrix} 1 & c_k \\ c_k & 1 \end{bmatrix}\begin{bmatrix} 1/\sqrt{Z} & 0 \\ 0 & \sqrt{Z} \end{bmatrix}\begin{bmatrix} U \\ D \end{bmatrix}_k \\ &= \frac{1}{t_k}\begin{bmatrix} 1/\sqrt{Z} & c_k\sqrt{Z} \\ c_k/\sqrt{Z} & \sqrt{Z} \end{bmatrix}\begin{bmatrix} U \\ D \end{bmatrix}_k \\ \begin{bmatrix} U \\ D \end{bmatrix}_{k+1} &= \frac{1}{\sqrt{Z}t_k}\begin{bmatrix} 1 & c_k Z \\ c_k & Z \end{bmatrix}\begin{bmatrix} U \\ D \end{bmatrix}_k \end{aligned} \tag{5.14}$$

5.3 ENERGY FLUX

If there is energy flowing through a stack of layers, there should be the same **energy flow** through the kth layer as through the $(k+1)$st layer. Otherwise, there would be an energy sink or source at the layer boundary. Let us guess that the net upward flow of energy (**energy flux**) at any frequency ω in the $(k+1)$st layer is given by

$$\text{flux}(\omega) \quad = \quad Y_k \left[U(Z)\bar{U}\left(\frac{1}{Z}\right) - D(Z)\bar{D}\left(\frac{1}{Z}\right) \right]_{k+1} \tag{5.15}$$

Inserting equation (5.14) into equation (5.15) (and using a few simplifications all spelled out in **FGDP**) we deduce the principle that the energy flux across any layer matches that across the neighboring layer.

$$Y_{k+1}\left[U\left(\frac{1}{Z}\right)U(Z) - D\left(\frac{1}{Z}\right)D(Z) \right]_{k+1} \quad = \quad Y_k \left[U\left(\frac{1}{Z}\right)U(Z) - D\left(\frac{1}{Z}\right)D(Z) \right]_{k} \tag{5.16}$$

Since the flux is the same in any two adjacent layers, it must be the same in all the layers. This energy flux theorem leads quickly to some sweeping statements about the waveforms scattered from layered structures.

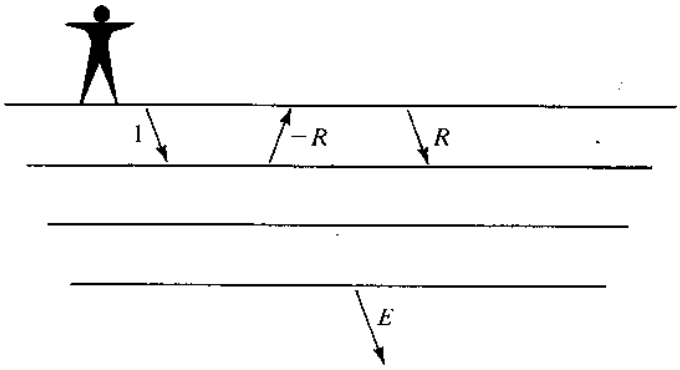

Figure 5.7: The reflection seismology geometry. The person sends an impulse down. The earth sends back $-R(Z)$. Since the surface is perfectly reflective, it sends $R(Z)$ back down. Escaping from the bottom is a wave $E(Z)$ heading to infinity. fgdp8-8-7 [NR]

Figure 5.7 shows the basic geometry of reflection seismology. Applying the energy flux theorem to this geometry we may say that the energy flux in the top layer equals that in the lower halfspace so

$$Y_1 \left\{ R\left(\frac{1}{Z}\right)R(Z) - \left[1 + R\left(\frac{1}{Z}\right)\right][1+R(Z)] \right\} \quad = \quad -Y_k E\left(\frac{1}{Z}\right)E(Z)$$

or

$$1 + R\left(\frac{1}{Z}\right) + R(Z) \quad = \quad \frac{Y_k}{Y_1}\, E\left(\frac{1}{Z}\right)E(Z) \tag{5.17}$$

This result is amazing. Nothing in life led me to suspect it before I found it. It says, if we were to observe the escaping wave $E(Z)$, we could by **autocorrelation** construct the waveform seen at the surface. This result has no practical use, but a slightly different arrangement does.

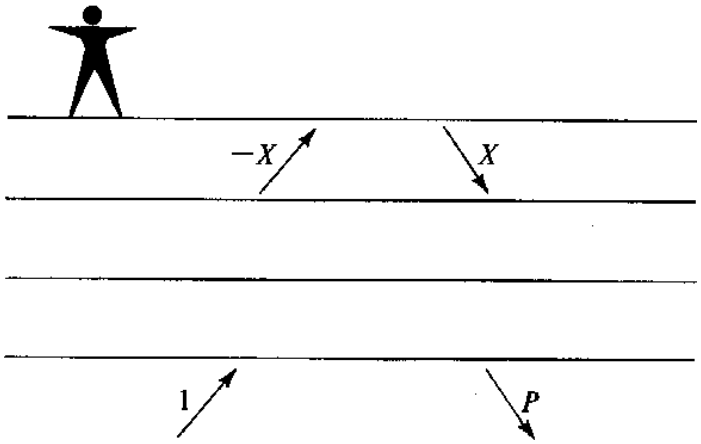

Figure 5.8: Earthquake seismology geometry. An impulse 1 is incident from below. The waveform $X(Z)$ is incident upon the free surface and is reflected back down. The waveform $P(Z)$ scatters back into earth. fgdp8-E8-3-5 [NR]

The escaping wave $E(Z)$ turns out to be the same as the earthquake wave $X(Z)$ in Figure 5.8. This can be shown algebraically (see exercises) or deduced as a consequence of the reciprocity principle (the idea that interchanging source and receiver location yields the same response). Notice that the incident earthquake "1" need not be an impulsive wave; it only need have the **autocorrelation** of an impulsive wave. This conclusion leads to the remarkable hypothesis that the autocorrelation of **natural** noises (microseism) gives reflection seismograms. This concept has proven useful in **solar seismology**, but a variety of practical factors (principally spatial distribution of stations) has limited its utility on the earth so far.

A more obvious result applies to the earthquake seismology geometry in Figure 5.8. Applying the energy flux theorem to this geometry we obtain

$$Y_1\left[X\left(\frac{1}{Z}\right)X(Z) - X\left(\frac{1}{Z}\right)X(Z)\right] \quad = \quad Y_k\left[1 - P(Z)P\left(\frac{1}{Z}\right)\right]$$

or

$$1 \quad = \quad P(Z)P\left(\frac{1}{Z}\right) \tag{5.18}$$

The interpretation of the result is that the backscattered waveform $P(Z)$ has the form of an all-pass filter. (An all-pass filter has a constant energy spectrum and an arbitrary phase spectrum.) This result may have been anticipated on physical grounds since all the energy which is incident is ultimately reflected without attenuation; thus the only thing which can happen is that there will be frequency-dependent delay.

5.4 GETTING THE WAVES FROM THE REFLECTION COEFFICIENTS

An important result of the last section was the development of a "layer matrix" (5.14) that is, a matrix which can be used to extrapolate waves observed in one layer to the waves observed in the next layer. This process may be continued indefinitely. To see how to extrapolate from

layer 1 to layer 3 substitute (5.14) with $k = 1$ into (5.14) with $k = 2$, obtaining

$$\begin{aligned}\begin{bmatrix} U \\ D \end{bmatrix}_3 &= \frac{1}{\sqrt{Z}t_2}\begin{bmatrix} 1 & Zc_2 \\ c_2 & Z \end{bmatrix}\frac{1}{\sqrt{Z}t_1}\begin{bmatrix} 1 & Zc_1 \\ c_1 & Z \end{bmatrix}\begin{bmatrix} U \\ D \end{bmatrix}_1 \\ &= \frac{1}{\sqrt{Z}^2 t_1 t_2}\begin{bmatrix} 1+Zc_1c_2 & Zc_1+Z^2c_2 \\ c_2+Zc_1 & Zc_1c_2+Z^2 \end{bmatrix}\begin{bmatrix} U \\ D \end{bmatrix}_1 \end{aligned} \tag{5.19}$$

Inspection of this product suggests the general form for a product of k layer matrices. We'll call it a **multilayer matrix**.

$$\begin{bmatrix} U \\ D \end{bmatrix}_{k+1} = \frac{1}{\sqrt{Z}^k \prod_{j=1}^{k} t_j}\begin{bmatrix} F(Z) & Z^k G\left(\frac{1}{Z}\right) \\ G(Z) & Z^k F\left(\frac{1}{Z}\right) \end{bmatrix}\begin{bmatrix} U \\ D \end{bmatrix}_1 \tag{5.20}$$

Starting from computer programs that add and multiply polynomials by manipulating their coefficients, it is easy enough to write a computer code to recursively build up the $F(Z)$ and $G(Z)$ polynomials by adding a layer at a time starting from $F(Z) = 1$ and $G(Z) = c_1$.

5.4.1 Layers between two halfspaces

Let a stack of layers be sandwiched in between two halfspaces Figure 5.9. An impulse is

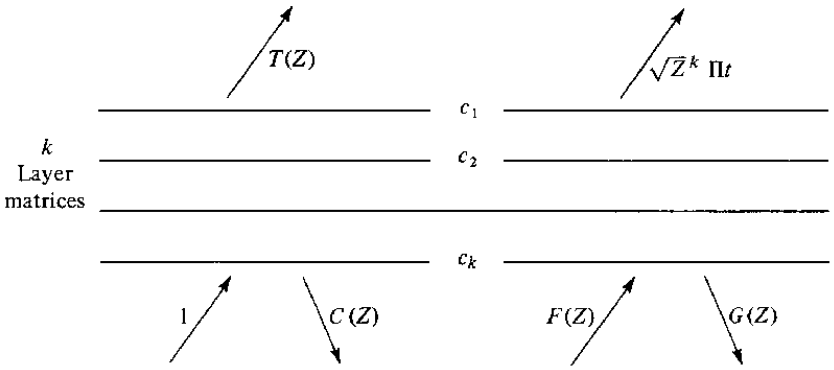

Figure 5.9: Waves incident, reflected, and transmitted from a stack of layers between two half-spaces. fgdp8-8-9 [NR]

incident from below. The backscattered wave is called $C(Z)$ and the transmitted wave is called $T(Z)$. We take these waves and load them into the **multilayer equation** (5.20).

$$\begin{bmatrix} 1 \\ C(Z) \end{bmatrix} = \frac{1}{\sqrt{Z}^k \Pi t}\begin{bmatrix} F(Z) & Z^k G\left(\frac{1}{Z}\right) \\ G(Z) & Z^k F\left(\frac{1}{Z}\right) \end{bmatrix}\begin{bmatrix} T(Z) \\ 0 \end{bmatrix} \tag{5.21}$$

We may solve the first of (5.21) for the transmitted wave $T(Z)$

$$T(Z) = \frac{\sqrt{Z}^k \Pi t}{F(Z)} \tag{5.22}$$

and introduce the result back into the second of (5.21) to obtain the backscattered wave

$$C(Z) \quad = \quad \frac{G(Z)T(Z)}{\sqrt{Z}^k \Pi t} \quad = \quad \frac{G(Z)}{F(Z)} \tag{5.23}$$

Let us be clear we understand the meaning of the solutions $T(Z)$ and $C(Z)$. They can be regarded two ways, either as functions of frequency $Z = e^{i\omega}$, or the polynomial coefficients (after dividing out the denominator $F(Z)$) as waveforms, as functions of time. Physical interpretations of $F(Z)$ and $G(Z)$ are on the right side of Figure 5.9.

5.4.2 Reflection seismology geometry

Identify the up and downgoing waves in the reflection seismology geometry in Figure 5.7 and load these waves into the multilayer matrix boundaries, equation (5.20).

$$\begin{bmatrix} 0 \\ E(Z) \end{bmatrix} \quad = \quad \frac{1}{\sqrt{Z}^k \Pi t} \begin{bmatrix} F(Z) & Z^k G\left(\frac{1}{Z}\right) \\ G(Z) & Z^k F\left(\frac{1}{Z}\right) \end{bmatrix} \begin{bmatrix} -R \\ 1+R \end{bmatrix} \tag{5.24}$$

From the first equation we may solve for $R(Z)$

$$R(Z) \quad = \quad +\frac{Z^k G(1/Z)}{F(Z) - Z^k G(1/Z)} \tag{5.25}$$

Given $R(Z)$, the second equation in the set (5.24) gives us $E(Z)$. Calculation of the wave fields on the earthquake geometry of Figure 5.8 is reserved for an exercise.

5.4.3 Getting reflection coefficients from the waves

The remainder of chapter 8 in **FGDP** shows how the algebraic techniques developed here can be solved in reverse. Given various waves such as the reflection data $R(Z)$ or the earthquake data $X(Z)$ we may deduce the reflection sequence $(c_1, c_2, c_3, \cdots)$. This appears to be the solution to the practical problem. It turns out however, for various complicated reasons, that this does not seem to be a methodology in widespread use in the seismic exploration business.

EXERCISES:

1 In Figure 5.9 let $c_1 = \frac{1}{2}$, $c_2 = -\frac{1}{2}$, and $c_3 = \frac{1}{3}$. What are the polynomial ratios $T(Z)$ and $C(Z)$?

2 For a simple interface, we had the simple relations $t = 1 + c$, $t' = 1 + c'$, and $c = -c'$. What sort of analogous relations can you find for the generalized interface of Figure 5.9? [For example, show $1 - T(Z)T'(1/Z) = C(Z)C(1/Z)$ which is analogous to $1 - tt' = c^2$.]

3 Show that $T(Z)$ and $T'(Z)$ are the same waveforms within a scale factor. Deduce that many different stacks of layers may have the same $T(Z)$.

4 Consider the earth to be modeled by layers over a halfspace. Let an impulse be incident from below (Figure 5.8). Given $F(Z)$ and $G(Z)$, elements of the product of the layer matrices, solve for X and for P. Check your answer by showing that $P(Z)\bar{P}(1/Z) = 1$. How is X related to E? This relation illustrates the principle of reciprocity which says source and receiver may be interchanged.

5 Show that $1 + R(1/Z) + R(Z) = (\text{scale factor})X(Z)X(1/Z)$, which shows that one may autocorrelate the transmission seismogram to get the reflection seismogram.

6 Refer to Figure 5.10. Calculate R' from R.

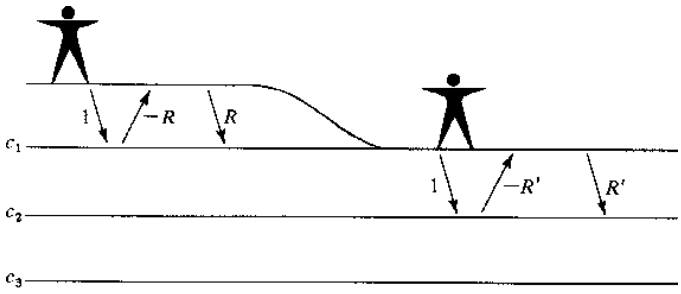

Figure 5.10: Stripping off the surface layer. fgdp8-E8-3-7 [NR]

Chapter 6

Multidimensional deconvolution examples

Deconvolution is a well-known process with deep and complex theoretical underpinnings. Here we'll begin with a superficial view of the theory and then jump to some examples in two-dimensional space.

One definition of science that distinguishes it from other views of the human condition is that science is those studies that offer demonstrable predictive power. That direction brings us too far afield. Never-the-less, science can be formalized as the construction of predicted data that can be compared with observed data. This requires a theory. Oft times, we really don't have a theory that we have much confidence in. Fortunately, there is an "all-purpose" theory of statistical signal processing that often allows us some predictability. Basically, it goes like this.

We have some data. From this data we can estimate a spectrum. By assuming the spectrum remains constant in time or space (stationary assumption), we have a certain predictive power. How much power this amounts to in practice depends on lots of things we won't examine in this brief introduction beyond saying that it depends a lot on both the observed data and the validity of the stationary assumption.

It often happens that the item of interest is not the prediction, but the error in the prediction. For example, we might be given the history of the stock market upon which to estimate the spectrum. Then we might be asked to predict (for each moment in time), the market price tomorrow based on today's price and all previous prices. The display that is often of interest is the error in prediction, did we overshoot or undershoot? Theoretically, it turns out that the prediction error is spectally white. If it had any other spectrum, that would imply suboptimum prediction. This leads to the simplest definition of prediction error, the definition we will adopt here. (The examples here come from one of my other books (GEE). These examples are based not on spectra, but on actual prediction error, but we will not examine the distinction, a distinction that is small compared to the essential ideas discussed here.)

6.1 2-D MODELING AND DECON

This chapter shows many examples to give some feeling for multidimensional spectra, deconvolution, and modeling. In each of Figures 6.2-6.7 we see three panels. The left panel is a two-dimensional data set which is an image of something that might be familiar. Let this data $d(x,y)$ have **Fourier Transform** $D(k_x,k_y)$.

Define noise $n(x,y)$ to be a plane of random numbers. In Fourier space, these numbers are $N(k_x,k_y)$. Now compute an amplitude spectrum by smoothing $A=\sqrt{\overline{N(k_x,k_y)}N(k_x,k_y)}$. Using this noise we compute the synthetic data in the center frame. It has the 2-D spectrum $N \times A$ which is the same 2-D *amplitude* spectrum of the given data but a different *phase* spectrum.

For each illustration, notice the similarities and differences between the real data and the synthetic data.

The rightmost frame has Fourier Transform D/A, the original data with its spectrum divided out. What we see is called "deconvolved data". It is also called "**prediction error**". Theoretically, its output is "white". Actually its spectrum is only approximately white because the spectrum is smoothed before it is divided out. (The name prediction error derives from a simple model that can be applied to all data. This simple model is that all data can be modeled as white noise into a filter where the filter has a spectrum that is a smoothed version of that of the data.)

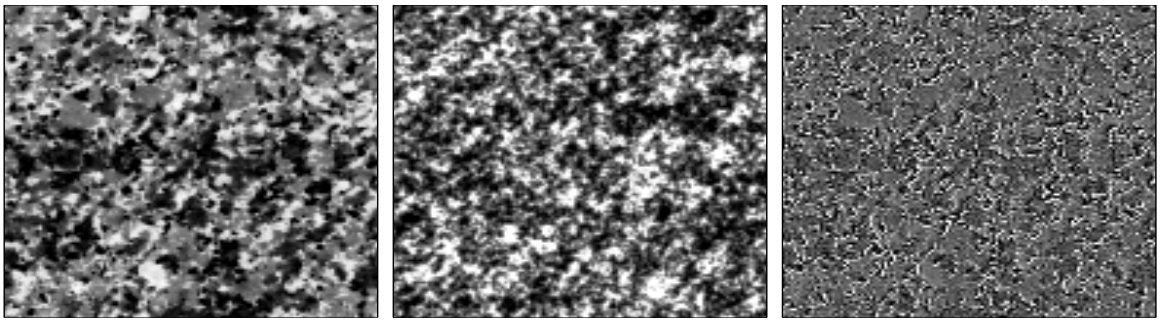

Figure 6.1: Synthetic granite matches the training image quite well. The prediction error is large at crystal grain boundaries and almost seems to outline the grains. mda-granite [NR]

6.2 DATA RESTORATION

Figures 6.8-6.11 demonstrate the success and lack of success of a process to replace missing data by data of the correct spectrum. Although real data is always better than synthetic data, especially synthetic data produced with a primitive theory involving only spectra, the fact is that synthetic data is often better than zero data.

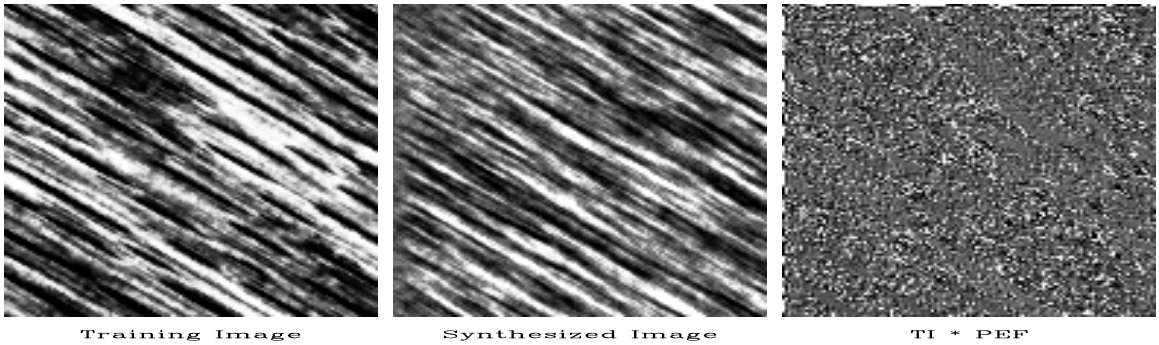

Figure 6.2: Synthetic wood grain has too little white. This is because of the nonsymmetric brightness histogram of natural wood. Again, the deconvolution looks random as expected. mda-wood [NR]

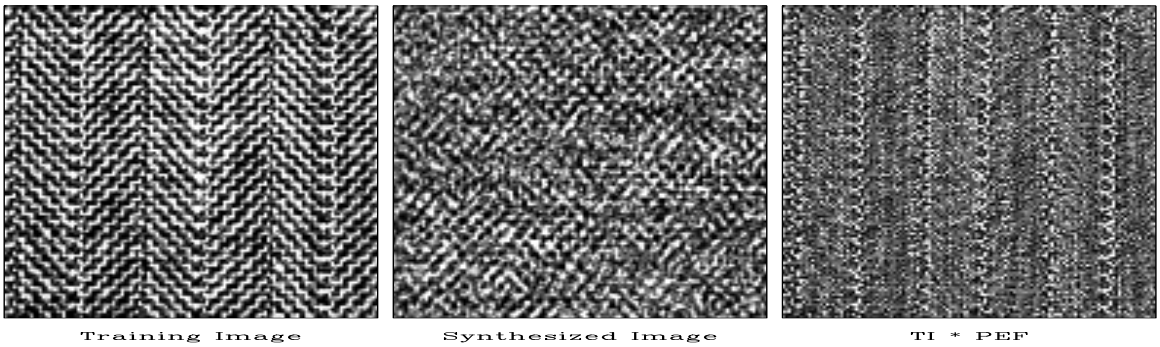

Figure 6.3: A banker's suit (left). A student's suit (center). My suit (right). The deconvolution is large where the weave changes direction (herring bone spine). mda-herr [NR]

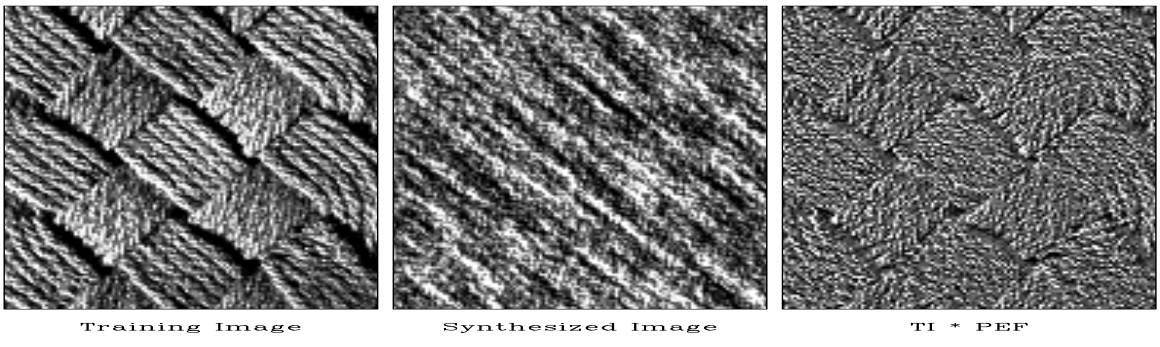

Figure 6.4: Basket weave. The synthetic data fails to segregate the two dips into a checkerboard pattern. The deconvolution looks structured. mda-basket [NR]

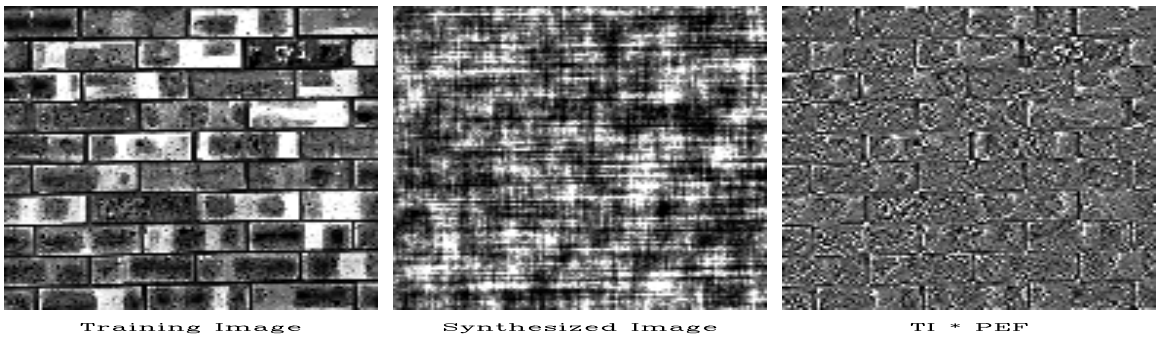

Figure 6.5: Brick. Synthetic brick edges are everywhere and do not enclose blocks containing a fixed color. PEF output highlights the mortar. mda-brick [NR]

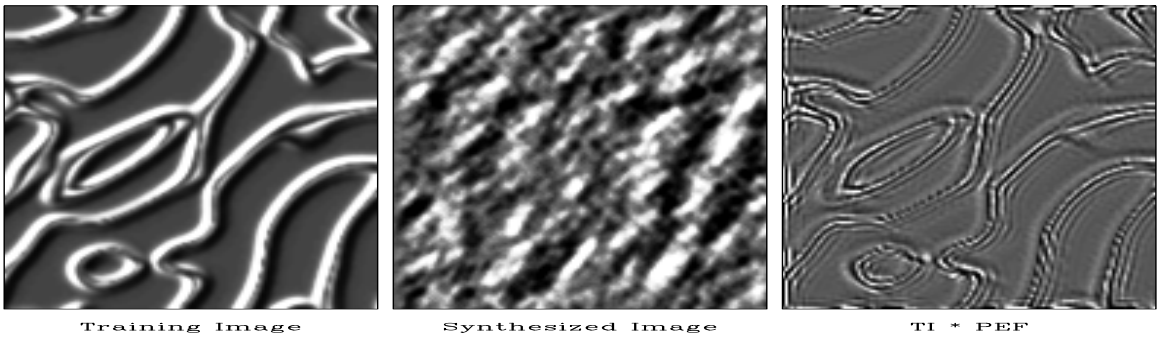

Figure 6.6: Ridges. A spectacular failure of the stationarity assumption. All dips are present but in different locations. The ridges have been sharpened by the deconvolution. mda-ridges [NR]

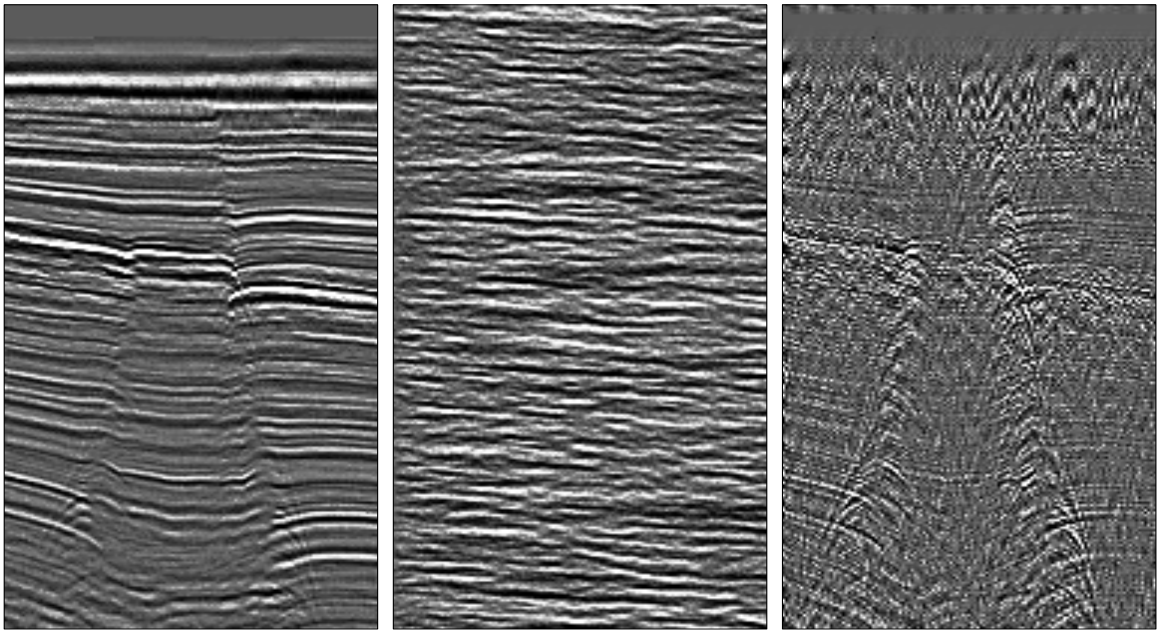

Figure 6.7: Gulf of Mexico seismic section, modeled, and deconvolved. Do you see any drilling prospects in the synthetic data? The deconvolution suppresses the strong horizontal layering giving a better view of the hyperbolas. mda-WGstack [NR]

In geophysical parameter fitting (inversion) you would not want to fit parameters to synthetic data, but in imaging processes like migration, synthetic data is better than zero data. This would be true for Figure 6.11.

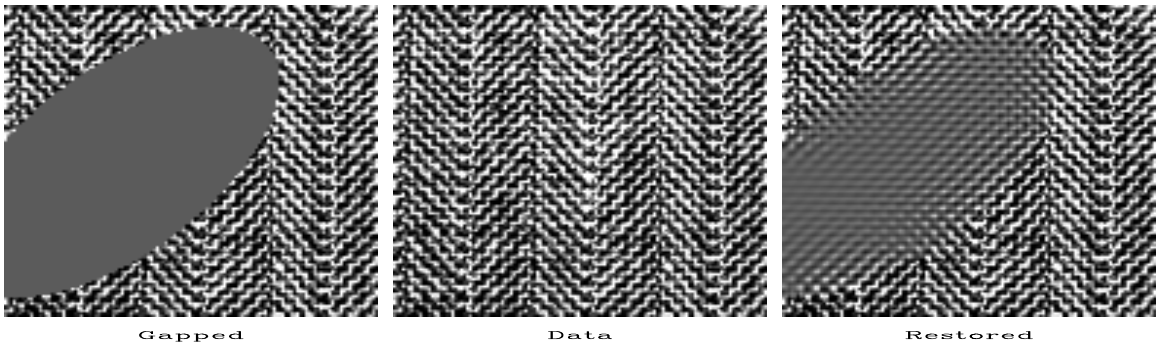

Figure 6.8: The herringbone texture is a patchwork of two textures. We notice that data missing from the hole tends to fill with the texture at the edge of the hole. The spine of the herring fish, however, is not modeled at all. mda-herr-hole-fill [NR,M]

Figure 6.9: The brick texture has a mortar part (both vertical and horizontal joins) and a brick surface part. These three parts enter the empty area but do not end where they should. mda-brick-hole-fill [NR,M]

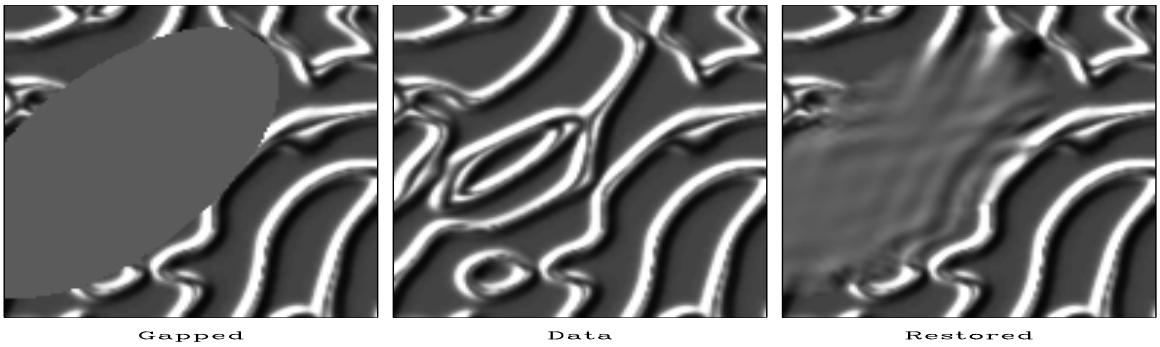

Figure 6.10: The theoretical model is a poor fit to the ridge data since the prediction must try to match ridges of all possible orientations. This data requires a broader theory which incorporates the possibility of nonstationarity (space variable slope). mda-ridges-hole-fill [NR,M]

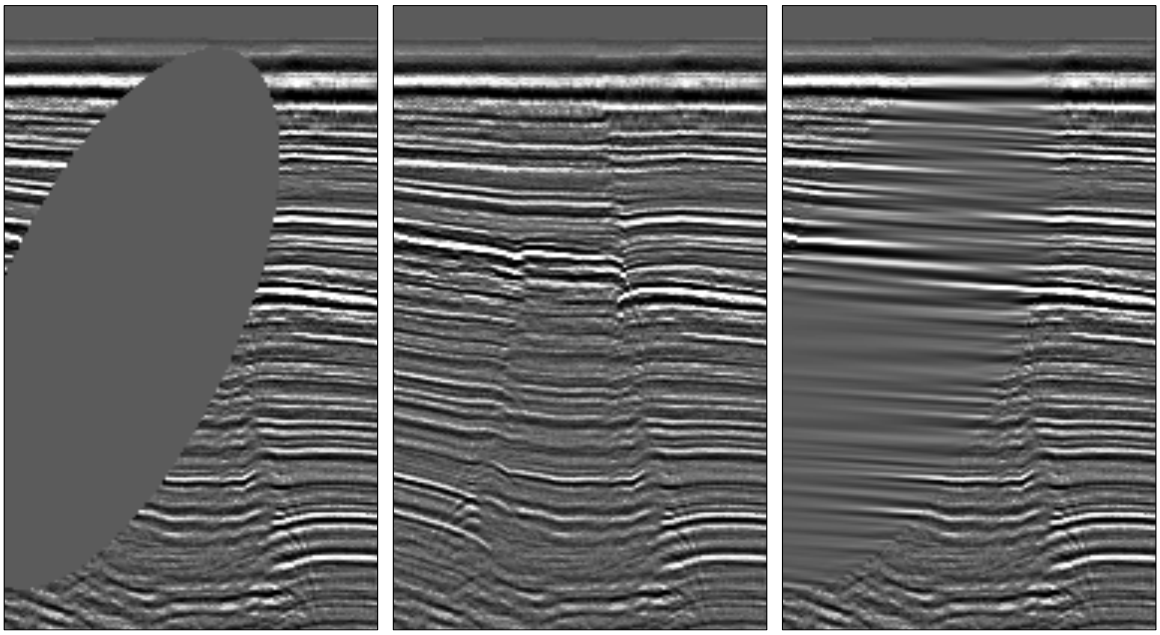

Figure 6.11: Filling the missing seismic data. The imaging process known as "migration" would suffer diffraction artifacts in the gapped data that it would not suffer on the restored data. mda-WGstack-hole-fill [NR,M]

Chapter 7

Resolution and random signals

[1]The accuracy of measurements on observed signals is limited not only by practical realities, but also by certain fundamental principles. The most famous example included in this chapter is the **time-bandwidth** product in Fourier-transformation theory, called the "**uncertainty principle**."

Observed signals often look random and are often modeled by filtered random numbers. In this chapter we will see many examples of signals built from random numbers and discover how the nomenclature of statistics applies to them. Fundamentally, this chapter characterizes "**resolution**," resolution of frequency and arrival time, and the statistical resolution of signal amplitude and power as functions of time and frequency.

We will see $\sqrt{n}$ popping up everywhere. This $\sqrt{n}$ enters our discussion when we look at spectra of signals built from random numbers. Also, signals that are theoretically uncorrelated generally appear to be weakly correlated at a level of $1/\sqrt{n}$, where n is the number of independent points in the signal.

Measures of resolution (which are variously called **variance**s, **tolerance**s, uncertainties, **bandwidth**s, **duration**s, **spread**s, **rise time**s, spans, etc.) often interact with one another, so that experimental change to reduce one must necessarily increase another or some combination of the others. In this chapter we examine basic cases where such conflicting interactions occur.

To avoid confusion I introduce the unusual notation Λ where Δ is commonly used. Notice that the letter Λ resembles the letter Δ, and Λ connotes length without being confused with wavelength. Lengths on the time and frequency axes are defined as follows:

dt, df	mesh intervals in time and frequency
$\Delta t, \Delta f$	mesh intervals in time and frequency
$\Delta T\ \Delta F$	extent of time and frequency axis
$\Lambda T, \Lambda F$	time duration and spectral **bandwidth** of a signal

[1]A more detailed and analytic version of this chapter is on the web at http://sep.stanford.edu/sep/prof/ in my book, Earth Soundings Analysis: Processing versus Inversion.

There is no mathematically tractable and universally acceptable definition for time span ΛT and **spectral bandwidth** ΛF. A variety of defining equations are easy to write, and many are in general use. The main idea is that the time span ΛT or the frequency span ΛF should be able to include most of the energy but need not contain it all. The time duration of a damped exponential function is infinite if by duration we mean the span of nonzero function values. However, for practical purposes the time span is generally defined as the time required for the amplitude to decay to e^{-1} of its original value. For many functions the span is defined by the span between points on the time or frequency axis where the curve (or its envelope) drops to half of the maximum value. Strange as it may sound, there are certain concepts about the behavior of ΛT and ΛF that seem appropriate for "all" mathematical choices of their definitions, yet these concepts can be proven only for special choices.

7.1 TIME-FREQUENCY RESOLUTION

A consequence of Fourier transforms being built from $e^{i\omega t}$ is that narrowing a function in one domain widens it in the other domain. Scaling ω implies inverse scaling of t to keep the product ωt constant. A pure sinusoidal wave has a clearly defined frequency, but it is spread over the infinitely long time axis. At the other extreme is an impulse function (often called a delta function), which is nicely compressed to a point on the time axis but contains a mixture of all frequencies. In this section we examine how the width of a function in one domain relates to that in the other. By the end of the section, we will formalize this into an inequality:

For any signal, the time duration ΛT and the spectral bandwidth ΛF are related by

$$\Lambda F \, \Lambda T \quad \geq \quad 1 \tag{7.1}$$

This **inequality** is the **uncertainty principle**.

Since we are unable to find a precise and convenient analysis for the definitions of ΛF and ΛT, the inequality (7.1) is not strictly true. What is important is that rough equality in (7.1) is observed for many simple functions, but for others the inequality can be extremely slack (far from equal). Strong *inequality* arises from **all-pass filter**s. A phase shifting (all-pass) filter leaves the spectrum unchanged, and hence ΛF unchanged, but it can spread out the signal arbitrarily, increasing ΛT arbitrarily. Some people say that the **Gaussian** function has the minimum product in (7.1), but that really depends on a particular method of measuring ΛF and ΛT.

7.1.1 A misinterpretation of the uncertainty principle

It is easy to misunderstand the uncertainty principle. An oversimplification of it is to say that it is "impossible to know the frequency at any particular time." This oversimplification leads us to think about a **truncated sinusoid**, such as in Figure 7.1. We know the frequency exactly, so ΛF seems zero, whereas ΛT is finite, and this seems to violate (7.1). But what the

figure shows is that the truncation of the sinusoid has broadened the frequency band. More particularly, the impulse function in the frequency domain has been convolved by the sinc function that is the Fourier transform of the truncating rectangle function.

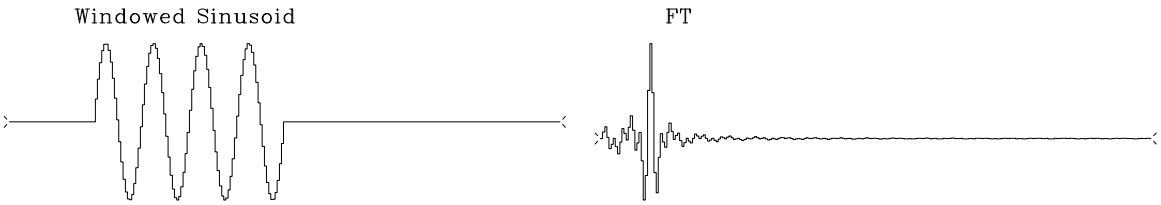

Figure 7.1: Windowed sinusoid and its Fourier transform. rnd-windcos [NR]

7.1.2 Measuring the time-bandwidth product

Now examine Figure 7.2, which contains sampled **Gaussian** functions and their Fourier transforms. The Fourier transform of a Gaussian is well known to be another Gaussian function, as the plot confirms. I adjusted the width of each Gaussian so that the widths would be about equal in both domains. The Gaussians were sampled at various values of n, increasing in steps by a factor of 4. You can measure the width dropping by a factor of 2 at each step. For those of you who have already learned about the uncertainty principle, it may seem paradoxical that the function's width is dropping in both time and frequency domains.

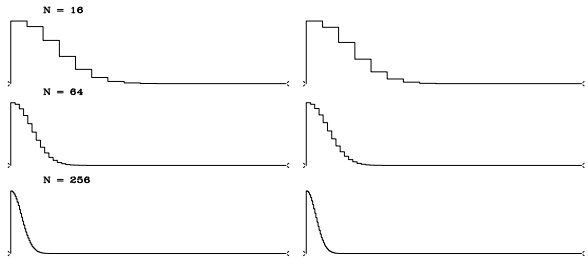

Figure 7.2: Sampled Gaussian functions and their Fourier transforms for vectors of length n = 16, 64, and 256. rnd-uncertain [NR]

The resolution of the paradox is that the physical length of the time axis or the frequency axis is varying as we change n (even though the plot length is scaled to a constant on the page). We need to associate a physical mesh with the computational mesh. In real physical space as well as in Fourier transform space, the object remains a constant size as the mesh is refined.

Let us read from Figure 7.2 values for the widths ΛF and ΛT. On the top row, where $N = 16$, I pick a width of about 4 points, and this seems to include about 90% of the area under the function. For this signal (with the widths roughly equal in both domains) it seems that $\Lambda T = \sqrt{N}dt$ and $\Lambda F = \sqrt{N}df$. Using the relation between dt and df found in equation (1.37), which says that $dt\,df = 1/N$, the product becomes $\Lambda T \Lambda F = 1$.

We could also confirm the inequality (7.1) by considering simple functions for which we know the analytic transforms—for example, an impulse function in time. Then $\Lambda T = dt$, and the Fourier transform occupies the entire frequency band from minus to plus the Nyquist frequency $\pm.5/dt$ Hz, i.e., $\Lambda F = 1/dt$. Thus again, the product is $\Lambda T \Lambda F = 1$.

7.1.3 The uncertainty principle in physics

The inequality (7.1) derives the name "**uncertainty principle**" from its interpretation in **quantum mechanics**. Observations of subatomic particles show they behave like waves with spatial frequency proportional to particle momentum. The classical laws of mechanics enable prediction of the future of a mechanical system by extrapolation from the currently known position and momentum. But because of the wave nature of matter, with momentum proportional to spatial frequency, such prediction requires simultaneous knowledge of both the location and the spatial frequency of the wave. This is impossible, as we see from (7.1); hence the word "uncertainty."

7.1.4 Gabor's proof of the uncertainty principle

Although it is easy to verify the uncertainty principle in many special cases, it is not easy to deduce it. The difficulty begins from finding a definition of the width of a function that leads to a tractable analysis. One possible definition uses a second moment; that is, ΛT is defined by

$$(\Lambda T)^2 \quad = \quad \frac{\int t^2 b(t)^2 \, dt}{\int b(t)^2 \, dt} \tag{7.2}$$

The spectral bandwidth ΛF is defined likewise. With these definitions, Dennis **Gabor** prepared a widely reproduced proof. I will omit his proof here; it is not an easy proof; it is widely available; and the definition (7.2) seems inappropriate for a function we often use, the **sinc** function, i.e., the FT of a step function. Since the sinc function drops off as t^{-1}, its width ΛT defined with (7.2) is infinity, which is unlike the more human measure of width, the distance to the first axis crossing.

7.1.5 My rise-time form of the uncertainty principle

A similar and possibly more basic concept than the product of time and frequency spreads is the relationship between spectral **bandwidth** and the "**rise time**" of a system-response function. The rise time ΛT of a system response is defined as follows: when we kick a physical system with an impulse function, it usually responds rapidly, rising to some maximum level, and then dropping off more slowly toward zero. The quantitative value of the rise time is generally, and somewhat arbitrarily, taken to be the span between the time of excitation and the time at which the system response is more than halfway up to its maximum.

"Tightness" (nearness to equality) in the inequality (7.1) is associated with minimum phase. "Slackness" (remoteness from equality) in the (7.1) would occur if a filter with an additional all-pass component were used. Slackness could also be caused by a decay time that is more rapid than the rise time, or by other combinations of rises and falls, such as random combinations. Physical systems generally respond rapidly compared to the rate at which they later decay. Focusing our attention on such systems, we can now seek to derive the inequality (7.1) applied to rise time and bandwidth.

The first step is to choose a definition for rise time. I have found a tractable definition of **rise time** to be

$$\frac{1}{\Lambda T} \quad = \quad \frac{\int_0^\infty \frac{1}{t} b(t)^2 \, dt}{\int_0^\infty b(t)^2 \, dt} \tag{7.3}$$

where $b(t)$ is the response function under consideration. Equation (7.3) defines ΛT by the first negative moment. Since this is unfamiliar, consider two examples. Taking $b(t)$ to be a step function, recognize that the numerator integral diverges, giving the desired $\Lambda T = 0$ rise time. As a further example, take $b(t)^2$ to grow linearly from zero to t_0 and then vanish. Then the rise time ΛT is $t_0/2$, again reasonable. It is curious that $b(t)$ could grow as $\sqrt{t}$, which rises with infinite slope at $t = 0$, and not cause ΛT to be pushed to zero.

Proof

We omit the details of the proof of the rise time form of the uncertainty principle which are found in PVI and mention instead a few insights found along with the proof. The “slack” in the uncertainty principle relates to comparing the duration of a signal ΛT to the duration of its autocorrelation ΛT_{auto}. Generally speaking, it is easy to find a long signal that has short autocorrelation. Just take an arbitrary short signal and convolve it using a lengthy all-pass filter. Conversely, we cannot get a long autocorrelation function from a short signal. A good example is the autocorrelation of a rectangle function, which is a triangle. The triangle appears to be twice as long, but considering that the triangle tapers down, it is reasonable to assert that the ΛT’s are the same. Thus, we conclude that

$$\Lambda T_{\text{auto}} \quad \leq \quad \Lambda T \tag{7.4}$$

From this inequality may be proven the inequality in the uncertainty relation

$$\Lambda T \; \Lambda F \quad \geq \quad 1 \tag{7.5}$$

7.2 FT OF RANDOM NUMBERS

Many real signals are complicated and barely comprehensible. In experimental work, we commonly transform such data. To better understand what this means, it will be worthwhile to examine signals made from **random** numbers.

Figure 7.3 shows discrete Fourier transforms of random numbers. The basic conclusion to be drawn from this figure is that transforms of random numbers look like more random numbers. A random series containing all frequencies is called a “white-noise" series, because the color white is made from roughly equal amounts of all colors. Any series made by independently chosen random numbers is said to be an “independent" series. An independent series must be white, but a white series need not be independent. Figure 7.4 shows Fourier transforms of random numbers surrounded by zeros (or zero padded). Since all the vectors of random numbers are the same length (each has 1024 points, including both sides of the even

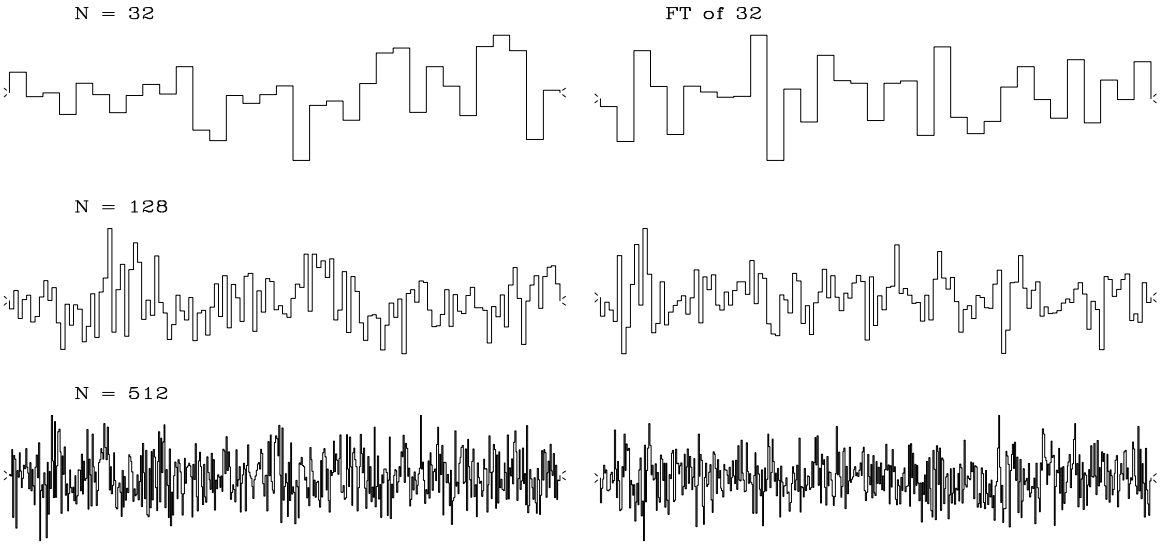

Figure 7.3: Fourier cosine transforms of vectors containing random numbers. N is the number of components in the vector. rnd-nrand [NR]

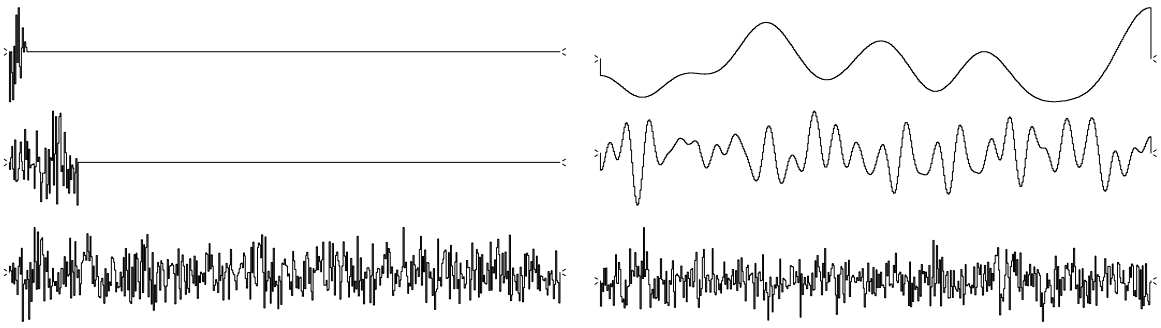

Figure 7.4: Zero-padded random numbers and their FTs. rnd-pad [NR]

function with the even part (513 points) shown), the transforms are also the same length. The top signal has less randomness than the second trace (16 random numbers versus 64). Thus the top FT is smoother than the lower ones. Although I introduced this figure as if the left panel were the time domain and the right panel were frequency, you are free to think of it the opposite way. This is more clear. With the left-hand signal being a frequency function, where higher frequencies are present, the right-hand signal oscillates faster.

7.2.1 Bandlimited noise

Figure 7.5 shows bursts of 25 random numbers at various shifts, and their Fourier transforms. You can think of either side of the figure as the time domain and the other side as the frequency domain. I like to think of the left side as the Fourier domain and the right side as the signals. Then the signals seem to be sinusoids of a constant frequency (called the “center" frequency) and of an amplitude that is modulated at a slower rate (called the “**beat**” frequency). Observe that the center frequency is related to the *location* of the random bursts, and that the beat frequency is related to the *bandwidth* of the noise burst.

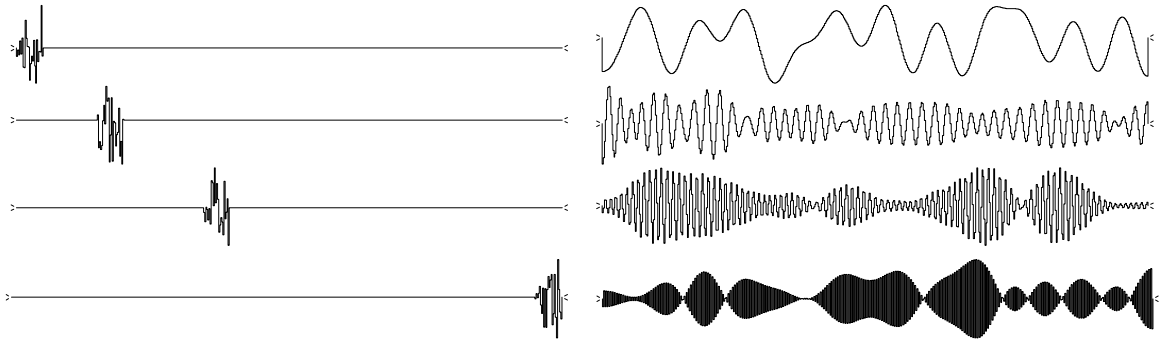

Figure 7.5: Shifted, zero-padded random numbers in bursts of 25 numbers. Left and right related by cosine transform. rnd-shift [NR]

You can also think of Figure 7.5 as having one-sided frequency functions on the left, and the right side as being the *real part* of the signal. The real parts are cosinelike, whereas the imaginary parts (not shown) are sinelike and have the same envelope function as the cosinelike part.

You might have noticed that the bottom plot in Figure 7.5, which has Nyquist-frequency modulated beats, seems to have about twice as many beats as the two plots above it. This can be explained as an end effect. The noise burst near the **Nyquist** frequency is really twice as wide as shown, because it is mirrored about the Nyquist frequency into negative frequencies. Likewise, the top figure is not modulated at all, but the signal itself has a frequency that matches the beats on the bottom figure.

7.3 TIME-STATISTICAL RESOLUTION

If we flipped a fair coin 1000 times, it is unlikely that we would get exactly 500 heads and 500 tails. More likely the number of heads would lie somewhere between 400 and 600. Or would it lie in another range? The theoretical value, called the "**mean**" or the "**expectation**," is 500. The value from our experiment in actually flipping a fair coin is called the "**sample mean**." How much difference Λm should we expect between the sample mean $\hat{m}$ and the true mean m? Both the coin flips x and our sample mean $\hat{m}$ are *random variables*. Our 1000-flip experiment could be repeated many times and would typically give a different $\hat{m}$ each time. This concept will be formalized in section 11.3.5. as the "**variance of the sample mean**," which is the expected squared difference between the true mean and the mean of our sample.

The problem of estimating the **statistic**al parameters of a time series, such as its mean, also appears in seismic processing. Effectively, we deal with seismic traces of finite duration, extracted from infinite sequences whose parameters can only be estimated from the finite set of values available in these seismic traces. Since the knowledge of these parameters, such as signal-to-noise ratio, can play an important role during the processing, it can be useful not only to estimate them, but also to have an idea of the error made in this estimation.

7.3.1 Ensemble

The "true value" of the mean could be defined as the mean that results when the coin is flipped n times, when n is conceived of as going to infinity. A more convenient definition of true value is that the experiment is imagined as having been done separately under identical conditions by an infinite number of people (an "**ensemble**"). The ensemble may seem a strange construction; nonetheless, much literature in statistics and the natural sciences uses the ensemble idea. Let us say that the ensemble is defined by a probability as a function of time. Then the ensemble idea enables us to define a time-variable mean (the sum of the values found by the ensemble weighted by the probabilities) for, for example, coins that change with time.

7.3.2 Expectation and variance

A conceptual average over the ensemble, or **expectation**, is denoted by the symbol E. The index for summation over the ensemble is never shown explicitly; every random variable is presumed to have one. Thus, the true mean at time t is defined as $m_x(t) = \mathrm{E}(x_t)$. The mean can vary with time:

$$m_x(t) \quad = \quad \mathrm{E}[x(t)] \tag{7.6}$$

The "**variance**" σ^2 is defined to be the power after the mean is removed, i.e.,

$$\sigma_x(t)^2 \quad = \quad \mathrm{E}\left[(x(t) - m_x(t))^2\right] \tag{7.7}$$

(Conventionally, σ^2 is referred to as the variance, and σ is called the "**standard deviation**.")

For notational convenience, it is customary to write $m(t)$, $\sigma(t)$, and $x(t)$ simply as m, σ, and x_t, using the verbal context to specify whether m and σ are time-variable or constant. For example, the standard deviation of the seismic amplitudes on a seismic trace before correction of spherical divergence decreases with time, since these amplitudes are expected to be "globally" smaller as time goes on.

When manipulating algebraic expressions, remember that the symbol E behaves like a summation sign, namely,

$$\mathrm{E} \quad \equiv \quad (\lim N \to \infty) \quad \frac{1}{N} \sum_{1}^{N} \tag{7.8}$$

Note that the summation index is not given, since the sum is over the ensemble, not time. To get some practice with the expectation symbol E, we can reduce equation (7.7):

$$\sigma_x^2 \quad = \quad \mathrm{E}[(x_t - m_x)^2] \quad = \quad \mathrm{E}(x_t^2) \; - \; 2m_x \mathrm{E}(x_t) + m_x^2 \quad = \quad \mathrm{E}(x_t^2) \; - \; m_x^2 \tag{7.9}$$

Equation (7.9) says that the energy is the variance plus the squared mean.

7.3.3 Probability and independence

A random variable x can be described by a **probability** $p(x)$ that the amplitude x will be drawn. In real life we almost never know the probability function, but theoretically, if we do know it, we can compute the **mean** value using

$$m \quad = \quad E(x) \quad = \quad \int x \, p(x) \, dx \tag{7.10}$$

"**Statistical independence**" is a property of two or more random numbers. It means the samples are drawn independently, so they are unrelated to each other. In terms of probability functions, the independence of random variables x and y is expressed by

$$p(x,y) \quad = \quad p(x) \, p(y) \tag{7.11}$$

From these, it is easy to show that

$$\mathrm{E}(xy) \quad = \quad \mathrm{E}(x)\mathrm{E}(y) \tag{7.12}$$

7.3.4 Sample mean

Now let x_t be a time series made up of identically distributed random numbers: m_x and σ_x do not depend on time. Let us also suppose that they are *independently* chosen; this means in particular that for any different times t and s ($t \neq s$):

$$\mathrm{E}(x_t x_s) \quad = \quad \mathrm{E}(x_t)\mathrm{E}(x_s) \tag{7.13}$$

Suppose we have a sample of n points of x_t and are trying to determine the value of m_x. We could make an estimate $\hat{m}_x$ of the mean m_x with the formula

$$\hat{m}_x = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{7.14}$$

A somewhat more elaborate method of estimating the mean would be to take a weighted average. Let w_t define a set of weights normalized so that

$$\sum w_t = 1 \tag{7.15}$$

With these weights, the more elaborate estimate $\hat{m}$ of the mean is

$$\hat{m}_x = \sum w_t x_t \tag{7.16}$$

Actually (7.14) is just a special case of (7.16); in (7.14) the weights are $w_t = 1/n$.

Further, the weights could be *convolved* on the random time series, to compute *local* averages of this time series, thus smoothing it. The weights are simply a filter response where the filter coefficients happen to be positive and cluster together. Figure 7.6 shows an example: a **random walk** function with itself smoothed locally.

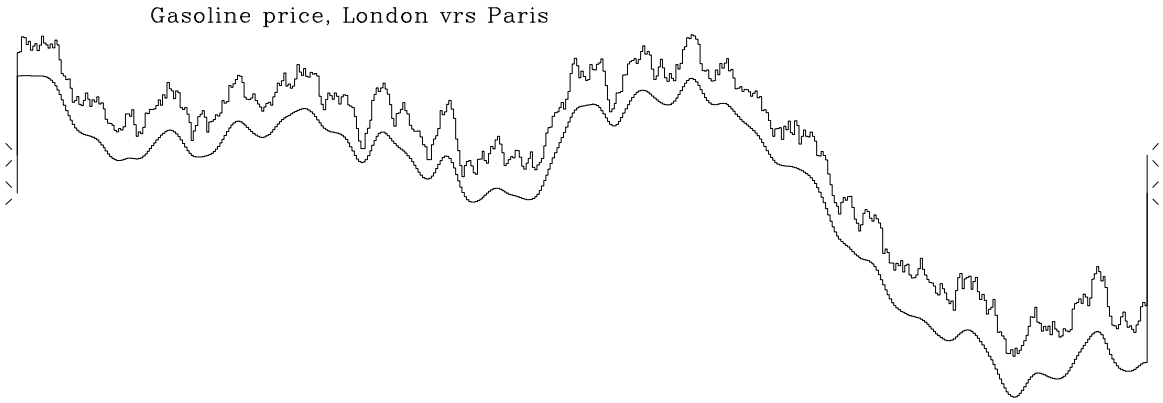

Figure 7.6: Random walk and itself smoothed (and shifted downward). rnd-walk [NR]

7.3.5 Variance of the sample mean

Ordinarily I like to squeeze all the equations out of my books, but I include them here because they show why random fluctions so often involve the square root of the number of variables. When you hear, for example, that a survey of the voters has a 3% accuracy, you can deduce that the surveyors spoke with about $1000 = 1/.03^2$ voters.

Formally now, we calculate how far the estimated mean $\hat{m}$ is likely to be from the true mean m for a sample of length n. This difference is the **variance of the sample mean** and is

given by $(\Lambda m)^2 = \sigma_{\hat{m}}^2$, where

$$\sigma_{\hat{m}}^2 = \mathrm{E}[(\hat{m}-m)^2] \tag{7.17}$$

$$= \mathrm{E}\left\{[(\sum w_t x_t)-m]^2\right\} \tag{7.18}$$

Now use the fact that $m = m\sum w_t = \sum w_t m$:

$$\sigma_{\hat{m}}^2 = \mathrm{E}\left\{\left[\sum_t w_t(x_t-m)\right]^2\right\} \tag{7.19}$$

$$= \mathrm{E}\left\{\left[\sum_t w_t(x_t-m)\right]\left[\sum_s w_s(x_s-m)\right]\right\} \tag{7.20}$$

$$= \mathrm{E}\left[\sum_t\sum_s w_t w_s(x_t-m)(x_s-m)\right] \tag{7.21}$$

The step from (7.20) to (7.21) follows because

$$(a_1+a_2+a_3)(a_1+a_2+a_3) = \text{sum of } \begin{bmatrix} a_1a_1 & a_1a_2 & a_1a_3 \\ a_2a_1 & a_2a_2 & a_2a_3 \\ a_3a_1 & a_3a_2 & a_3a_3 \end{bmatrix} \tag{7.22}$$

The expectation symbol E can be regarded as another summation, which can be done after, as well as before, the sums on t and s, so

$$\sigma_{\hat{m}}^2 = \sum_t\sum_s w_t\, w_s\, \mathrm{E}\,[(x_t-m)(x_s-m)] \tag{7.23}$$

If $t \neq s$, since x_t and x_s are independent of each other, the expectation $\mathrm{E}[(x_t-m)(x_s-m)]$ will vanish. If $s = t$, then the expectation is the variance defined by (7.7). Expressing the result in terms of the Kronecker delta, δ_{ts} (which equals unity if $t = s$, and vanishes otherwise) gives

$$\sigma_{\hat{m}}^2 = \sum_t\sum_s w_t\, w_s\, \sigma_x^2 \delta_{ts} \tag{7.24}$$

$$\sigma_{\hat{m}}^2 = \sum_t w_t^2\, \sigma_x^2 \tag{7.25}$$

$$\sigma_{\hat{m}} = \sigma_x \sqrt{\sum_t w_t^2} \tag{7.26}$$

For n weights, each of size $1/n$, the **standard deviation of the sample mean** is

$$\Lambda m_x = \sigma_{\hat{m}_x} = \sigma_x \sqrt{\sum_{t=1}^{n}\left(\frac{1}{n}\right)^2} = \frac{\sigma_x}{\sqrt{n}} \tag{7.27}$$

This is the most important property of random numbers that is not intuitively obvious. Informally, the result (7.27) says this: given a sum y of terms with random polarity, whose

theoretical mean is zero, then

$$y \quad = \quad \underbrace{\pm 1 \pm 1 \pm 1 \cdots}_{n \text{ terms}} \tag{7.28}$$

The sum y is a random variable whose standard deviation is $\sigma_y = \sqrt{n} = \Lambda y$. An experimenter who does not know the mean is zero will report that the mean of y is $\mathrm{E}(y) = \hat{y} \pm \Lambda y$, where $\hat{y}$ is the experimental value.

If we are trying to estimate the mean of a random series that has a time-variable mean, then we face a basic dilemma. Including many numbers in the sum in order to make Λm small conflicts with the possibility of seeing m_t change during the measurement.

The "**variance of the sample variance**" arises in many contexts. Suppose we want to measure the storminess of the ocean. We measure water level as a function of time and subtract the mean. The storminess is the variance about the mean. We measure the storminess in one minute and call it a sample storminess. We compare it to other minutes and other locations and we find that they are not all the same. To characterize these differences, we need the *variance of the sample variance* $\sigma^2_{\hat{\sigma}^2}$. Some of these quantities can be computed theoretically, but the computations become very cluttered and dependent on assumptions that may not be valid in practice, such as that the random variables are independently drawn and that they have a Gaussian probability function. Since we have such powerful computers, we might be better off ignoring the theory and remembering the basic principle that a function of random numbers is also a random number. We can use simulation to estimate the function's mean and variance. Basically we are always faced with the same dilemma: if we want to have an accurate estimation of the variance, we need a large number of samples, which limits the possibility of measuring a time-varying variance.

EXERCISES:

1 Suppose the mean of a sample of random numbers is estimated by a triangle weighting function, i.e.,

$$\hat{m} \quad = \quad s \sum_{i=0}^{n} (n-i)\, x_i$$

Find the scale factor s so that $\mathrm{E}(\hat{m}) = m$. Calculate Λm. Define a reasonable ΛT. Examine the uncertainty relation.

2 A random series x_t with a possibly time-variable mean may have the mean estimated by the feedback equation

$$\hat{m}_t \quad = \quad (1-\epsilon)\hat{m}_{t-1} + b x_t$$

a. Express $\hat{m}_t$ as a function of $x_t, x_{t-1}, \ldots$, and not $\hat{m}_{t-1}$.

b. What is ΛT, the effective averaging time?

c. Find the scale factor b so that if $m_t = m$, then $\mathrm{E}(\hat{m}_t) = m$.

d. Compute the random error $\Lambda m = \sqrt{\mathrm{E}(\hat{m} - m)^2}$. (HINT: Λm goes to $\sigma\sqrt{\epsilon/2}$ as $\epsilon \to 0$.)

e. What is $(\Lambda m)^2 \Lambda T$ in this case?

7.4 SPECTRAL FLUCTUATIONS

Recall the basic model of time-series analysis, namely, random numbers passing through a filter. A sample of input, filter, and output amplitude spectra is shown in Figure 7.7. From the

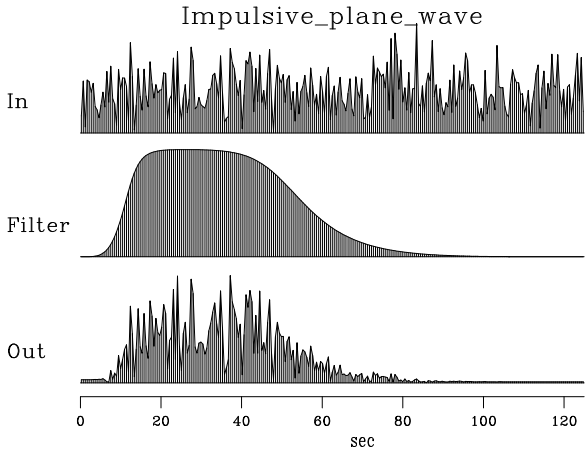

Figure 7.7: Random numbers into a filter. Top is a spectrum of random numbers. Middle is the spectrum of a filter. Bottom is the spectrum of filter output. rnd-filter [NR]

spectrum of the output we can guess the spectrum of the filter, but the figure shows there are some limitations in our ability to do so. Let us analyze this formally.

Observations of sea level over a long period of time can be summarized in terms of a few statistical averages, such as the mean height m and the variance σ^2. Another important kind of statistical average for use on geophysical time series is the "**power spectrum**." Many mathematical models explain only statistical averages of data and not the data itself. To recognize certain **pitfall**s and understand certain fundamental limitations on work with power spectra, we first consider the idealized example of random numbers.

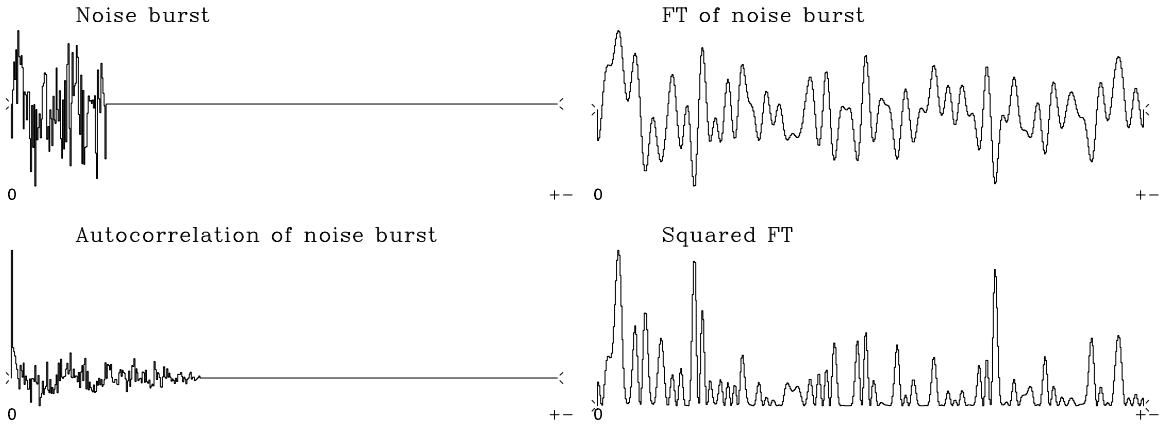

Figure 7.8: Autocorrelation and spectra of random numbers. rnd-auto [NR]

Figure 7.8 shows a signal that is a burst of noise; its Fourier transform, and the transform squared; and its inverse transform, the autocorrelation. Here the FT squared is the same as the more usual FT times its complex conjugate—because the noise-burst signal is even, its FT is real.

Notice that the **autocorrelation** has a big spike at zero lag. This spike represents the correlation of the random numbers with themselves. The other lags are much smaller. They represent the correlation of the noise burst with itself shifted. Theoretically, the noise burst is *not* correlated with itself shifted: these small fluctuations result from the finite extent of the noise sample.

Imagine many copies of Figure 7.8. Ensemble averaging would amount to adding these other autocorrelations or, equivalently, adding these other spectra. The fluctuations aside the central lobe of the autocorrelation would be destroyed by ensemble averaging, and the fluctuations in the spectrum would be smoothed out. The **expectation of the autocorrelation** is that it is an impulse at zero lag. The **expectation of the spectrum** is that it is a constant, namely,

$$\mathrm{E}[\hat{S}(Z)] \quad = \quad S(Z) \quad = \quad \text{const} \tag{7.29}$$

7.4.1 Paradox: large n vs. the ensemble average

Now for the paradox. Imagine $n \to \infty$ in Figure 7.8. Will we see the same limit as results from the **ensemble average**? Here are two contradictory points of view:

- For increasing n, the fluctuations on the nonzero autocorrelation lags get smaller, so the autocorrelation should tend to an impulse function. Its Fourier transform, the spectrum, should tend to a constant.

- On the other hand, for increasing n, as in Figure 7.3, the spectrum does not get any smoother, because the FTs should still look like random noise.

We will see that the first idea contains a false assumption. The autocorrelation does tend to an impulse, but the fuzz around the sides cannot be ignored—although the fuzz tends to zero amplitude, it also tends to infinite extent, and the product of zero with infinity here tends to have the same energy as the central impulse.

To examine this issue further, let us discover how these autocorrelations decrease to zero with n (the number of samples). Figure 7.9 shows the autocorrelation samples as a function of n in steps of n increasing by factors of four. Thus $\sqrt{n}$ increases by factors of two. Each autocorrelation in the figure was normalized at zero lag. We see the sample variance for nonzero lags of the autocorrelation dropping off as $\sqrt{n}$. We also observe that the ratios between the values for the first nonzero lags and the value at lag zero roughly fit $1/\sqrt{n}$. Notice also that the fluctuations drop off with lag. The drop-off goes to zero at a lag equal to the sample length, because the number of terms in the autocorrelation diminishes to zero at that lag. A first impression is that the autocorrelation fits a triangular envelope. More careful inspection, however, shows that the triangle bulges upward at wide offsets, or large values of k (this is slightly clearer in Figure 7.8). Each of these observations has an analytic explanation found in **PVI**.

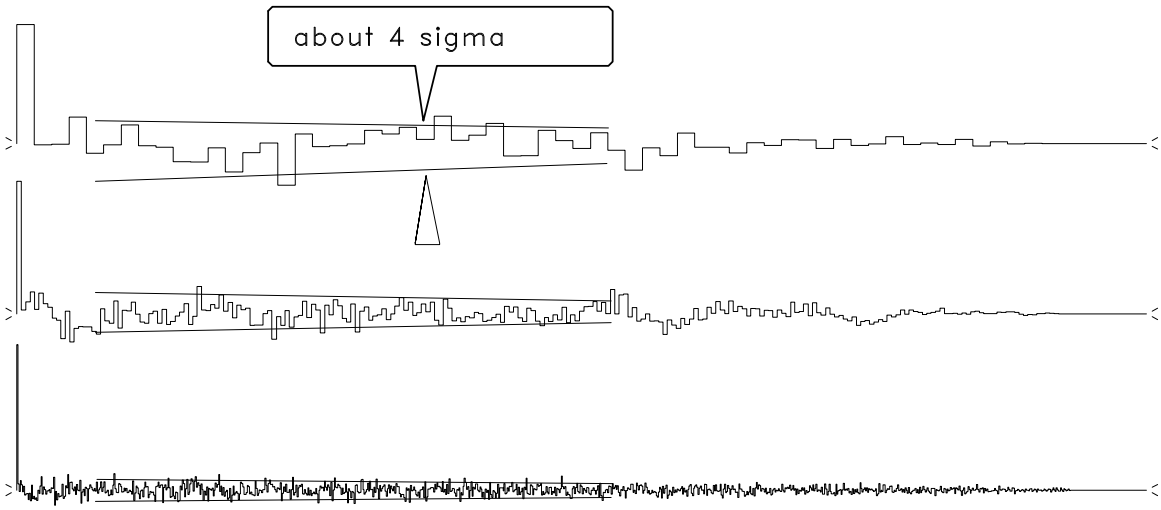

Figure 7.9: Autocorrelation as a function of number of data points. The random-noise-series (even) lengths are 60, 240, 960. rnd-fluct [NR]

7.4.2 An example of the bandwidth/reliability tradeoff

Letting n go to infinity does not take us to the expectation $\hat{S} = \sigma^2$. The problem is, as we increase n, we increase the **frequency resolution** but not the **statistical resolution** (i.e., the fluctuation around $\hat{S}$). To increase the statistical resolution, we need to simulate ensemble averaging. There are two ways to do this:

1. Take the sample of n points and break it into k equal-length segments of n/k points each. Compute an $S(\omega)$ for each segment and then average all k of the $S(\omega)$ together. The variance of the average spectrum is equal to the variance of each spectrum (σ_x^2) *divided* by the number of segments, and so the fluctuation is substantially reduced.

2. Form $S(\omega)$ from the n-point sample. Replace each of the $n/2$ independent amplitudes by an average over its k nearest neighbors. This could also be done by tapering the autocorrelation.

The second method is illustrated in Figure 7.10. This figure shows a **noise burst** of 240 points. Since the signal is even, the burst is effectively 480 points wide, so the autocorrelation is 480 points from center to end: the number of samples will be the same for all cases. The spectrum is very rough. Multiplying the autocorrelation by a triangle function effectively smooths the spectrum by a sinc-squared function, thus reducing the **spectral resolution** ($1/\Lambda F$). Notice that ΛF is equal here to the width of the sinc-squared function, which is inversely proportional to the length of the triangle (ΛT_{auto}).

However, the first taper takes the autocorrelation width from 480 lags to 120 lags. Thus the spectral fluctuations ΛS should drop by a factor of 2, since the count of terms s_k in $S(\omega)$ is reduced to 120 lags. The width of the next weighted autocorrelation width is dropped from 480

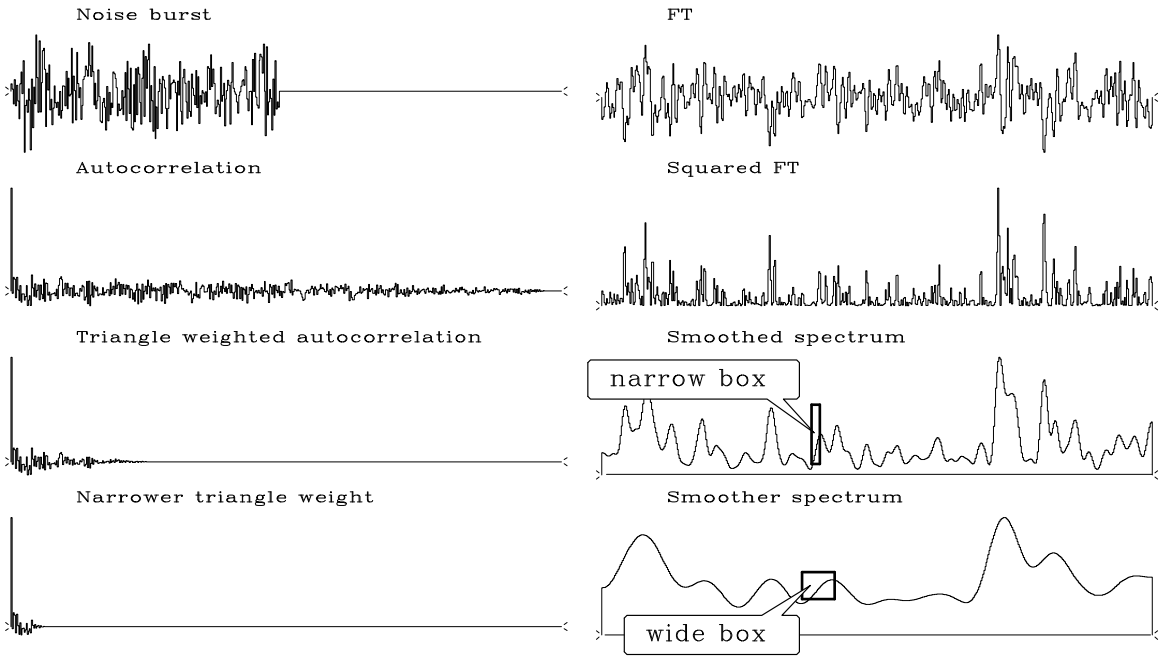

Figure 7.10: Spectral smoothing by tapering the autocorrelation. ΛT is constant and specified on the top row. Successive rows show ΛF increasing while ΛS decreases. The width of a superimposed box roughly gives ΛF, and its height roughly gives ΛS. rnd-taper [NR]

to 30 lags. Spectral roughness should consequently drop by another factor of 2. In all cases, the *average* spectrum is unchanged, since the first lag of the autocorrelations is unchanged. This implies a reduction in the relative spectral fluctuation proportional to the square root of the length of the triangle ($\sqrt{\Lambda T_{\text{auto}}}$).

Our conclusion follows:

The trade-off among **resolution**s of time, frequency, and spectral amplitude is

$$\Lambda F \, \Lambda T \left(\frac{\Lambda S}{S} \right)^2 \quad > \quad 1 \tag{7.30}$$

7.4.3 Spectral estimation

In Figure 7.10 we did not care about **spectral resolution**, since we knew theoretically that the spectrum was white. But in practice we do not have such foreknowledge. Indeed, the random factors we deal with in nature rarely are white. A widely used model for naturally occurring random functions, such as microseism, or sometimes reflection seismograms, is **white noise** put into a filter. The spectra for an example of this type are shown in Figure 7.7. We can see that smoothing the envelope of the power spectrum of the output gives an estimate of the

spectrum of the filter. But we also see that the estimate may need even more smoothing.

7.5 CROSSCORRELATION AND COHERENCY

With two time series we can see how crosscorrelation and coherency are related.

7.5.1 Correlation

"**Correlation**" is a concept similar to cosine. A cosine measures the angle between two vectors. It is given by the dot product of the two vectors divided by their magnitudes:

$$c = \frac{(\mathbf{x}\cdot\mathbf{y})}{\sqrt{(\mathbf{x}\cdot\mathbf{x})(\mathbf{y}\cdot\mathbf{y})}} \tag{7.31}$$

This is the **sample normalized correlation**.

Formally, the **normalized correlation** is defined using x and y as zero-mean, scalar, random variables instead of sample vectors. The summation is thus an **expectation** instead of a dot product:

$$c = \frac{\mathrm{E}(xy)}{\sqrt{\mathrm{E}(x^2)\mathrm{E}(y^2)}} \tag{7.32}$$

A practical difficulty arises when the **ensemble averaging** is simulated over a sample. The problem occurs with small samples and is most dramatically illustrated when we deal with a sample of only one element. Then the sample correlation is

$$\hat{c} = \frac{xy}{|x||y|} = \pm 1 \tag{7.33}$$

regardless of what value the random number x or the random number y should take. For any n, the sample correlation $\hat{c}$ scatters away from zero. Such scatter is called "bias." The topic of bias and variance of coherency estimates is a complicated one, but a rule of thumb seems to be to expect bias and variance of $\hat{c}$ of about $1/\sqrt{n}$ for samples of size n. Bias, no doubt, accounts for many false "discoveries," since cause-and-effect is often inferred from correlation.

7.5.2 Coherency

The concept of "**coherency**" in time-series analysis is analogous to correlation. Taking x_t and y_t to be time series, we find that they may have a mutual relationship which could depend on time delay, scaling, or even filtering. For example, perhaps $Y(Z) = F(Z)X(Z) + N(Z)$, where $F(Z)$ is a filter and n_t is unrelated noise. The generalization of the correlation concept is to define coherency by

$$C = \frac{\mathrm{E}\left[X\left(\frac{1}{Z}\right)Y(Z)\right]}{\sqrt{\mathrm{E}(\overline{X}X)\mathrm{E}(\overline{Y}Y)}} \tag{7.34}$$

Correlation is a real scalar. ***Coherency*** is a complex function of frequency; it expresses the frequency dependence of correlation. In forming an estimate of coherency, it is always essential to simulate ensemble averaging. Note that if the ensemble averaging were to be omitted, the coherency (squared) calculation would give

$$|C|^2 \quad = \quad \overline{C}C \quad = \quad \frac{(\overline{\overline{X}Y})(\overline{X}Y)}{(\overline{\overline{X}X})(\overline{\overline{Y}Y})} \quad = \quad 1 \tag{7.35}$$

which states that the coherency squared is unity, independent of the data. Because correlation scatters away from zero, we find that coherency squared is biased away from zero.

Index

Jon F. Claerbout (M.I.T., B.S. physics, 1960; M.S. 1963; Ph.D. geophysics, 1967), professor at Stanford University, 1967. Best Presentation Award from the Society of Exploration Geophysicists (SEG) for his paper, *Extrapolation of Wave Fields.* Honorary member and SEG Fessenden Award "in recognition of his outstanding and original pioneering work in seismic wave analysis." Founded the Stanford Exploration Project (SEP) in 1973. Elected Fellow of the American Geophysical Union. Authored three published books and five internet books. Elected to the National Academy of Engineering. Maurice Ewing Medal, SEG's highest award. Honorary Member of the European Assn. of Geoscientists & Engineers (EAGE). EAGE's highest recognition, the Erasmus Award.